



HAL
open science

**Paramétrisations stochastiques de processus
biogéochimiques non résolus dans un modèle couplé
NEMO/PISCES de l'Atlantique Nord : Applications
pour l'assimilation de données de la couleur de l'océan**

Florent Garnier

► **To cite this version:**

Florent Garnier. Paramétrisations stochastiques de processus biogéochimiques non résolus dans un modèle couplé NEMO/PISCES de l'Atlantique Nord : Applications pour l'assimilation de données de la couleur de l'océan. Océanographie. Université Grenoble Alpes, 2016. Français. NNT : 2016GREAU044 . tel-01661414

HAL Id: tel-01661414

<https://theses.hal.science/tel-01661414>

Submitted on 11 Dec 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

THÈSE

Pour obtenir le grade de

DOCTEUR DE LA COMMUNAUTÉ UNIVERSITÉ GRENOBLE ALPES

Spécialité : **Océanographie, Atmosphère et Hydrologie** Arrêté
ministériel : 7 août 2006

Présentée par

Florent Garnier

Thèse dirigée par **Pierre Brasseur**
et codirigée par **Emmanuel Cosme**

préparée au sein du **Laboratoire de Glaciologie et Géophysique
de l'Environnement (LGGE)**
et de l'école doctorale **Terre, Univers et Environnement (TUE)**

Paramétrisations stochastiques de proces- sus biogéochimiques non résolus dans un modèle couplé NEMO/PISCES de l'Atlan- tique Nord.

Applications pour l'assimilation de données de
la couleur de l'océan

Thèse soutenue publiquement le , 10 MAI 2016
devant le jury composé de :

M. Didier Voisin

Professeur UJF (LGGE, France), Président

M. Frédéric Chevallier

Ingénieur-chercheur CEA (LSCE, France), Rapporteur

Mme Marilaure Grégoire

Directrice de Recherche FNRS (MARE, Belgique), Rapporteur

M. Pierre-Yves Le Traon

Directeur de Recherche IFREMER (Technopôle Brest-Iroise, France),
Examinateur

M. Fabrizio D'Ortenzio

Chargé de Recherche CNRS (LOV, France), Examinateur

M. Pierre Brasseur

Directeur de Recherche CNRS (LGGE, France), Directeur de thèse

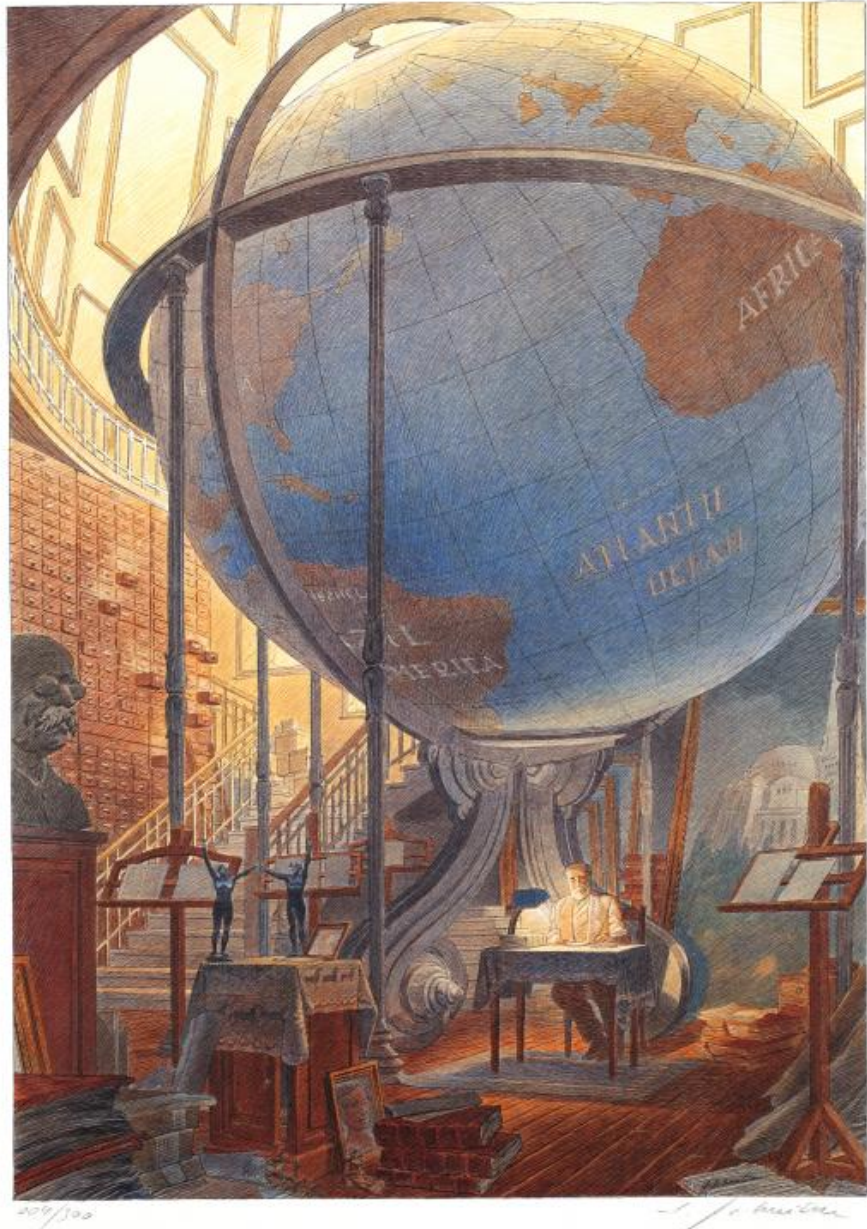
M. Emmanuel Cosme

Maître de conférence UJF (LGGE, France), Co-Directeur de thèse



*Quels que soient les progrès des connaissances humaines,
il y aura toujours place pour l'ignorance et par suite,
pour le hasard et la probabilité*

Émile Borel (1871-1956)



Paul Otlet¹ et le Mundaneum, 1998.
L'homme qui voulait classer le monde.
Lithographie illustrée de **François Schuiten**.

¹ Paul Otlet (1864-1944) est un documentaliste belge, souvent perçu comme l'un des précurseurs européens d'internet. En 1895, il créa notamment avec Henri Lafontaine, prix nobel de la paix en 1913, l'Office International de Bibliographie à Bruxelles, avec pour ambition de cataloguer l'ensemble des publications existantes à travers le monde. Sa quête perpétuelle de connaissances est aujourd'hui reconnue pour avoir été à la base du développement de la science de l'information.

Remerciements

Il est vrai qu'il n'est pas rare de se sentir quelque peu face à soi-même, pendant les derniers mois d'écriture d'une thèse. Ce fut également mon cas. Un peu à l'image de Paul Otlet sur l'illustration en page de gauche, submergé par le poids de la complexité de la planète Océan... Un projet scientifique n'est pourtant jamais le fruit d'un travail solitaire, comme en atteste la liste impressionnante d'auteurs ayant participé à l'observation récente des ondes gravitationnelles (Abbott et al., 2016). Remercier les personnes qui m'ont accompagné de près ou de loin dans cette enrichissante expérience est le moindre des hommages que je puisse leur rendre.

Tout d'abord, je souhaite remercier profondément mon directeur de thèse, *Pierre Bresseur*, pour m'avoir accordé sa confiance et offert l'opportunité de réaliser cette thèse. Merci de m'avoir soutenu avec respect. La seconde personne que je souhaite remercier très chaleureusement est *Jean-Michel Brankart*, sans qui je serais encore "probablement" en train de peaufiner les paramétrisations stochastiques. Et pour nos échanges... Merci également à *Emmanuel Cosme*, mon co-directeur de thèse, qui a su se rendre disponible pour moi et me faciliter le lien avec l'assimilation de données. Enfin, un grand merci à *l'ensemble des personnes de l'équipe MEOM*, avec qui les discussions m'ont permis de me positionner dans un cadre global d'océanographie et de pouvoir avancer personnellement. Un remerciement particulier à *Bernard Barnier* pour les précisions de sa relecture sur la partie NEMO.

Évidemment une attention sincère à tous les étudiants que j'ai pu croiser, et avec qui j'ai maintes fois eu l'occasion de discuter de science et de bien d'autres choses. Un clin d'oeil plus spécifique à *Gildas Mainsant*, *Giovanni Ruggiero*, *Sammy Metref*, *Guillaume Serazin*, *Pierre Labreuche*, *Lucile Gaultier*, *Natacha Djath*, *Nicolas Ducouso* et bonne chance aux futurs docteurs *Nacho Merino*, *Christophe Calone* et *Marina Durand*.

Merci également à toute l'équipe de MERCATOR-Océan de s'être intéressée à mon travail et de m'avoir permis d'accéder à la simulation BIOMER. Un geste plus personnel à *Charles Emmanuel Testut*, *Coralie Perruche*, *Marion Ghelen* et *Yann Le Drillet*. Merci aussi à *Raphaëlle Sauzede* du LOV pour l'utilisation de sa climatologie verticale de chlorophylle. Pour cette magique expérience de campagne en mer de Salomon, merci beaucoup à *Sophie Cravatte* et *Alexandre Ganachaud* du LEGOS. Je tiens également à remercier *Laurent Memery* du LEMAR et *Patrick Lehodey* de CLS, d'avoir accepté de faire partie de mon comité de thèse. Enfin, je remercie bien évidemment les membres de mon jury de thèse : *Didier Voisin*, *Frédéric Chevallier*, *Marilaure Grégoire*, *Pierre-Yves Le Traon* et *Fabrizio D'Ortenzio*.



Je profite de l'occasion pour envoyer un message très fort d'amitié à tous *mes amis de Charente Maritime*, dont une grande partie va me faire l'honneur de se déplacer jusqu'ici pour ma soutenance. Merci à vous tous, pour nos longues années d'amitié passées et d'avance pour celles qui restent à venir. Je serre du plus fort que je peux dans mes bras ma famille, qui m'aide chaque jour pour tout depuis toujours : merci *Maman*, merci *Tantine* (vous m'avez même aidé à corriger ma thèse...), merci *Philippe* mon oncle, merci mon frère *Valérian* et merci *Gautier* mon cousin. Enfin je te remercie du fond du cœur *Anaïs*, pour avoir été présente à mes côtés et pour ta bienveillance dans cette difficile épreuve de rédaction. Les étincelles de couleur de tes peintures m'ont accompagné tous les jours et le feront j'espère encore longtemps.

À tous ceux qu'on oublie parfois, mais qui participent en permanence à notre construction personnelle : merci à vous tous. Un dernier hommage pour *l'océan* qui m'a toujours accompagné, ne cessera de me fasciner, et va même aujourd'hui jusqu'à me faire travailler...



Abstract

In spite of recent advances, biogeochemical models are still unable to represent the full complexity of marine ecosystems. Since mathematical formulations are still based on empirical laws involving many parameters, it is now well established that the uncertainties inherent to the biogeochemical complexity strongly impact the model response. Improving model representation therefore requires to properly describe model uncertainties and their consequences. Moreover, in the context of ocean color data assimilation, one of the major issue rely on our ability to characterize the model uncertainty (or equivalently the model error) in order to maximize the efficiency of the assimilation system. This is exactly the purpose of this PhD which investigates the potential of using random processes to simulate some biogeochemical uncertainties of the $1/4^\circ$ coupled physical–biogeochemical NEMO/PISCES model of the North Atlantic ocean.

Starting from a deterministic simulation performed with the original PISCES formulation, we propose a generic method based on AR(1) random processes to generate perturbations with temporal and spatial correlations. These perturbations are introduced into the model formulations to simulate 2 classes of uncertainties : the uncertainties on biogeochemical parameters and the uncertainties induced by unresolved scales in the presence of non-linear processes. Using these stochastic parameterizations, a probabilistic version of PISCES is designed and a 60-member ensemble simulation is performed.

The implications of this probabilistic approach is assessed using the information of the probability distributions given of this ensemble simulation. The relevance and the impacts of the stochastic parameterizations are assessed from a comparison with SeaWiFS satellite data. In particular, it is shown that the ensemble simulation is able to produce a better estimate of the surface chlorophyll concentration than the first guess deterministic simulation. Using SeaWiFS ocean color data observations, the statistical consistency (reliability) of this prior ensemble is demonstrated using rank histograms.

Finally, the relevance of our approach in the prospect of ocean color data assimilation is demonstrated by considering a 3D optimal analysis of the ensemble (one update at one time step) performed from the statistic errors of the stochastic ensemble simulation previously stated. During this experiment, the high resolution SeaWiFS ocean color data are assimilated using a Ensemble Transform Kalman Filter (ETKF) analysis scheme and the non gaussian behaviour and non linear relationship between variables are taken into account using anamorphic transformations. More specifically, we show that the analysis of SeaWiFS data improves the representation and the ensemble statistics of chlorophyll concentrations.

Principales activités scientifiques menées durant la thèse

Publications

Garnier, F., Brankart, J.-M., Brasseur, P., Cosme, E., 2016. Stochastic parameterizations of biogeochemical uncertainties in a 1/4° nemo/pisces model for probabilistic comparisons with ocean color data. *Journal of Marine Systems* vol.155, p.59–72.

Brankart, J., Candille, G., **Garnier, F.**, Calone, C., Melet, A., Bouttier, P., Brasseur, P., Verron, J., 2015. A generic approach to explicit simulation of uncertainty in the nemo ocean model. *Geoscientific Model Development* vol.8 (5), p.1285–1297.

Principales Conférences

International Liège Colloquium of Ocean Dynamics, Liège, mai 2013. *Présentation poster.*

Ocean Sciences Meeting, Hawaii, février 2014. *Présentation poster.*

MyOcean Science Days, Toulouse, septembre 2014. *Présentation poster.*

International Liège Colloquium of Ocean Dynamics, Liège, mai 2015. *Présentation orale.*

GODAE OceanView MEOPAR/GOV workshop, Halifax, juin 2015. *Présentation orale.*

Atlantos OSSE workshop, Toulouse, décembre 2015. *Présentation orale.*

Réunion annuelle de l'équipe MEOM avec le CNES, 2013-2015. *Présentations orales.*

Campagne océanographique

Mars-avril 2014 : participation à la campagne océanographique Moor SPICE à bord du R.V Thompson de l'université de Washington. Relevés et analyses journaliers des teneurs en oxygène et en nutriments de l'eau en mer de Salomon. Déploiement de balises Argos et de bouées flottantes. Coordinateurs : *Sophie Cravatte*, IRD Nouméa et *Janet Springtall*, Scripps USA.

Résumé

En dépit de progrès croissants durant la dernière décennie, la complexité des écosystèmes marins est encore imparfaitement simulée par les modèles. Les formulations des processus biogéochimiques sont en général établies de manière empirique et contraintes par une multitude de paramètres. Il est ainsi généralement admis que leurs incertitudes impactent fortement l'estimation de la production primaire, dont le rôle dans le cycle du carbone est primordial. Analyser les impacts de l'incertitude des modèles est donc nécessaire pour améliorer la représentation des caractéristiques biogéochimiques de l'océan. Dans le contexte d'assimilation de données de la couleur de l'océan, la définition des erreurs de prévision représente de plus un important verrou aux performances des systèmes. Ces points seront analysés dans cette thèse. L'objectif sera d'examiner, dans un contexte de modélisation/assimilation, la pertinence d'utiliser une approche probabiliste basée sur une simulation explicite des incertitudes biogéochimiques du modèle couplé au 1/4° NEMO/PISCES sur l'océan Atlantique Nord.

À partir d'une simulation déterministe du modèle PISCES, nous proposerons une méthode pour générer des processus aléatoires, AR(1), permettant d'inclure des structures spatiales et temporelles de corrélations. À chaque pas de temps, ces perturbations aléatoires seront ensuite introduites dans le modèle par l'intermédiaire de paramétrisations stochastiques. Elles simuleront 2 différentes classes d'incertitudes : les incertitudes sur les paramètres biogéochimiques du modèle et les incertitudes dues aux échelles non résolues dans le cas d'équations non linéaires. L'utilisation de paramétrisations stochastiques permettra ainsi d'élaborer une version probabiliste du modèle PISCES, à partir de laquelle nous pourrons réaliser une simulation d'ensemble de 60 membres.

La pertinence de cette simulation d'ensemble sera évaluée par comparaison avec les observations de la couleur de l'océan SeaWIFS. Nous montrerons en particulier que la simulation d'ensemble conserve les structures de grande échelle présentes dans la simulation déterministe. En utilisant les distributions de probabilité définies par les membres de l'ensemble, nous montrerons que l'ensemble capture l'information des observations avec une bonne estimation de leurs statistiques d'erreur (fiabilité statistique). L'intérêt de l'approche probabiliste sera ainsi d'abord évalué dans un contexte de modélisation biogéochimique.

Dans la dernière partie de cette thèse, les bénéfices de cette approche probabiliste seront démontrés dans un contexte d'assimilation de données. En utilisant les statistiques d'erreur définies par la simulation d'ensemble, nous réaliserons une séquence d'analyses d'ensemble des observations SeaWIFS à haute résolution (1/24°) basée sur le schéma d'analyse du Filtre d'Ensemble de Kalman Transformé (ETKF). En plus de démontrer la faisabilité de ce type d'expérience, nous verrons que cette méthode permet d'améliorer significativement les caractéristiques statistiques des distributions de la chlorophylle. L'impact des corrections sur les distributions verticales de la chlorophylle et sur les variables biogéochimiques non observées sera également étudié.

Table des matières

1	Introduction	7
2	Conception probabiliste des incertitudes et assimilation de données	15
2.1	Hasards mathématiques	17
2.2	Modéliser la biogéochimie marine	20
2.2.1	Omniprésence des incertitudes	20
2.2.2	Complexité des modèles	21
2.3	L'approche probabiliste	23
2.3.1	Simulation d'ensemble	24
2.3.2	Simulation stochastique des incertitudes	26
2.4	L'assimilation de données	27
2.4.1	Formalisme probabiliste	28
2.4.2	Étape d'analyse : l'interpolation statistique	29
2.4.3	D'un filtre de rang réduit à un filtre d'ensemble	30
3	Les observations satellites et le modèle couplé NEMO/PISCES	33
3.1	L'observation satellite de la couleur de l'océan	34
3.1.1	Principe de la mesure et incertitudes	34
3.1.2	Les différentes missions	36
3.1.3	État de l'art en assimilation de données couleur de l'eau	38
3.2	Le modèle couplé NEMO/PISCES	40
3.2.1	Le modèle de circulation océanique OPA	41
3.2.2	Le modèle biogéochimique PISCES	45
3.2.2.1	Complexité biogéochimique	45
3.2.2.2	Équations de la production primaire	48
3.3	Mise en place du modèle couplé au 1/4° sur l'Atlantique Nord	53
3.3.1	Caractéristiques du couplage	53
3.3.2	Configuration du modèle	54
3.4	Conclusion	59
4	Une simulation déterministe du modèle couplé NEMO/PISCES dans l'océan Atlantique Nord	61
4.1	La production primaire dans l'océan Atlantique Nord	62
4.2	De LOBSTER à PISCES	65
4.2.1	Chlorophylle de surface	65
4.2.2	Distribution verticale de la chlorophylle	68

4.3	Variabilité de la simulation déterministe	71
4.3.1	Variabilité interannuelle de la chlorophylle	71
4.3.2	Comparaison avec les observations SeaWIFS	77
4.3.3	Diversité des espèces	81
4.4	Conclusion	84
5	Mise en place d'une version probabiliste du modèle PISCES	87
5.1	Simulation explicite d'incertitudes biogéochimiques	89
5.1.1	Une approche stochastique	89
5.1.1.1	Démarche et objectifs	89
5.1.1.2	Génération des perturbations	91
5.1.2	Pour quelles incertitudes ?	93
5.1.2.1	Incertitudes sur les paramètres biologiques	93
5.1.2.2	Incertitudes des échelles non résolues	95
5.2	Ajustement conjugué des paramètres stochastiques et déterministes du modèle	97
5.2.1	Des paramétrisations indissociables du modèle...	98
5.2.2	...impliquant un réajustement des paramètres déterministes	100
5.2.3	Conclusion partielle	103
5.3	Impacts des paramétrisations stochastiques	103
5.3.1	Répartition de la chlorophylle de surface	104
5.3.2	Structures verticales de la chlorophylle	106
5.4	Conclusion	109
6	Réponse probabiliste aux incertitudes biogéochimiques	111
6.1	Dispersion statistique de la chlorophylle de surface	112
6.1.1	Quantiles de surface	112
6.1.2	Évolution temporelle	116
6.1.3	Conclusion partielle	118
6.2	Fiabilité statistique de l'ensemble	119
6.2.1	Démarche	119
6.2.2	Histogrammes de rangs	120
6.2.2.1	Caractéristiques spatiales	120
6.2.2.2	Caractéristiques temporelles	123
6.3	Distributions verticales de la chlorophylle	126
6.3.1	Dispersion d'ensemble	126
6.3.2	Comparaison aux observations	129
6.3.2.1	Climatologie verticale	129
6.3.2.2	Diversité des comportements biologiques	131
6.4	Conclusion	135
7	Une séquence d'analyses 3D de données de la couleur de l'eau	137
7.1	Méthode d'analyse d'ensemble	139
7.1.1	Le filtre de Kalman	139
7.1.2	Le filtre de Kalman transformé (ETKF)	140
7.1.3	Prise en compte des non-gaussianités	143
7.2	Un nouvel ensemble d'états actualisés	145

7.2.1	Impact des analyses sur l'état moyen	146
7.2.2	Incertitudes des analyses	148
7.2.3	Extrapolation verticale	151
7.3	Comparaison statistique aux données de la couleur de l'eau	155
7.3.1	Rangs des observations MERIS	155
7.3.2	Score probabiliste CRPS	159
7.3.2.1	Variabilité temporelle	160
7.3.2.2	Décomposition du score	161
7.4	Correction des variables non observées	164
7.4.1	Dispersion des séries temporelles	164
7.4.2	Dispersion des profils verticaux	166
7.5	Conclusion	168
8	Conclusion	171
8.1	Démarche, développements et résultats	172
8.2	Perspectives	176
A	Namelist du modèle OPA	203
B	Namelist du modèle PISCES	213
C	Namelist TOP	217
D	Publication Garnier et al. (2016)	221

Chapitre 1

Introduction

Le travail de thèse présenté dans ce document fait partie d'une thématique large visant à améliorer la représentation des caractéristiques biogéochimiques de l'océan. Cette introduction a pour vocation d'expliquer précisément le contexte, les motivations et les objectifs de cette étude, ainsi que l'organisation de ce manuscrit.

Telle que définie dans le livre de Sarmiento (2013), la biogéochimie marine regroupe l'ensemble des activités s'articulant autour de la distribution des éléments chimiques de l'océan et des processus qui la contrôlent. À l'origine de l'étude de la composition chimique des océans, les travaux d'Albert Redfield (Redfield, 1934) ont, dès le départ, mis en avant le rôle clé des micro-organismes marins : le plancton. Les analyses, qu'il effectue pour déterminer le contenu de l'eau de mer révèlent, en effet, un rapport constant entre les teneurs en nitrate (NO_3) et en phosphate (PO_4), retrouvées dans les mêmes proportions au sein du plancton (Redfield, 1958). En partant du constat que leur production de matière organique se fait par photosynthèse (production primaire), c'est-à-dire en utilisant la lumière comme énergie, l'étude de Redfield et al. (1963) popularise ensuite le concept des rapports de Redfield. Elle démontre que l'activité des micro-organismes végétaux, appelés phytoplancton, modifie les propriétés chimiques de l'environnement marin selon des proportions stœchiométriques bien définies : $\Delta\text{C}/\Delta\text{N}/\Delta\text{P}/\Delta\text{O}_2 = 106/16/1/-138$. Si cette vision relativement simpliste a depuis largement évolué, l'influence du phytoplancton sur les cycles des éléments chimiques n'en reste pas moins cruciale, s'étendant jusqu'à l'ensemble du système climatique. Rappeler que cette vie microscopique représente presque 50% de la production végétale de la planète (Lévy and Bopp, 2007) et qu'elle fournit environ les 2/3 de l'oxygène présent dans l'atmosphère¹ (plus que toutes les forêts réunies, Mollo and Noury, 2013) suscite en général l'étonnement. Le rôle de la production primaire des océans dans le contrôle du CO_2 atmosphérique est pourtant aujourd'hui devenu indéniable.

La proposition faite par John Martin² : « *Give me a half a tanker of iron and I will give you another ice age* » illustre bien l'impact planétaire que peut prendre l'activité biologique

1. Le phytoplancton est le premier type d'organisme vivant apparu sur terre. Sa capacité à rejeter de l'oxygène dans l'atmosphère est à l'origine de la création de la couche d'ozone qui a permis par la suite le développement de formes de vie plus élaborées.

2. John Martin a proposé cette « Iron hypothesis » en 1988, lors d'une conférence à l'institut océanographique de Woods Hole. Il suggère de réduire le CO_2 atmosphérique en stimulant la croissance du phytoplancton dans les régions où elle est limitée par l'absence de fer. Difficilement réalisable, l'impact de cette fertilisation de l'océan sur le climat a toutefois pu être démontrée théoriquement (Martin et al., 1990).

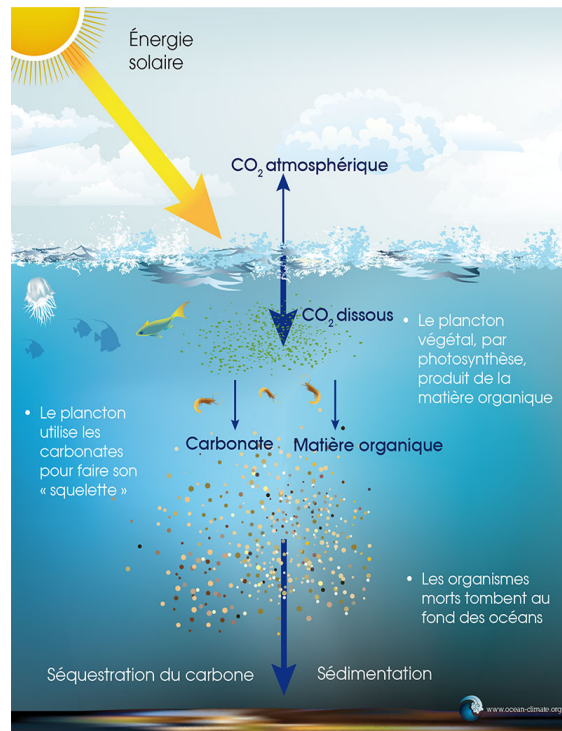


FIGURE 1.1 – Représentation simplifiée de la séquestration du carbone atmosphérique par l'océan.
Source : www.ocean-climate.org

des océans. Lors de la photosynthèse, le phytoplancton extrait le carbone inorganique dissous dans les hautes couches de l'océan pour se développer. Cette production primaire est régulée par l'apport de lumière et la disponibilité en éléments nutritifs tels que l'azote (N), le phosphate (P) ou encore le fer : d'où la suggestion de John Martin de stimuler la croissance du phytoplancton en fertilisant l'océan. À la mort de ces végétaux, une partie de leur matière organique carbonée est transportée vers le fond des océans où elle reste stockée sous forme de sédiments (cf. figure 1.1). Cette séquestration du carbone conduit à une diminution de pression partielle qui favorise la dissolution du CO_2 de l'atmosphère dans les eaux de surface (à l'interface océan-atmosphère). Ce carbone, dissous dans l'océan, se retrouve alors à la disposition des organismes marins.

Si les processus physico-chimiques qui contrôlent la séquestration du carbone sont évidemment bien plus complexes, la production primaire aboutit in fine à une absorption de CO_2 atmosphérique par l'océan. Les études mentionnent des quantités de carbone comprises entre 9 et 28 Gt.C.an^{-1} , ainsi exportées vers l'océan intérieur (Najjar et al., 2007). Si l'évocation d'une telle gamme de valeurs souligne d'ores et déjà obligatoirement le caractère incertain de ce type de prévisions, il est généralement admis que l'océan aurait absorbé, depuis le début de l'ère industrielle, environ 30% des émissions de carbone atmosphérique d'origine anthropique (Bopp et al., 2015). Sans l'existence d'un tel réservoir, les concentrations de l'atmosphère en CO_2 seraient aujourd'hui nettement plus élevées (Sabine et al., 2004). Prévoir la réponse des écosystèmes marins aux futures évolutions climatiques est donc aujourd'hui un enjeu majeur pour l'étude du cycle du carbone (p.ex. Sarmiento and Bender, 1994; Sigman and Boyle, 2000; Goodwin et al., 2008; Ridgwell and Schmidt,

2010).

Modéliser la biogéochimie

Face aux besoins de prévoir l'évolution des stocks de carbone, comprendre les processus qui contrôlent la production primaire apparaît être un élément essentiel. Un effort de plus en plus conséquent est ainsi mené pour tenter de modéliser les fonctionnements biogéochimiques de l'océan. Mais mathématiser le vivant est loin d'être une chose aisée. Même dans le cas d'organismes majoritairement unicellulaires comme le plancton, les processus qui régulent les comportements biologiques sont complexes, mal connus et fortement dépendants de la dynamique des courants marins dont la modélisation repose, elle aussi, sur d'importantes approximations. Au premier ordre, caractériser l'état biogéochimique de l'océan nécessite donc une représentation réaliste de la physique de l'océan (Doney, 1999). L'évaluation de la production primaire présuppose ainsi de coupler la biogéochimie à des modèles de circulation océanique tel le modèle NEMO-OPA³ (Madec et al., 1998).

Parmi les mécanismes physiques, les travaux de Oschlies and Garçon (1998b) font partie des premiers à avoir mis en évidence l'importance de l'activité océanique de petite échelle (turbulence) sur la production primaire. Disposer d'un modèle capable de simuler en partie cette activité de submésos-échelle⁴ (modèle au minimum « eddy-permitting ») est un pré requis essentiel à une représentation réaliste de la biogéochimie marine (p.ex. Lévy and Martin, 2013). Mais même en imaginant pouvoir représenter parfaitement les caractéristiques physiques de l'océan, faut-il encore être capable de modéliser correctement les processus biogéochimiques.

Les premiers modèles simplifiés d'écosystèmes marins (p.ex. Fasham et al., 1990) étaient basés sur la simulation des interactions entre les grands compartiments biogéochimiques : les Nutriments, le Phytoplancton, le Zooplancton et les Détritus. L'inconvénient de ce type de modèles, dits modèles NPZD, est qu'ils ne permettent pas de distinguer la diversité des espèces, qui joue pourtant un rôle important dans le cycle du carbone (Maier-Reimer et al., 1996; Bopp et al., 2005). L'approche développée aujourd'hui est donc de spécifier, à l'intérieur de ces grands compartiments, l'existence de différents groupes fonctionnels possédant chacun un rôle distinct au sein de l'écosystème (p.ex. le modèle PISCES, Aumont et al., 2003). Ces modèles permettent également d'aller au-delà de l'hypothèse simplificatrice de Redfield. Toujours est-il que leurs formulations impliquent une multitude de paramètres dont les valeurs sont, en général, définies de manière approximative. Le niveau de diversité biologique, comme la quantité de processus que l'on peut simuler restent, de plus, limités par notre connaissance a priori des phénomènes et par les ressources en calcul numérique.

En définitive, la représentation des caractéristiques biogéochimiques de l'océan reste sujette à de nombreuses contraintes qui rendent leur modélisation largement incertaine. Tenir compte de ces incertitudes et évaluer leurs impacts sur l'estimation de la production primaire sera un aspect fondamental de cette étude.

3. NEMO : Nucleus for European Modelling of the Ocean ; OPA : Ocean PARallélisé

4. L'activité de submésos-échelle regroupe l'ensemble des processus dont la taille caractéristique est comprise entre une dizaine et une centaine de kilomètres. Parmi ces processus, le mélange vertical est certainement celui qui a le plus d'influence sur la production primaire puisqu'il régule en grande partie les apports de nutriments vers la surface.

Observer la production primaire

Compte tenu de l'incertitude des modèles, caractériser la production primaire de l'océan nécessite de recourir à des moyens supplémentaires. L'autre type d'information dont nous disposons est l'observation. Outre les mesures in-situ, réalisées en général par bateau, mouillages ou à l'aide de flotteurs lagrangiens de type BIO-ARGO, la mesure spatiale (par satellite) est le système d'observation le plus perfectionné actuellement. Elle est perçue comme une avancée scientifique majeure, qui a profondément influencé notre conception de l'océanographie (Yoder et al., 2010). Depuis 1978 et le lancement du capteur expérimental CZCS (Coastal Zone Color Scanner) embarqué à bord du satellite NIMBUS-7, une partie des caractéristiques biogéochimiques de l'océan peut désormais être mesurée depuis l'espace. En analysant les propriétés optiques de la lumière réfléchie à la surface de l'océan, on peut en déduire la partie du spectre de la lumière ayant été absorbée. L'anomalie par rapport à une absorption "standard" (en absence de vie) permet de mesurer la part de lumière utilisée par les organismes marins pour réaliser la photosynthèse : c'est la mesure de la couleur de l'océan. La qualité des résultats du premier radiomètre CZCS a encouragé de nombreux programmes spatiaux à s'intéresser à la couleur de l'océan. Nous pouvons citer en exemple le satellite SeaStar, muni du capteur SeaWIFS (Sea-viewing Wide Field-of-view Scanner) ou encore le satellite Envisat avec le capteur MERIS (Medium Resolution Imaging Spectrometer). Ces missions se sont surtout penchées sur la détermination de la chlorophylle de surface, principal pigment photosynthétique du phytoplancton des hautes couches de l'océan.

Si les observations par satellite produisent des données bénéficiant d'une couverture spatio-temporelle de haute résolution, elles possèdent toutefois les inconvénients d'une mesure à grande distance. Lors de son trajet entre la surface de l'océan et le capteur satellite, le signal optique réfléchi est, en effet, fortement perturbé par ses interactions avec l'atmosphère. D'importants traitements, plus ou moins empiriques (corrections atmosphériques), sont alors nécessaires avant d'aboutir à des données exploitables. Le nombre de variables pouvant être extraites du signal optique est de plus relativement faible comparé à la complexité des cycles biogéochimiques. Dans l'objectif de quantifier la production primaire, l'élément le plus désavantageux reste la restriction des mesures à la surface de l'océan, une partie non négligeable de la photosynthèse se produisant en subsurface.

Malgré les qualités indéniables des observations satellites, elles ne peuvent proposer qu'une vision partielle de l'océan. Elles n'ont pas vocation à remplacer les modèles mais apportent des informations complémentaires importantes. Dans cette thèse, une attention particulière sera portée sur l'évaluation de la pertinence du modèle en le comparant avec les observations de la couleur de l'eau. Les difficultés que nous rencontrerons pour confronter ces 2 sources d'information permettront de justifier l'utilisation d'une approche intégrée modèle/observation.

L'approche intégrée

Que ce soit dans un contexte climatique ou dans un cadre de gestion des ressources environnementales, les besoins croissants en prévisions opérationnelles ont poussé le dé-

veloppement d’approches dites intégrées. L’introduction d’observations satellites dans des modèles 3D couplés physique biogéochimie (BGCMS) est aujourd’hui l’un des moyens les plus prometteurs pour quantifier la production primaire et contrôler les différents flux biogéochimiques (Brasseur et al., 2009). Le principe est relativement simple : combiner l’information des modèles et celle des observations pour produire une représentation aussi précise que possible des caractéristiques biogéochimiques de l’océan. Par l’intermédiaire de méthodes d’assimilation de données, les sorties de modèles sont ainsi corrigées par les observations. La dynamique des modèles permet d’extrapoler ensuite ces améliorations sur la verticale, dans les régions non couvertes par les observations et sur les variables biogéochimiques non observées.

L’assimilation de données se place aujourd’hui au cœur de tous les systèmes opérationnels tel que le système de prévision océanique MERCATOR-Océan. Elle est un élément primordial au déploiement des services d’information environnementale tels que le programme international GODAE Ocean View (Global Ocean Data Assimilation Experiment) ou le programme européen GMES (Global Monitoring for Environment and Security), renommé depuis 2013 COPERNICUS. C’est dans ce cadre de prévision opérationnelle à grande échelle que se place ce projet de thèse. Il bénéficie des moyens mis en place dans le projet MyOcean2 (au sein du programme COPERNICUS) pour participer au développement de méthodes permettant de fournir l’information la plus pertinente sur l’état et l’évolution de la biogéochimie marine. Plus précisément, cette thèse s’inscrit dans la prospective de recherche et développement du projet support SANGOMA (Stochastic Assimilation for the New Generation of Ocean Model Applications) qui vise spécifiquement le développement d’outils d’assimilation pour assurer les besoins futurs des systèmes opérationnels.

Objectifs et démarche de cette thèse

Puisque l’approche intégrée découle du caractère imparfait des modèles et des observations, ses performances dépendent fortement de la définition des incertitudes. Modéliser l’incertitude devient donc aussi important que de modéliser l’état. Ce constat est particulièrement valable dans le cas de la biogéochimie marine. La quantification des incertitudes demeure en effet un important verrou à notre problématique de développement d’outils pour l’assimilation de données de la couleur de l’eau. Si l’erreur d’observation est en général “livrée avec le produit”, une des questions qui constituera le fil conducteur de cette thèse est :

Comment caractériser et représenter les incertitudes des modèles biogéochimiques (ou erreur modèle) dont les équations, comme les paramètres qu’elles contiennent, sont elles-mêmes établies de manière approximative ?

L’analyse des incertitudes sur la formulation des processus biogéochimiques et de leurs impacts sur l’estimation de la chlorophylle nous amènera vers un des points fondateurs de cette étude :

Les incertitudes du modèle permettent-elles d’expliquer les écarts avec les observations de la couleur de l’eau et de produire des distributions d’erreur

pertinentes pour l'assimilation de données ?

Pour répondre à ces questions, une approche d'ensemble sera utilisée et développée dans le cadre de modélisation/assimilation biogéochimique. Le principe est de réaliser simultanément plusieurs simulations (ou simulation d'ensemble) à partir de données d'entrée (paramètres, conditions initiales, forçages...) légèrement différentes. Les écarts entre les différentes simulations caractérisent (de manière réduite) l'erreur des prévisions. Dans notre cas, ils permettront également d'estimer l'incertitude du modèle.

Dans le cas d'une configuration réaliste eddy-permitting du modèle couplé physique-biogéochimie NEMO/LOBSTER (Lévy et al., 2005), plusieurs études menées par l'équipe MEOM⁵ ont déjà abordé la caractérisation d'incertitudes biogéochimiques dans un contexte d'assimilation de données. Au niveau de l'océan Atlantique Nord, Béal et al. (2010) ont décrit l'erreur de prévision à partir d'un ensemble de simulations produit en perturbant le forçage par le vent. Dans une configuration identique, Doron et al. (2013) ont montré qu'à l'échelle d'un bassin, la variabilité spatiale des paramètres du modèle peut expliquer certains écarts avec les observations. L'assimilation de données de la couleur de l'eau a permis de corriger ces paramètres biogéochimiques et d'améliorer la distribution spatiale de la chlorophylle de surface. Enfin, les travaux de Fontana et al. (2013) ont implémenté avec succès un schéma d'assimilation de données permettant, sur la période 1998-2006, d'intégrer tous les 10 jours des données satellites SeaWiFS.

Pour s'accorder aux besoins de l'opérationnel (cadre GMMC⁶/MERCATOR-vert), il devient nécessaire de poursuivre ces travaux sur un modèle de plus grande complexité biogéochimique, en l'occurrence le modèle PISCES (Pelagic Interaction Scheme for Carbon and Ecosystem Studies). Comparé aux 6 variables biogéochimiques de LOBSTER, le modèle PISCES comporte 24 variables dont 2 différents types de phytoplancton. Il permet ainsi une représentation améliorée de la production primaire et des flux de carbone associés. Compte tenu de ces précédents travaux, les objectifs de départ de cette thèse sont donc :

- Mettre en place une configuration réaliste au 1/4° du modèle couplé NEMO/PISCES adaptée à l'océan Atlantique Nord et évaluer sa pertinence pour la représentation de la production primaire.
- Faire évoluer cette configuration vers un système d'ensemble capable de présenter un niveau d'incertitude cohérent avec les écarts entre le modèle et les observations satellites.
- Examiner l'aptitude de ce système dans un contexte d'assimilation de données de la couleur de l'océan.

Pour réaliser ces objectifs, une part importante de cette thèse a été consacrée à une réflexion autour du concept d'incertitude. Quelles sont les principales sources d'incertitudes de la production primaire? Comment en tenir compte? Qu'est-ce que cela implique? Ces questions, initialement destinées à l'assimilation de données, nous ont amenés à reconsidérer

5. Modélisation des Ecoulements Océaniques à Moyenne et grande échelle

6. Le Groupe Mission Mercator Coriolis est un programme scientifique dont l'objectif est de faire le lien entre la recherche et l'océanographie opérationnelle

l'approche déterministe⁷ conventionnellement utilisée en modélisation biogéochimique.

Les travaux de Brankart (2013) sont le point de départ de l'approche probabiliste que nous utiliserons dans cette thèse. Ils montrent que l'introduction de processus aléatoires (stochastiques) dans les équations d'un modèle de circulation de l'océan permet de simuler certains effets dus aux échelles non résolues. Suite à ces résultats, l'idée est d'exploiter le potentiel de ces processus aléatoires (i.e d'une approche probabiliste) pour simuler explicitement certaines incertitudes biogéochimiques du modèle PISCES. Une simulation d'ensemble pourra ainsi être réalisée. Elle permettra de caractériser l'erreur de modélisation sous la forme d'une distribution de probabilité et de comparer statistiquement les écarts entre le modèle et les observations. Dans le contexte d'assimilation de données, l'intérêt de cette démarche probabiliste sera d'estimer l'erreur de prévision à partir de l'incertitude du modèle. La confrontation entre le modèle et les observations permettra de soulever un autre point important de cette étude : montrer que l'utilisation d'une approche probabiliste peut être une solution appropriée à la modélisation des processus biogéochimiques.

Organisation du document

Dans les 2 chapitres qui suivent cette introduction, nous présenterons l'ensemble des éléments à la base de ce travail de thèse. Les notions importantes, évoquées brièvement dans cette introduction, seront approfondies. Le chapitre 2 s'attardera sur des points de réflexion et de contexte, nécessaires pour définir notre démarche probabiliste dans l'approche de modélisation/assimilation. Le troisième chapitre présente les principaux outils que nous utiliserons : d'une part les observations satellites de la couleur de l'océan et d'autre part la configuration couplée au 1/4° NEMO/PISCES.

Dans le chapitre 4, nous présenterons les résultats d'une simulation déterministe du modèle PISCES réalisée entre janvier 2005 et décembre 2010. La pertinence de remplacer le modèle LOBSTER par PISCES pour représenter les distributions en chlorophylle de l'océan Atlantique Nord sera démontrée. Nous exposerons également quelques-unes des caractéristiques spatiales et temporelles de la variabilité en chlorophylle de cette simulation déterministe.

Le chapitre 5 marque le passage vers une approche probabiliste. Nous proposerons une méthode permettant de simuler certaines incertitudes biogéochimiques en introduisant des processus aléatoires, via des paramétrisations stochastiques, dans les équations du modèle PISCES. L'impact de ces paramétrisations stochastiques sera ensuite analysé à partir d'une simulation d'ensemble à 60 membres. Dans le chapitre 6, cette simulation d'ensemble sera évaluée de manière statistique en considérant les distributions de probabilité définies par ses différents membres. Nous étudierons en particulier la cohérence statistique (fiabilité) entre cette simulation et les observations SeaWiFS.

Dans le dernier chapitre, nous montrerons que la mise en place de cette configuration constitue une base prometteuse pour l'assimilation de données de la couleur de l'eau. Nous présenterons pour cela les résultats d'une séquence d'analyses de données satellites SeaWiFS, faite tous les 5 jours sur l'année 2005. Ces analyses seront réalisées à partir de la définition des statistiques d'erreur déduites de la simulation d'ensemble décrite dans les 2 chapitres précédents.

7. Le déterminisme est l'approche habituelle de résolution des équations qui aboutit à un résultat unique.

Finalement, les résultats présentés dans ce manuscrit, la démarche utilisée ainsi que les perspectives ouvertes par ces travaux dans le cadre modélisation/assimilation, seront récapitulés dans une conclusion générale.

Chapitre 2

Conception probabiliste des incertitudes et assimilation de données

Sommaire

2.1	Hasards mathématiques	17
2.2	Modéliser la biogéochimie marine	20
2.2.1	Omniprésence des incertitudes	20
2.2.2	Complexité des modèles	21
2.3	L'approche probabiliste	23
2.3.1	Simulation d'ensemble	24
2.3.2	Simulation stochastique des incertitudes	26
2.4	L'assimilation de données	27
2.4.1	Formalisme probabiliste	28
2.4.2	Étape d'analyse : l'interpolation statistique	29
2.4.3	D'un filtre de rang réduit à un filtre d'ensemble	30

Si la lecture de ce manuscrit devait laisser transparaître un message, je souhaiterais qu'il apparaisse comme un encouragement au développement d'une approche probabiliste en océanographie. J'aimerais parvenir à souligner que la conception, selon laquelle il n'existe plus de certitudes mais seulement des événements probables, doit véritablement faire partie intégrante de la pensée géo-scientifique. Ne pas intégrer la notion d'incertitude au sein d'un modèle de prévision pourrait même « ne pas être conforme à la démarche scientifique » (Tennekes, 1991), dans le sens où tout résultat restera incomplet sans une estimation de la quantité d'erreur que l'on peut lui associer. En définitive, il me paraît essentiel d'accepter qu'il puisse exister différentes manières d'appréhender les notions d'incertitude et de hasard. Et selon la signification qu'on leur prête, notre manière mathématique de raisonner, qui résulte encore largement des principes de causalités de notre éducation newtonienne, n'a pas forcément besoin d'être totalement remise en question. L'aléatoire déterministe de la théorie du chaos (dont nous parlerons dans la prochaine section), fait par exemple intervenir le hasard d'une manière bien différente de la mécanique quantique.

En se situant au voisinage entre philosophie et mathématiques, les subtiles nuances de l'incertain se retrouvent régulièrement contestées, rejetées par les critères de réalité de notre perception naturellement instinctive du monde. La raison humaine peine pour ne pas associer automatiquement au hasard des connotations métaphysiques, alors qu'il exprime parfois simplement « l'expression de notre ignorance », comme le définissait Pierre-Simon de Laplace. Le concept d'obstacle épistémologique cher à Gaston Bachelard (Bachelard, 1938) prendrait ici tout son sens. Les principales explications de notre résistance à l'approche probabiliste seraient ainsi à chercher dans son assimilation inconsciente avec l'imprévisible, l'absence de causes ou le surréalisme. « Un mécanisme qui prend l'apparence d'une intention » disait Henri Bergson, un état de désordre qui échappe à notre « instinct naturel d'induction » préférait le philosophe David Hume, parce qu'il ne se peut pas que notre dieu, comme celui d'Einstein, puisse « jouer avec les dés ». Niels Bohr se permit tout de même de répondre : « Ne dites pas à dieu ce qu'il doit faire ! ». Mais alors que penser ? Cette vision du désordre de l'aléatoire est-elle obligatoirement incompatible avec un ordre physique et mathématique de la nature ? Puisque, après tout, on pense à la fois que le hasard fait bien les choses, et que la nature est bien faite, peut-être pas forcément...

Quoi qu'il en soit, mon intention n'est certainement pas d'établir ici un traité philosophique sur l'intelligibilité scientifique du hasard et de l'incertitude. Le point que je souhaite soulever dès le début de cette thèse, parce que moi-même contraint de le réaliser, est que le concept de démarche probabiliste n'est pas forcément intuitif, n'a pas un sens conventionnellement établi. Le principal obstacle à son acceptation provient alors sûrement du fait qu'il nous oblige à réadapter certaines de nos manières habituelles de concevoir les choses. Dans un domaine comme l'océanographie, où la quasi-totalité des modèles est entièrement déterministe, c'est-à-dire que la résolution des équations se fait sans l'intervention de nombres aléatoires, il paraît d'autant plus important de définir clairement la nature de notre démarche probabiliste.

Ce chapitre a été conçu dans ce sens et doit donc être perçu comme un prolongement à l'introduction. L'objectif est uniquement de présenter les éléments à la base de l'approche probabiliste que nous utiliserons dans la suite de ce manuscrit. Une description détaillée des outils, c'est-à-dire du modèle couplé NEMO/PISCES et des observations de la couleur de l'océan, sera l'objet du chapitre suivant.

Dans la première section de ce chapitre, nous dresserons un aperçu chronologique de l'évolution de la perception scientifique du hasard. L'intérêt sera de préciser certaines nuances qui entourent la notion d'approche probabiliste afin de mieux délimiter le contexte de notre démarche. Dans la deuxième section, nous discuterons des principales sources d'incertitude dans le cas particulier de la modélisation biogéochimique. Pour rendre compte d'une partie de ces incertitudes, une simulation d'ensemble sera réalisée et analysée. Nous verrons dans la troisième section de ce chapitre ce que représente la modélisation d'ensemble. Cela nous amènera à définir l'approche de modélisation stochastique que nous utiliserons plus particulièrement. Dans la dernière section de ce chapitre, les bases de l'assimilation de données seront présentées dans un cadre probabiliste (problème de Bayes). Nous ferons notamment le rapprochement entre la conception probabiliste des incertitudes et l'assimilation de données. Je rappelle enfin que l'objectif de cette thèse reste, avant tout, de démontrer la pertinence d'une caractérisation probabiliste des incertitudes dans une approche intégrée modèle/observation.

Un rapide retour en arrière à travers l'histoire de la théorie des probabilités s'impose d'abord à nous...

2.1 Hasards mathématiques

Bien que présente dès l'Antiquité dans certains jeux et divertissements, l'histoire de la théorie des probabilités ne débute véritablement qu'à partir du XVII^{ème} siècle avec la correspondance entre Blaise Pascal et Pierre Fermat au sujet du «problème des partis». Historiquement attribué au mathématicien italien de la Renaissance Luca Paciolo, l'énoncé de ce problème issu de conflits dans des ruptures de contrats entre marchands est le suivant : 2 joueurs s'affrontent dans une partie de pile ou face. Celui qui remporte la mise est celui qui totalise N prédictions correctes. La question est de savoir comment répartir les gains équitablement si la partie s'arrête avant que l'un des joueurs ait atteint les N prévisions correctes. Sa résolution en 1654 par Pascal qui suggère simplement de continuer virtuellement le jeu, lui permettra d'établir le Traité du triangle arithmétique dans lequel apparaît, entre autres, une représentation géométrique des coefficients binomiaux : le célèbre triangle de Pascal. Huygens élaborera en 1657 à partir de ces travaux le premier véritable ouvrage de théorie des probabilités avec *De ratiociniis in ludo aleae*. La notion d'espérance mathématique était dès lors introduite.

Il faudra ensuite attendre 1712 et la publication de *Ars Conjectandi* pour que Jacques Bernoulli propose une avancée majeure à la théorie des probabilités avec l'énoncé de la loi des grands nombres. Si l'échantillon d'un tirage aléatoire est suffisamment grand, on peut maintenant s'appuyer sur les caractéristiques statistiques, réaliser des sondages et évaluer des risques (premières apparitions des assurances). La moyenne empirique devient un estimateur convergent de l'espérance, amorçant l'émergence de la future loi normale. Dans la foulée, le mathématicien français Abraham De Moivre exprime en 1718 dans *Théorie du hasard* son intérêt pour les distributions binomiales dans l'étude des probabilités. Ses recherches sur les problèmes de passage à la limite feront ensuite l'objet en 1730 de *Miscellanea analytica*, œuvre dans laquelle apparaît pour la première fois l'évaluation de $n!$. Dès lors, tous les ingrédients étaient réunis pour que Pierre Simon de Laplace puisse démontrer

en 1812 dans *Théorie analytique des probabilités* l'existence de l'un des théorèmes les plus importants de la théorie des probabilités, le théorème de la limite centrale¹ : *toute suite de variables aléatoires indépendantes définies sur le même espace de probabilité et suivant une même loi d'espérance μ et d'écart-type σ converge vers une loi normale $\mathcal{N}(\mu, \sigma^2)$* . La fameuse courbe en cloche, ou courbe de Gauss² était née.

Mais si l'essentiel des bases de la théorie des probabilités était posé, cette «mathématisation du hasard» n'apparaissait toujours pas dans la physique. Le hasard ne restait encore qu'une illusion mathématique, une théorie de l'erreur. Le raisonnement probabiliste ne commencera à véritablement intégrer la physique qu'un peu plus tard, dans les travaux de Clausius sur le désordre moléculaire (l'entropie) puis en 1866 avec l'énoncé de la loi de distribution statistique des vitesses de la théorie cinétique des gaz de Maxwell. Il ne restait plus qu'à Ludwig Eduard Boltzmann de lui ajouter une équation d'évolution pour qu'un grand nombre de lois expérimentales de la thermodynamique puissent, enfin, être démontrées à partir de théories probabilistes. La physique venait de subir une avancée fondamentale, elle devenait une physique statistique. Pourtant, rien ne contredisait véritablement le déterminisme et la vision Laplacienne du hasard : les probabilités ne sont que le reflet de notre ignorance. À chaque instant, l'état du gaz reste complètement déterminé par une valeur fixe de la vitesse et de la quantité de mouvement de chaque particule. L'approche statistique ne reflète finalement ici que notre incapacité à avoir accès à l'ensemble de ces informations lorsque le nombre de particules est très grand. Malgré cela, déduire des lois macroscopiques à partir de comportements microscopiques témoigne d'une véritable prise de conscience de la complexité des systèmes physiques. Une nouvelle conception du hasard pouvait voir le jour, résultat des travaux de Henri Poincaré à la fin du XIX^{ème} sur les problèmes à 3 corps publiés dans *Méthodes nouvelles de la Mécanique Céleste*, celle de l'aléatoire déterminisme : *une cause très petite, qui nous échappe, détermine un effet considérable que nous ne pouvons pas voir....Il peut arriver que de petites différences dans les conditions initiales en engendrent de très grandes dans les phénomènes finaux, la prédiction devient impossible et nous avons le phénomène fortuit*. Henri Poincaré, *Théorie des probabilités*, 1905. Le terme d'ignorance de Laplace est remplacé par une insuffisante précision des données, par l'impossibilité d'accéder à cette précision. La nuance peut sembler étroite mais elle est fondamentale car elle permet de considérer conjointement hasard et déterminisme. Comme je ne pourrai jamais avoir accès à une précision infinie, certains systèmes complexes, même s'ils restent déterministes, ne pourront être résolus que de manière probabiliste. Des effets connus ne pourront pas être prédits, ils nous apparaîtront alors « par hasard ». Cette approche aboutira en 1961 à la théorie du chaos attribuée au météorologue Edward Lorenz. Cette conception des incertitudes, issue d'une hypersensibilité des conditions initiales, sera même un succès auprès du grand public avec l'illustration bien connue de l'effet papillon, dont le battement d'aile au Brésil pourrait provoquer un ouragan le long des côtes américaines.

Entre temps, un autre genre de conception avait également vu le jour puisqu'aux alentours de 1843, l'économiste Antoine Cournot avait déjà introduit, à partir des idées de Spinoza, une nouvelle définition du hasard scientifique. Selon lui, le hasard apparaît lorsque

1. La désignation de théorème de la limite centrale est donnée en 1920 par le mathématicien Georges Polya.

2. Le nom courbe de Gauss provient des travaux de Carl Friedrich Gauss de 1809 publiés dans *Theoria motus corporum coelestium*. Il retrouve la loi normale en calculant les mouvements des corps célestes.

deux séries causales indépendantes se télescopent. C'est la nouvelle causalité qui en résulte qui crée le hasard. L'image régulièrement citée est celle de l'homme qui répare une tuile sur un toit. Un autre homme passe en dessous et la tuile lui tombe sur la tête. Sa thèse est donc qu'il existe plusieurs ordres de causalité, plusieurs séries temporelles qui ne devraient pas se mélanger et que, de temps en temps, par hasard, certaines de ces séries se rejoignent et il se passe quelque chose d'inattendu. Critiquée par Poincaré, qui considère que l'existence des différentes séries temporelles provient d'un découpage humain de l'univers, cette thèse eut au moins le mérite de faire apparaître pour la première fois le hasard comme faisant partie de l'ordre des choses. Et c'est précisément cette conception du hasard dans la nature qui va connaître un essor fulgurant avec ce qui reste pour moi l'une des œuvres scientifiques majeures du XIX^{ème} siècle, *De l'origine des espèces*, publiée en 1859 par le biologiste Charles Darwin. Non seulement toutes les espèces vivantes descendent d'un même ancêtre mais l'évolution se fait sans direction prédéterminée. Ses mécanismes sont dirigés au hasard de mutations génétiques et l'évolution est renommée sélection naturelle, conférant ainsi une connotation naturelle à l'aléatoire.

Sans savoir réellement si un lien pouvait exister entre les deux, débute peu de temps après une ère nouvelle pour la physique, celle de la remise en cause du déterminisme : l'ère du bouleversement quantique. Ses initiateurs en sont les physiciens Max Planck et Albert Einstein, entre 1900 et 1905, à partir de l'hypothèse selon laquelle, pour expliquer l'effet photoélectrique, les échanges d'énergies entre la matière se font de manière discontinue, par quanta. Le vieux débat du XVII^{ème} siècle qui avait vu s'opposer Newton et Huygens à propos de la dualité onde-corpuscule de la lumière allait être remis sur la table, modifiant cette fois-ci radicalement la conception classique de la physique. Il faudra désormais concevoir comme possible que 2 événements puissent être corrélés sans que l'un soit la cause de l'autre (intrication quantique) et qu'un atome soit simultanément dans plusieurs états à la fois (principe de superposition) puisque c'est la mesure, l'observation, qui définit l'existence même de la notion d'état. L'incertitude et le hasard étaient ainsi mis au cœur de la théorie scientifique, appuyés par le principe d'indétermination de Heisenberg qui affirme en 1927 l'impossibilité (et non l'incapacité comme en physique statistique) d'avoir accès en simultané à la vitesse et à la position. La physique entrait dans un nouveau paradigme, le paradigme quantique. L'approche probabiliste n'était plus une expression de notre ignorance ou de notre incapacité à disposer d'informations assez précises mais celle d'un comportement réel dans lequel l'aléatoire et le hasard existent bien et font partie de l'ordre des choses, de l'ordre de la nature.

J'espère que ce détour ne nous éloigne pas trop de la thématique générale de ce manuscrit. Bien évidemment, nous n'utiliserons dans cette thèse aucune considération scientifique qui pourrait sortir du cadre de la physique classique. La réalité existe, elle est unique et nous pouvons l'observer. Ce point n'est jamais remis en question. Toutefois, par "nature", c'est-à-dire parmi le champ des possibles, une autre réalité aurait pu exister. Un système est par nature imprévisible car nous sommes incapables de discerner parfaitement l'ensemble des causes à l'origine de l'état observé. C'est l'approche qui sera utilisée ici. La démarche probabiliste ne remettra donc pas entièrement en cause le déterminisme des lois d'évolution de la dynamique biogéochimique de l'océan. Malgré l'introduction de processus aléatoires, elles resteront toujours caractérisées par des équations aux dérivées partielles.

2.2 Modéliser la biogéochimie marine

Si nous considérons que la réalité (observée) n'est qu'une réalisation possible, l'avantage de l'approche probabiliste est de pouvoir attribuer une probabilité d'existence aux différents événements possibles. Toutefois, pour que nous puissions comparer les distributions de probabilité, caractérisant l'erreur de prévision du modèle, avec les observations de la couleur de l'eau (SeaWIFS), elles doivent être le reflet de sources d'incertitude clairement identifiées. Les probabilités seront donc uniquement le moyen de rendre compte d'une partie des incertitudes sur la représentation des caractéristiques biogéochimiques par un modèle. Nous proposerons ensuite de simuler ces incertitudes par l'introduction de processus aléatoires dans les équations du modèle PISCES (cf. chapitre 5).

Dans la première partie de cette section, nous présenterons les principales sources d'incertitude biogéochimiques dont découle, en définitive, l'incapacité des modèles à simuler un état en accord avec les observations. Discerner ces différentes incertitudes permettra d'identifier celles qui nous intéresseront plus particulièrement dans la suite de cette thèse. Comme le système (la biogéochimie marine) est complexe et chaotique, sa modélisation déterministe sera toujours entachée d'erreur (aléatoire déterministe). Sous cet angle, la notion de complexité se rapporte à l'existence de multiples processus en interaction. Un modèle à vocation réaliste sera donc défini par son degré de complexité, c'est-à-dire la part de processus explicitement résolus, et son niveau d'incertitude, relative aux caractéristiques que nous souhaitons représenter. Dans la seconde partie de cette section, nous développerons la notion de complexité biogéochimique des modèles en faisant le lien avec celle d'incertitude.

2.2.1 Omniprésence des incertitudes

Lorsque l'on s'attelle à représenter le vivant, il n'existe pas, malgré la présence de processus bien connus, d'équivalent théorique comparable au principe fondamental de la dynamique ou aux équations de Navier-Stokes. La modélisation biogéochimique repose sur des formulations empiriques, contraintes par une multitude de paramètres déterminées la plupart du temps à partir d'expériences en laboratoire. Cette absence de « grands principes généraux » est une difficulté majeure à la modélisation des caractéristiques biogéochimiques de l'océan (Oschlies and Schertau, 2005). Si on veut, par exemple, simuler la quantité de nitrate consommée par une population d'organismes photosynthétiques, on utilise une équation, un modèle, caractérisant l'absorption de ce nutriment à partir de conditions initiales (du système nitrate + phytoplancton) et de données extérieures comme l'intensité lumineuse ou la température. Si cette équation est exacte, alors le degré de précision de la prévision dépend uniquement de celui des conditions initiales et des données extérieures. Notre connaissance de ces 2 données étant forcément non absolues, la représentation du phénomène sera inévitablement incertaine.

Les modèles ayant généralement vocation à être utilisés dans des cadres plus réalistes, les dépendances aux propriétés physico-chimiques de l'environnement extérieur sont, la plupart du temps, établies de manière approximative. Le comportement d'un système biologique au sein d'un milieu différent du laboratoire ne pourra jamais être parfaitement représenté. Certains processus comme la diversité des espèces ou l'adaptation biologique seront tout simplement absents des formulations du modèle. Pour compenser ces incertitudes, une multitude de paramètres sont alors introduits dans les équations pour contraindre les

comportements biogéochimiques de la manière la plus proche possible de la réalité. En définitive, les lois mathématiques ne peuvent retranscrire qu’une vision approximative de la réalité (parfois suffisante), n’englobant jamais la multiplicité des phénomènes ayant véritablement de l’influence sur l’absorption du nitrate par le phytoplancton, pour reprendre notre exemple précédent. Notre manque de connaissance des fonctionnements biogéochimiques se manifeste par des formulations imparfaites et incomplètes, rendant les prévisions qui en découlent nécessairement incertaines. Une approche probabiliste sera proposée dans cette thèse pour simuler explicitement ce type d’incertitudes « structurelles » car inhérentes aux formulations biogéochimiques des modèles. Quelques paramètres du modèle PISCES, clés pour l’estimation de la production primaire (contexte assimilation de données de la couleur de l’eau), seront pour cela considérés comme incertains. Nous présenterons ces paramètres et la démarche utilisée pour simuler leur incertitude dans le chapitre 5.

Les organismes photosynthétiques ne possédant que très peu de mouvements propres, leur répartition spatiale, tout comme l’accessibilité aux nutriments sont de plus conditionnées par la dynamique chaotique et turbulente de l’océan (Oschlies and Garçon, 1998a) dont la modélisation est elle aussi évidemment incertaine. L’illustration la plus convaincante est certainement celle de la représentation des petites échelles. Dans un courant comme le Gulf Stream (cf. section 4.1), l’étirement des tourbillons de méso-échelle crée des structures instables turbulentes beaucoup plus fines, dites de subméso-échelle. Ces petites structures génèrent un fort mélange turbulent (Dubos and Babiano, 2002) qui modifie la distribution des vitesses verticales à l’origine de l’advection des nutriments vers la subsurface (Lévy et al., 2011; Klein and Lapeyre, 2009). Comme conséquence de cette dynamique, on peut observer par image satellitaire des « filaments » de chlorophylle de l’ordre de la dizaine de kilomètres (figure 2.1), au sein desquels l’activité biogéochimique est particulièrement intense (Legal et al., 2006). Mais malgré l’amélioration croissante des performances informatiques, la résolution des modèles ne dépasse que très rarement la dizaine de kilomètres. Ce type de régimes dynamiques sous-maille ne peut donc pas être résolu explicitement et représente alors une importante source d’incertitude sur l’estimation des caractéristiques dynamiques et a fortiori biogéochimiques de l’océan.

De manière générale, les échelles non résolues sont une source d’erreur permanente en modélisation, quel que soit le système considéré. La discrétisation numérique d’équations non linéaires ne permet, en effet, pas de rendre compte précisément de l’ensemble des fluctuations spatio-temporelles réellement existantes, induisant des biais dans le calcul des variables d’état physiques (exemple précédent), mais aussi biogéochimiques. Dans cette thèse, nous présenterons une paramétrisation (stochastique) permettant de simuler les incertitudes sur l’estimation des variables biogéochimiques dues aux échelles non résolues. Cette paramétrisation, conjuguée avec celle simulant l’incertitude des paramètres biogéochimiques constituera la base qui nous permettra de mettre en place une version probabiliste du modèle PISCES.

2.2.2 Complexité des modèles

Compte tenu de l’omniprésence des incertitudes, caractériser le niveau de complexité d’un modèle biogéochimique nécessite de pouvoir définir clairement la séparation entre les processus résolus, partiellement résolus et non résolus. Dans ce sens, incertitude et com-

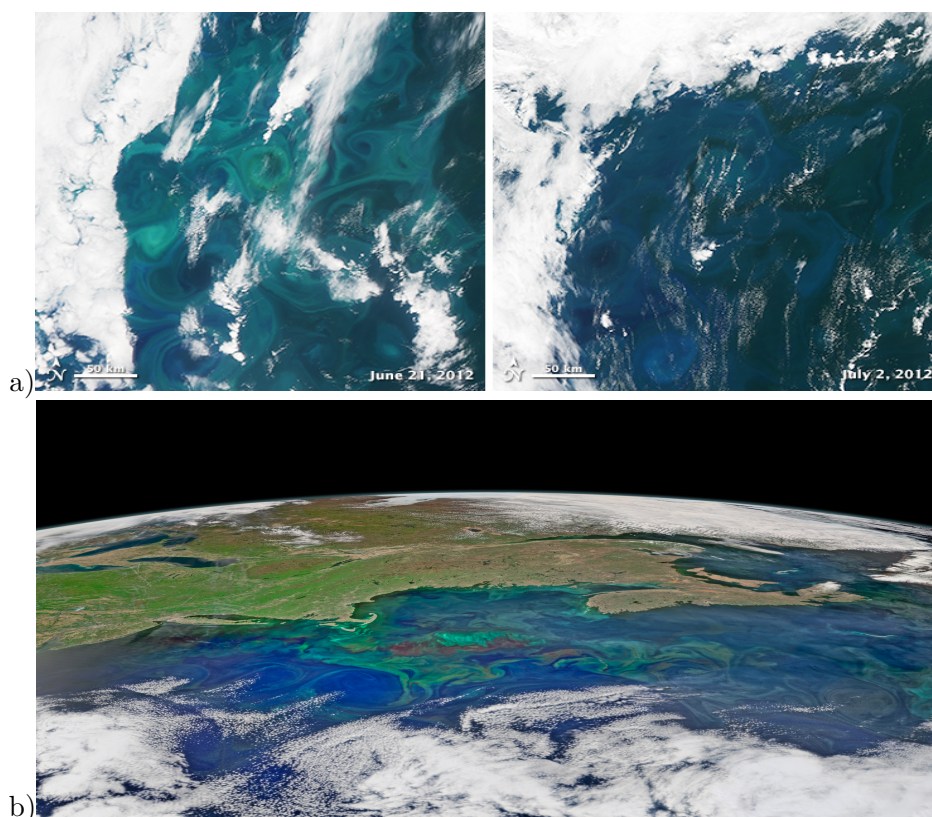


FIGURE 2.1 – a) Images satellites présentant des structures en filaments de la chlorophylle de surface (source <http://earthobservatory.nasa.gov>). a) Bloom de coccolithophoridés observé au large du Groënland avec le capteur MODIS. b) Bloom de phytoplancton dans le Gulf Stream observé le 14 Mai 2015 au large de la Nouvelle Ecosse avec le capteur VIIRS.

plexité sont étroitement liées. Plus un système est complexe, plus le nombre de processus non résolus et partiellement résolus sera faible. Pour améliorer les performances des modèles biogéochimiques, une technique courante consiste à réduire les incertitudes en élargissant le nombre de variables, c'est-à-dire en augmentant le niveau de complexité. Si cette approche peut s'avérer fructueuse, déterminer le niveau optimal de complexité à inclure dans les modèles est régulièrement sujet à débats. Dans son article, Anderson (2005) assimile l'utilisation de modèles toujours plus complexes à essayer de courir avant de savoir marcher : « running before we can walk ». Le principe du rasoir d'Ockham (les multiples ne doivent pas être utilisés sans nécessité) est une expression couramment entendue dans les conférences de modélisation biogéochimique. Les intercomparaisons entre modèles de différentes complexités peinent, en effet, à prouver qu'ajouter des variables (donc de la complexité) permet véritablement d'améliorer les prévisions (Kwiatkowski et al., 2014). En optimisant les paramètres (en général par assimilation de données), des modèles "simples" sont souvent capables de donner des résultats comparables à ceux obtenus avec des modèles plus compliqués (Raick et al., 2006; Ward et al., 2013). Si la portabilité de ces modèles reste généralement limitée (Friedrichs et al., 2007), travailler avec un plus faible niveau de complexité est donc certainement préférable. L'identification des processus est plus aisée et les

coûts de calcul numérique peuvent être considérablement réduits. La course effrénée à la complexité n’est donc peut-être pas la meilleure solution pour améliorer les prévisions des modèles. La formulation des processus biogéochimiques restant toujours approximative, il est donc surtout nécessaire d’utiliser un modèle possédant : 1) un niveau de complexité en accord avec les besoins de l’étude, pour lequel 2) une caractérisation des principales incertitudes est possible (Arhonditsis et al., 2008).

Compte tenu du contexte d’assimilation de données de la couleur de l’océan de ce travail, la première étape sera de vérifier que la plate-forme de modélisation utilisée, le modèle couplé NEMO/PISCES, permet de représenter l’essentiel des caractéristiques de grande échelle présentes dans ces observations. En particulier, on s’assurera qu’il fournit une meilleure estimation de la chlorophylle de surface que le modèle LOBSTER, utilisé dans les études qui ont précédé ce travail. Nous proposerons ensuite d’améliorer la représentation de la chlorophylle en prenant explicitement en compte certaines sources d’incertitude biogéochimiques du modèle (cf. paragraphe précédent). L’approche probabiliste que nous utiliserons se traduira, en pratique, par une démarche de modélisation d’ensemble que l’on introduit dans la section qui suit.

2.3 L’approche probabiliste

Si les probabilités sont le langage mathématique de l’incertain, les simulations d’ensemble en sont l’expression numérique. Telle qu’effectuée aujourd’hui, l’introduction de probabilité dans les modèles de climat provient principalement des travaux de Epstein (1969) et Leith (1974). Depuis le début des années 1990, la prévision opérationnelle du Centre Européen pour les Prévisions Météorologiques à Moyen Terme (CEPMMT/ECMWF, Palmer et al., 1993; Molteni et al., 1996) et du Centre National de Prévision Environnementale (NCEP, Toth and Kalnay, 1993) est probabiliste, c’est-à-dire faite à partir de simulations d’ensemble.

En océanographie, les simulations d’ensemble sont le plus souvent associées à l’assimilation de données (cf. section 3.1.3). Dans le cadre de la modélisation biogéochimique, nous pouvons toutefois mentionner les travaux de Woods and Onken (1982) et Wolf and Woods (1988), dans lesquels la population de phytoplancton d’un modèle 1D est représentée par un ensemble de particules (lagrangiennes) aux mouvements verticaux différents. Après avoir été presque oubliée car jugée numériquement trop coûteuse, cette approche réapparaît aujourd’hui pour simuler la diversité biologique dans des modèles de compétition d’espèces (Follows et al., 2007; Follows and Dutkiewicz, 2011). D’autres études se sont penchées sur l’utilisation d’approches stochastiques pour la modélisation du climat. Nous en évoquerons quelques-unes dans le chapitre 5.

Dans ce manuscrit, nous présenterons une simulation d’ensemble réalisée en introduisant des processus stochastiques dans les équations du modèle PISCES. Si l’objectif sera de caractériser les erreurs de prévision par l’intermédiaire de distributions de probabilité (contexte assimilation de données), nous développerons également le bénéfice de cette approche probabiliste/stochastique pour la modélisation biogéochimique. Nous montrerons en particulier qu’elle permet une comparaison objective aux observations de la couleur de l’eau et de représenter des comportements biogéochimiques non explicitement simulés par le modèle (cf. chapitre 6). Dans cette section, nous développerons les principes de base de la simulation d’ensemble et de la modélisation stochastique.

2.3.1 Simulation d'ensemble

Démarche

Conceptuellement, le principe de la modélisation d'ensemble est relativement simple. Au lieu de travailler avec une simulation unique, on lance un ensemble de simulations avec des données d'entrée légèrement différentes. L'échantillonnage des réalisations fait par les différents membres de l'ensemble définit une distribution de probabilité (PDF) qui rend compte de l'incertitude de la prévision. Comparé à l'approche déterministe, l'intérêt est donc d'associer aux prévisions un niveau d'incertitude tenant compte de la dynamique du modèle et de le représenter sous la forme d'une distribution de probabilité. En définitive, une simulation d'ensemble permet de définir, à l'instant t_k , la distribution de probabilité a priori $P(\mathbf{x}_{t_k})$ du vecteur d'état \mathbf{x}_{t_k} .

La figure 2.2 illustre schématiquement le principe de la modélisation d'ensemble. À partir de conditions initiales légèrement différentes (le niveau d'incertitude est représenté par un cercle bleu), plusieurs simulations sont réalisées (traits bleu clair). La réponse à cette incertitude est définie par un espace des réalisations (espace bleu ciel) dans lequel la densité des membres caractérise une distribution de probabilité représentative de l'erreur faite sur la prévision. La simulation d'ensemble propage ainsi l'incertitude a priori sur l'état initial et évalue son impact sur l'incertitude des prévisions.

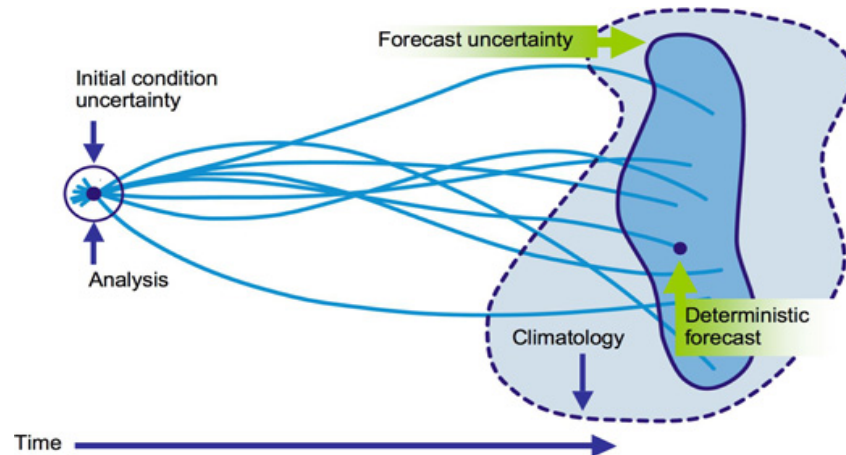


FIGURE 2.2 – Principe d'une simulation d'ensemble tel que présenté dans le blog de Steve Easterbrook "Tracking down the uncertainties in weather and climate prediction". Source : <http://www.easterbrook.ca/steve/?s=stochastic>

De la manière dont nous avons présenté jusqu'ici la modélisation d'ensemble, c'est uniquement le caractère chaotique du système (sensibilité aux conditions initiales) qui est évalué. N'importe quelle source d'incertitude peut néanmoins être simulée avec une approche d'ensemble à condition d'utiliser une méthode appropriée pour générer les différents membres. Dans notre cas, tous les membres de l'ensemble posséderont la même condition initiale. Par contre, nous introduirons des processus aléatoires (stochastiques) pour simuler explicitement l'incertitude associée au modèle. Les membres se différencieront donc les uns des autres par leur séquence de nombres aléatoires.

Méthode de Monte-carlo

Pour échantillonner de manière discrète les distributions de probabilité et produire les N membres d’une simulation d’ensemble, on utilise généralement des méthodes de type Monte-Carlo³. Le terme générique de Monte-Carlo désigne l’ensemble des méthodes d’approximation numérique dont le principe est la répétition d’expériences. Plus le nombre d’expériences est grand, meilleur sera l’échantillonnage. Ces méthodes reposent le plus souvent⁴ sur l’utilisation de nombres aléatoires (ou pseudo-aléatoires) pour échantillonner l’espace des états. Pour calculer numériquement une quantité Q par méthode de Monte-Carlo, cette quantité doit être écrite sous la forme d’une espérance telle que $Q = \mathcal{E}[f(x)]$, avec f une fonction quelconque. Si X_i est une suite de variables aléatoires tirées au hasard dans cet intervalle, l’hypothèse de la loi des grands nombres permet d’avoir :

$$Q_N = \frac{1}{N} \sum_{i=1}^N f(X_i) \xrightarrow[N \text{ grand}]{} \mathcal{E}[f(x)]. \quad (2.1)$$

Selon le théorème central limite, la convergence de Q_N vers Q est en \sqrt{N} et dépend de la variance de la fonction f . Cela explique que le coût numérique d’une simulation d’ensemble puisse rapidement devenir conséquent.

Lorsque les méthodes de Monte-Carlo sont utilisées pour rendre compte de l’incertitude d’un modèle, les échantillons aléatoires sont tirés selon une loi de probabilité, souvent gaussienne, définie a priori. Ces nombres aléatoires sont alors des perturbations que l’on applique à un modèle. Pour tenir compte de l’incertitude des conditions initiales d’un modèle (comme sur la figure 2.2), l’ensemble est ainsi initialisé par $x(t=0)$ modulo une petite quantité aléatoire qui diffère d’un membre à l’autre. La dynamique du modèle propage ensuite cette incertitude pour obtenir une incertitude sur la prévision. Si c’est souvent l’état initial qui est supposé incertain (caractère chaotique des systèmes), l’incertitude des paramètres du modèle peut également être considérée (Murphy et al., 2004). Chaque membre de l’ensemble possède alors une condition initiale identique mais les paramètres qui régissent la dynamique du modèle sont légèrement différents.

Cette démarche est proche de celle utilisée dans cette thèse puisque nous considérerons l’incertitude de certains paramètres biogéochimiques du modèle PISCES (cf. section 2.2). Toutefois, au lieu de restreindre l’utilisation des nombres aléatoires à l’initialisation du modèle, ils seront introduits dans les équations à chaque pas du temps du modèle par l’intermédiaire de processus stochastiques.

3. Le nom de Monte-Carlo a été suggéré en 1947 par Nicholas Metropolis en référence aux jeux de hasard des casinos du quartier de la principauté de Monaco. Ces méthodes sont célèbres pour avoir été un point central des simulations du projet Manhattan pendant la seconde guerre mondiale. Leur développement est généralement attribué au mathématicien John von Neumann. Ces méthodes sont aujourd’hui appliquées dans des domaines aussi variés que la finance, les sciences sociales, la sismologie, l’ingénierie.

4. En pratique, on peut choisir un tirage déterministe. L’espace des états est alors une suite de nombres préalablement définie. L’intérêt est de contrôler l’homogénéité de la répartition des valeurs. On parle de méthode de quasi Monte-Carlo.

2.3.2 Simulation stochastique des incertitudes

L'utilisation de processus stochastiques dans les équations des modèles de climat est généralement attribuée à Hasselmann (Hasselmann, 1976). Si cette approche reste encore très peu développée en dehors de la météorologie, elle apparaît si prometteuse que plusieurs études plaident en faveur de systèmes de prévision entièrement stochastiques⁵ (p.ex. Palmer, 2012). De manière simplifiée, un processus stochastique \mathbf{w}_{t_k} décrit l'évolution temporelle d'une variable aléatoire. C'est donc une suite, indexée sur le temps, de variables aléatoires définies sur un même espace de probabilité. L'exemple le plus connu de processus stochastique est certainement celui du processus de Wiener, à la base de la modélisation du mouvement brownien⁶.

Pour expliciter la démarche de modélisation stochastique, plaçons nous tout d'abord dans un contexte de prévision déterministe. Le principe est d'évaluer l'évolution d'un système décrit par un vecteur d'état $\mathbf{x}^t(t)$ (état vrai) à partir d'un modèle \mathcal{M} qui en fait une représentation \mathbf{x}_{t_k} à l'instant t_k (équation 2.2).

$$\mathbf{x}_{t_k} = \mathcal{M}_{t_{k-1}, t_k}(\mathbf{x}_{t_{k-1}}) \quad ; \quad \mathbf{x}(0) = \mathbf{x}_{t_0} \quad (2.2)$$

En raison des imperfections du modèle, $\mathbf{x}_{t_k} \neq \mathbf{x}^t(t_k)$. L'approche de modélisation stochastique, telle que celle que nous utiliserons, consiste à introduire des processus stochastiques dans les équations d'un modèle de prévision (équation 2.3) pour rendre compte de l'incertitude du modèle. Avec cette formulation, l'évolution de l'état du système va dépendre du processus aléatoire \mathbf{w}_{t_k} . Le vecteur d'état du modèle devient donc un vecteur aléatoire.

$$\mathbf{x}_{t_k} = \mathcal{M}_{t_{k-1}, t_k}(\mathbf{x}_{t_{k-1}}, \mathbf{w}_{t_k}) \quad ; \quad \mathbf{x}(0) = \mathbf{x}_{t_0} \quad (2.3)$$

Afin de réaliser une simulation d'ensemble, il suffit d'introduire dans le modèle un ensemble de processus stochastiques. Les membres d'une simulation d'ensemble diffèrent donc les uns des autres par la séquence des nombres aléatoires (possédant des caractéristiques statistiques identiques). Une description approfondie des processus stochastiques utilisés dans cette thèse pour produire une simulation d'ensemble de 60 membres sera l'objet du chapitre 5. À ce stade, nous pouvons retenir qu'une simulation d'ensemble (échantillonnage de type Monte-Carlo) sera générée par des processus aléatoires introduits dans les équations du modèle PISCES. L'objectif de cette approche d'ensemble (stochastique) est d'obtenir une distribution de probabilité qui caractérise l'erreur de prévision en tenant compte de l'incertitude du modèle.

Ce qui précède explique la différence principale entre l'approche d'ensemble habituelle et l'approche d'ensemble stochastique telle qu'adoptée dans cette étude. Dans le premier cas, les nombres aléatoires dessinent une distribution de probabilité qui caractérise une incertitude a priori sur les conditions initiales, les forçages ou les paramètres du modèle.

5. C'est-à-dire dans lesquels des processus aléatoires interviendraient dans toutes les équations du modèle

6. Le mouvement brownien est un mouvement aléatoire de particules découvert en 1827 par le botaniste Charles Brown en observant les mouvements des grains de pollen. Il sera expliqué physiquement par Einstein et modélisé à partir de processus stochastiques par Norbert Wiener en 1923.

Chacune des trajectoires des membres de l’ensemble reste ensuite foncièrement déterministe. Les nombres aléatoires n’interviennent que de manière « extérieure » au modèle. Avec une approche stochastique, les nombres aléatoires sont introduits à chaque pas de temps du modèle. Les trajectoires dynamiques sont ainsi dictées par l’ajout successif de nombres aléatoires dans les équations du modèle. La formulation du modèle lui-même n’est donc plus (entièrement) déterministe, le modèle devient (partiellement) probabiliste.

Considérer de cette manière l’incertitude des formulations du modèle nous a semblé particulièrement approprié à la problématique de caractérisation des incertitudes des processus biogéochimiques. Cela permet, non pas uniquement d’évaluer l’impact de l’incertitude d’une donnée d’entrée (p.ex. paramètre) sur la représentation du modèle, mais de réaliser une simulation dans laquelle l’incertitude sur cette donnée est explicitement considérée lors de la modélisation. En plus de pouvoir améliorer l’estimation des statistiques d’erreur du modèle, proposer ce type d’approche probabiliste présente une alternative à la modélisation déterministe qui n’existe pas, à notre connaissance, dans une configuration biogéochimique réaliste de l’océan. Ce sera un point important de cette thèse.

2.4 L’assimilation de données

L’assimilation de données peut être définie comme l’ensemble des méthodes mathématiques permettant de combiner au mieux l’ensemble des connaissances que l’on possède sur un système dynamique. Elle se présente classiquement comme la recherche du meilleur estimé possible de l’état du système à partir des observations (mesures), d’un modèle d’équations et des incertitudes sur ces 2 sources d’information. Elle peut, en ce sens, être qualifiée de « science des compromis fructueux », pour reprendre une expression employée par Jacques Verron⁷. Si l’assimilation de données est souvent associée à la prévision météorologique, ses champs d’application couvrent des disciplines aussi variées que l’astronomie, la médecine, la robotique ou encore la surveillance des centrales nucléaires. Comme nous l’évoquions en introduction, l’émergence de l’assimilation de données en océanographie s’explique en grande partie par le développement, relativement récent, des réseaux d’observation, en particulier l’observation spatiale. Elle prend aujourd’hui place au coeur de tous les systèmes opérationnels tel que le système de prévision océanique de MERCATOR-Océan.

L’utilisation historique et principale de l’assimilation de données est d’améliorer la représentation des conditions initiales. Les systèmes étudiés étant fortement chaotiques, la correction de l’état estimé par le modèle permet d’établir un nouvel état initial, plus réaliste, améliorant les prévisions futures. L’incertitude des prévisions peut également être réduite par l’assimilation de données en déterminant des valeurs ”optimales“ des paramètres d’un modèle. On parle de problème inverse⁸ (par opposition au problème direct) dans le sens où ce sont les causes qui sont recherchées à partir des effets. Une autre application courante en océanographie est la production de réanalyses, c’est-à-dire la reconstruction d’états du passé sur de longues périodes. Cela permet entre autres d’améliorer la compréhension des processus océaniques, d’étudier la variabilité de l’océan et de construire des climatologies. Enfin, l’assimilation de données peut également servir à l’évaluation des systèmes d’obser-

7. Jacques Verron est un océanographe émérite de l’équipe MEOM

8. L’assimilation de données est parfois conçue comme un problème inverse à part entière.

vation.

On distingue essentiellement 2 grandes classes d'approche d'assimilation de données, pouvant être équivalentes sous certaines hypothèses : 1) l'approche séquentielle, qui repose sur l'interpolation statistique optimale et procède par la succession d'étapes de correction (d'analyse) et 2) l'approche variationnelle, qui découle de la théorie du contrôle optimal. Dans cette thèse, il sera uniquement question d'assimilation séquentielle de données. On se limitera à l'étape d'analyse (cf. chapitre 7), c'est-à-dire la mise à jour de l'état prédit par le modèle (ou de sa distribution de probabilité) à partir de l'ensemble des observations disponibles.

2.4.1 Formalisme probabiliste

Dans la section précédente, nous avons présenté, par l'intermédiaire de l'équation 2.3, l'approche stochastique de modélisation (d'ensemble). Pour se placer dans un cadre d'assimilation de données, il suffit d'ajouter les observations. L'objectif est donc désormais d'évaluer l'évolution d'un système pour lequel on dispose d'observations \mathbf{y}^0 . L'équivalent du vecteur d'état du modèle \mathbf{x}_{t_k} dans l'espace des observations est noté \mathbf{y}_{t_k} . La projection entre ces 2 espaces est réalisée via l'opérateur d'observation \mathcal{H} . En reprenant les notations utilisées pour présenter la modélisation stochastique, le problème peut se poser de la manière suivante :

$$\begin{cases} \mathbf{x}_{t_k} &= \mathcal{M}_{t_{k-1}, t_k}(\mathbf{x}_{t_{k-1}}, \mathbf{w}_{t_k}) & ; & \mathbf{x}(0) = \mathbf{x}_{t_0} \\ \mathbf{y}_{t_k} &= \mathcal{H}(\mathbf{x}_{t_k}, \mathbf{n}_{t_k}) \end{cases} \quad (2.4)$$

L'évolution vers l'état \mathbf{x}_{t_k} s'accompagne d'une incertitude caractérisée par les processus stochastiques \mathbf{w}_{t_k} . L'équivalent modèle des observations \mathbf{y}_{t_k} est déterminé avec une incertitude \mathbf{n}_{t_k} , provenant à la fois des erreurs sur l'opérateur d'observation (souvent associées à des erreurs d'interpolation pour passer de la grille du modèle à la grille des observations) et des erreurs de mesure des observations (cf. chapitre 3). À noter que les conditions initiales \mathbf{x}_{t_0} peuvent également être considérées comme incertaines. L'objectif de l'assimilation de données est de réduire les incertitudes de ce système grâce à l'information supplémentaire apportée par les observations \mathbf{y}^0 .

Sous la contrainte du système 2.4, une façon appropriée de formaliser l'assimilation de données de manière probabiliste est de l'exprimer sous la forme d'un problème d'estimation bayésienne (Lorenç, 1986; Tarantola, 1987).

$$P(\mathbf{x}_{t_k} | \mathbf{y}^o) \propto P(\mathbf{x}_{t_k}) P(\mathbf{y}^o | \mathbf{x}_{t_k}) \quad (2.5)$$

Ainsi formulée, la procédure d'analyse consiste "simplement" à déterminer la distribution de probabilité conditionnelle $P(\mathbf{x}_{t_k} | \mathbf{y}^o)$ du système d'être dans l'état \mathbf{x}_{t_k} connaissant les observations \mathbf{y}^o à partir de :

- $P(\mathbf{x}_{t_k})$, la distribution de probabilité a priori de l'état du modèle, c'est-à-dire de l'incertitude sur l'état \mathbf{x}_{t_k} , déterminée dans notre cas à partir de la simulation d'ensemble.

- $P(\mathbf{y}^o|\mathbf{x}_{t_k})$, la vraisemblance de l'état \mathbf{x}_{t_k} , c'est-à-dire la distribution de probabilité conditionnelle des observations \mathbf{y}^o connaissant \mathbf{x}_{t_k} . Elle dépend à la fois des opérateurs \mathcal{M} et \mathcal{H} et de l'incertitude sur les observations.

En réalité, il est presque impossible de connaître exactement ces distributions de probabilité⁹. Une des difficultés majeures à la résolution du problème de Bayes réside donc dans la détermination des incertitudes a priori du modèle et des observations. Dans cette thèse, l'évaluation de l'incertitude de l'erreur de prévisions du modèle sera déterminée à partir de la simulation d'ensemble. Nous considérerons donc que l'ensemble des incertitudes du modèle peuvent être caractérisées par notre approche stochastique. Nous faisons également l'hypothèse que l'échantillonnage de la distribution de probabilité par les membres de l'ensemble est suffisante pour représenter la distribution de probabilité.

2.4.2 Étape d'analyse : l'interpolation statistique

Dans cette thèse, l'assimilation de données se réduira à l'étape d'analyse. La description précise du schéma que nous utiliserons fera l'objet de la première partie du chapitre 7. Dans ce paragraphe, nous présentons la méthode d'interpolation statistique qui est l'une des techniques les plus simples pour répondre au problème de Bayes (équation 2.5) en considérant des distributions de probabilité gaussiennes des erreurs. Cette méthode peut être perçue comme un rappel aux observations utilisant des statistiques d'erreur.

On suppose qu'un modèle donne une prévision (ébauche) \mathbf{x}^b de l'état vrai \mathbf{x}^{true} d'un système observé par \mathbf{y}^o . L'erreur sur l'ébauche s'écrit alors $\varepsilon^b = \mathbf{x}^b - \mathbf{x}^{true}$ et celle sur les observations $\varepsilon^o = \mathbf{y}^o - \mathcal{H}\mathbf{x}^{true}$. On peut alors définir la matrice de covariance d'erreur a priori sur l'ébauche $\mathbf{P}^b = \mathbb{E}[(\varepsilon^b)(\varepsilon^b)^T]$ et la matrice de covariance des erreurs d'observation $\mathbf{R} = \mathbb{E}[(\varepsilon^o)(\varepsilon^o)^T]$. Pour estimer l'état analysé \mathbf{x}^a , le principe de l'interpolation optimale est de supposer qu'il peut s'écrire comme la combinaison linéaire de l'ébauche \mathbf{x}^b et des observations \mathbf{y}^o . Sous les hypothèses que l'opérateur d'observation \mathcal{H} est linéaire (i.e $\mathcal{H} \equiv \mathbf{H}$) et que les erreurs sont additives, indépendantes et sans biais (i.e $\mathbb{E}[\varepsilon^b] = \mathbb{E}[\varepsilon^o] = 0$), l'état analysé peut s'écrire :

$$\mathbf{x}^a = \mathbf{x}^b + \mathbf{K} \overbrace{(\mathbf{y}^o - \mathbf{H}\mathbf{x}^b)}^{innovation} \quad (2.6)$$

Avec ces approximations, l'analyse correspond à une régression linéaire. \mathbf{K} est un opérateur linéaire appelé matrice de gain qui passe de l'espace des observations dans l'espace du modèle. Pour le déterminer, on suit le schéma d'analyse BLUE (Best linear unbiased estimator) qui consiste à trouver un estimateur \mathbf{x}^a sans biais (i.e $\mathbb{E}[\varepsilon^a] = 0$ avec $\varepsilon^a = \mathbf{x}^a - \mathbf{x}^{true}$) et de variance minimale. L'expression de la matrice de gain (équation 2.7) s'obtient donc en minimisant la trace de la matrice de covariance d'erreur d'analyse $\mathbf{P}^a =$

9. Même lorsque $P(\mathbf{x}_{t_k})$ et $P(\mathbf{y}^o|\mathbf{x}_{t_k})$ prennent une forme analytique simple, la complexité des opérateurs \mathcal{M} et \mathcal{H} rend en général impossible de pouvoir expliciter $P(\mathbf{x}_{t_k}|\mathbf{y}^o)$

$E[(\varepsilon^a)(\varepsilon^a)^T]$ sous la supposition que les erreurs ε^b et ε^o sont décorréliées. Elle permet ensuite d'exprimer la matrice de covariance d'erreur d'analyse \mathbf{P}^a selon l'équation 2.8.

$$\mathbf{K} = \mathbf{P}^b \mathbf{H}^T (\mathbf{R} + \mathbf{H} \mathbf{P}^b \mathbf{H}^T)^{-1} \quad (2.7)$$

$$\mathbf{P}^a = (\mathbf{I} - \mathbf{K} \mathbf{H}) \mathbf{P}^b \quad (2.8)$$

L'ensemble du processus décrit ci-dessus constitue une étape d'analyse par interpolation statistique optimale. On constate qu'elle repose sur un grand nombre d'hypothèses, généralement difficiles à satisfaire. De plus, le coût informatique requis pour les calculs d'inversion de matrices rend l'application directe de cette méthode impossible en océanographie où la taille des matrices prend rapidement des proportions faramineuses. Un grand nombre de travaux se penche ainsi sur la mise au point de moyens pour réduire la taille des matrices de covariance en altérant le moins possible les résultats de l'analyse.

2.4.3 D'un filtre de rang réduit à un filtre d'ensemble

En océanographie, l'assimilation séquentielle de données est essentiellement fondée sur des méthodes dérivées de la théorie du filtre de Kalman (Kalman, 1960). L'une des plus utilisées (du moins en France) est le filtre SEEK (Sequential Evolutive Extended Kalman filter, Pham et al., 1998a) adapté aux problèmes de l'océanographie dans un contexte de modélisation déterministe (Brasseur and Verron, 2006). Il est basé sur l'approximation des matrices de rang réduit, permettant de diminuer la taille des matrices à inverser. Les covariances d'erreur sont calculées dans un sous-espace d'erreur généré la plupart du temps à partir du calcul des EOF (fonctions orthogonales empiriques, Kutzbach, 1967). Le coût numérique est ainsi réduit en ne tenant compte que des modes dominants des erreurs.

Toutefois, dans le filtre SEEK, la propagation de la matrice des covariances d'erreur repose encore sur l'hypothèse d'un modèle linéaire. Une fois la mise à jour effectuée (cf. paragraphe précédent), l'étape suivante de l'assimilation de données consiste à propager l'expression de \mathbf{P}^a afin d'obtenir une estimation des covariances d'erreur de l'état ébauche pour le pas de temps suivant. Cette propagation nécessite de linéariser le modèle de prévision (i.e $\mathcal{M} \equiv \mathbf{M}$). Dans le contexte de modélisation biogéochimique, la forte non-linéarité des équations rend cette approximation difficilement justifiable. Pour y remédier, des méthodes de filtre de Kalman d'ensemble ont été mises au point (p.ex. Evensen, 1994, 2003; Burgers et al., 1998). Dans ces méthodes, les matrices de covariance de rang réduit sont représentées par un ensemble d'états. À chaque pas de temps, la distribution statistique des erreurs est ainsi évaluée à partir des membres de l'ensemble. La propagation des covariances d'erreur ne nécessite donc plus la contrainte de linéarité. Subsiste néanmoins l'hypothèse que nous puissions disposer de suffisamment de membres pour échantillonner correctement les distributions de probabilité.

Quelle que soit la méthode utilisée, une définition appropriée des matrices de covariance d'erreur reste essentielle pour les performances de l'assimilation de données (Nerger et al., 2005). Le plus souvent, le sous-espace des erreurs a priori est échantillonné uniquement à partir de l'incertitude sur les données d'entrée du modèle (cf. section 2.3). Dans l'objectif d'améliorer la représentation des covariances d'erreur, nous développerons dans cette

thèse une approche stochastique. L'introduction à chaque itération du modèle de processus stochastiques permettra de générer des structures d'erreur dynamiques directement associées aux incertitudes des formulations biogéochimiques. L'un des principaux objectifs de cette thèse sera de démontrer la pertinence des covariances d'erreur générées selon cette approche. Dans le dernier chapitre, nous présenterons pour cela des analyses réalisées à partir du schéma d'analyse du Filtre de Kalman d'Ensemble Transformé (ETKF, Bishop et al., 2001).

Fon-li,

Chapitre 3

Les observations satellites et le modèle couplé NEMO/PISCES

Sommaire

3.1	L'observation satellite de la couleur de l'océan	34
3.1.1	Principe de la mesure et incertitudes	34
3.1.2	Les différentes missions	36
3.1.3	État de l'art en assimilation de données couleur de l'eau	38
3.2	Le modèle couplé NEMO/PISCES	40
3.2.1	Le modèle de circulation océanique OPA	41
3.2.2	Le modèle biogéochimique PISCES	45
3.2.2.1	Complexité biogéochimique	45
3.2.2.2	Équations de la production primaire	48
3.3	Mise en place du modèle couplé au 1/4° sur l'Atlantique Nord	53
3.3.1	Caractéristiques du couplage	53
3.3.2	Configuration du modèle	54
3.4	Conclusion	59

Dans le chapitre précédent, nous avons présenté le cadre probabiliste de conception des incertitudes qui sera développé dans la suite de ce manuscrit. Au préalable, l'un des premiers objectifs de cette thèse est de mettre en place une configuration réaliste du modèle couplé NEMO/PISCES sur l'Atlantique Nord, pouvant être confrontée aux données satellites de la couleur de l'océan. Dans ce chapitre, nous décrirons ces 2 composantes : le modèle et les observations.

Dans la première partie de ce chapitre, nous aborderons le principe de la télédétection spatiale de la couleur de l'océan. Dans le même temps, nous présenterons les principales caractéristiques des données qui seront exploitées par la suite. Je précise dès maintenant que nous prendrons un positionnement "utilisateur" vis à vis de ces observations. Outre des expériences d'assimilation de données, elles seront utilisées pour vérifier la pertinence des simulations que nous réaliserons à partir de la plate-forme de modélisation NEMO. Les composantes de la physique (OPA) et de la biogéochimie (PISCES) seront décrites dans la deuxième section. Nous terminerons ce chapitre en présentant la configuration couplée au $1/4^\circ$, mise en place sur l'Atlantique Nord dans le but d'atteindre une représentation de la chlorophylle comparable avec les observations de la couleur de l'eau.

3.1 L'observation satellite de la couleur de l'océan

L'observation spatiale est aujourd'hui le seul moyen de disposer d'une vision quasi synoptique de l'océan. Elle permet l'accès à des données de résolution spatiale pouvant aller jusqu'au kilomètre et de résolution temporelle quasi journalière. Les observations satellites sont, de ce fait, un élément incontournable de la compréhension globale de l'océan.

Dans le cas de la couleur de l'océan, c'est le contenu en chlorophylle-a (principal pigment du phytoplancton) qui est évalué. La première partie de cette section explique comment cette mesure est effectuée, en insistant sur les difficultés (incertitudes) d'une observation réalisée depuis l'espace. Nous présenterons ensuite les principales caractéristiques des données satellites que nous utiliserons dans cette thèse. La dernière partie de cette section dresse un bilan des principaux travaux d'assimilation de données menés à partir d'observations de la couleur de l'eau dans une configuration réaliste d'océan.

3.1.1 Principe de la mesure et incertitudes

La mesure de la couleur de l'océan

Les observations de la couleur de l'eau proviennent de la partie du spectre de la lumière du soleil réfléchi par les premières couches de l'océan. Lorsqu'elle pénètre dans l'océan, la lumière interagit avec les différents éléments qui composent l'eau de mer¹. Parmi ces interactions, une partie du spectre de la lumière (principalement les parties rouges et bleues) est absorbée par la chlorophylle du phytoplancton pour réaliser la photosynthèse. La partie non absorbée du rayonnement peut ensuite être réémise vers l'espace, où elle est alors détectée par les capteurs (passifs) des satellites. Cette énergie radiative est ensuite convertie en signal électrique dont l'intensité est fonction du rayonnement reçu. L'information qui

1. La lumière interagit également avec l'eau pure

permet d’estimer la concentration en chlorophylle provient donc de la diffusion de la lumière juste en dessous de la surface de l’océan, c’est-à-dire de la “couleur de l’eau”.

Pour être plus précis, la couleur de l’océan est déterminée par les variations spectrales à la longueur d’onde λ de la réflectance $\mathbf{R}(\lambda)$ diffusée par la couche supérieure de l’océan. Cette réflectance est définie comme le rapport entre l’éclairement ascendant juste à la surface de l’océan $E_a(\lambda)$ et l’éclairement descendant pénétrant dans l’océan $E_d(\lambda)$: $\mathbf{R}(\lambda) = E_a(\lambda)/E_d(\lambda)$. Les variations de cette réflectance dépendant de la composition de l’eau et de la nature des particules en suspension, l’inversion du signal de réflectance permet de retrouver les caractéristiques de l’eau de mer. Comme les interactions entre la lumière et l’eau de mer sont multiples, il est nécessaire de disposer d’instruments spécifiques à la quantité que l’on souhaite observer. Pour la couleur de l’océan, plusieurs capteurs (observation multi spectrale) sont ainsi utilisés pour balayer les parties du spectre de la lumière s’étendant de la limite du visible au proche infrarouge (en général entre 400 et 900 nm, voir table 3.1). Les concentrations en chlorophylle sont déduites du rapport de réflectance entre la longueur d’onde correspondant à la couleur verte et celle correspondant au bleu. Les autres longueurs d’ondes servent aux corrections atmosphériques (voir paragraphe suivant) et à l’évaluation d’autres composants de l’eau de mer (autres pigments, matériel particulaire, etc...).

Incertitudes des données

Si les données satellites ont “révolutionné” l’observation de l’océan, déduire des informations pertinentes sur l’état de l’océan à partir d’un signal lumineux détecté depuis l’espace est loin d’être évident. En plus des lourdes contraintes d’ingénierie que nous n’aborderons pas ici, la mesure de réflectances convertibles en concentration de chlorophylle pose de nombreuses difficultés. Les données satellites reposent donc inévitablement sur d’importantes approximations et possèdent alors un degré d’incertitude non négligeable. Dans la suite de ce paragraphe, nous présentons quelques-unes des principales sources d’incertitude des observations de la couleur de l’océan.

- Une contrainte majeure à l’observation de la couleur de l’eau vient de la faiblesse du signal que l’on cherche à identifier. La fraction à mesurer ne représente en effet qu’environ 10% de l’intensité lumineuse reçue par le capteur (Antoine, 1998). Le reste du signal est issu de la réflexion de la lumière à l’interface océan-atmosphère (avant qu’elle ne pénètre dans l’océan), mais surtout de la diffusion atmosphérique. Ces contributions doivent donc être supprimées du signal total avant de pouvoir en extraire une estimation de la chlorophylle de surface. L’ensemble des traitements se regroupe sous le terme de corrections atmosphériques. À noter que les nuages bloquent toutes les émissions dans le visible et l’infrarouge. Lorsque la couverture nuageuse est trop importante, l’océan ne peut donc pas être observé. Il en résulte des “zones d’ombres” ou “trous” dans les données satellites (cf. figure 3.1).
- L’estimation de la part du signal due à la diffusion par les aérosols constitue la plus importante source d’incertitude des corrections atmosphériques. Elle est désormais évaluée à partir de capteurs qui mesurent le signal dans le domaine du proche infrarouge. L’océan absorbant intégralement cette bande spectrale de la lumière, le signal détecté ne provient que de l’atmosphère. Les propriétés optiques dans le proche infrarouge sont ensuite extrapolées dans le visible pour corriger la partie du spectre dont la chlorophylle est déduite.

- En plus des erreurs dites « de mesure », dues aux imprécisions des instruments, la conversion d'un signal optique en données de chlorophylle constitue également une importante source d'incertitude. Elle repose sur l'utilisation d'algorithmes fondés sur la théorie du transfert radiatif. Les propriétés optiques de l'océan pouvant fortement varier pour une concentration en chlorophylle donnée, ils restent relativement approximatifs. L'algorithme de reconstruction OC4 (Ocean Chlorophyll 4, O'Reilly et al., 2000) est aujourd'hui l'un des plus utilisés. Pour rendre compte de l'impact des différentes compositions de l'eau de mer sur l'estimation de la chlorophylle, les eaux sont généralement divisées en 2 catégories (Morel and Prieur, 1977).

- Les eaux de « Cas1 », constituées par les eaux du large (95% de l'océan), non affectées par les influences de la côte et du fond des océans. Les propriétés optiques de ces eaux sont caractérisées par la présence de phytoplancton (chlorophylle-a) et des substances particulaires et dissoutes qui lui sont associées .
- Les eaux de « Cas2 », constituées par les eaux qui subissent l'influence du littoral. Les propriétés optiques sont caractérisées par la présence de nombreux sédiments particulaires et dissous tels que les CDOM (Colored Dissolved Organic Matter) affectant la mesure de la couleur de l'eau. La variabilité spatiale de la composition de ces eaux rend l'inversion du signal beaucoup plus difficile que pour les eaux de Cas1.

Enfin, il est évident que les paramètres du satellite (angle d'observation, inclinaison, traitement de la polarisation, etc...) ont une influence non négligeable sur la qualité du signal optique reçu par le capteur. Les détailler ici est toutefois hors propos. Ce qui nous intéresse plus particulièrement, est le niveau d'incertitude final des observations. Compte tenu de la multiplicité des sources potentielles d'incertitude, l'estimation de cette erreur est difficile à établir. Elle varie de manière conséquente en fonction de la localisation géographique et de la période d'observation. De manière générale, le niveau d'erreur admis pour les eaux du large (Cas1) avoisine les 30% (Gregg and Casey, 2004; Moore et al., 2009). Pour les eaux du Cas2, il peut aller jusqu'à plus de 100%.

Lorsqu'il faudra quantifier précisément le niveau d'incertitude des observations de la couleur de l'eau, une erreur de 30% sera conventionnellement considérée. Il faudra néanmoins toujours garder à l'esprit que les mesures satellites sont elles aussi incertaines et ne représentent donc pas une vérité absolue sur la chlorophylle de surface. Pour plus de détails sur l'observation par satellite en océanographie, le lecteur peut se référer à l'ouvrage très complet de Robinson (2004).

3.1.2 Les différentes missions

Comme déjà évoqué en introduction, l'observation de la couleur de l'océan a débuté en 1978, avec le lancement par la NASA² du capteur CZCS³ à bord du satellite Nimbus 7. Initialement prévu pour un fonctionnement d'une année, ce capteur a poursuivi ses mesures jusqu'en 1986. La réussite de cette première mission a ouvert la voie à plus d'une dizaine de satellites capables de mesurer la couleur de l'eau. Parmi ces missions, les capteurs qui ont observé le plus longtemps l'océan sont : SeaWiFS-SeaStar, MERIS-Envisat

2. National Agency Space Administration

3. Coastal Zone Color Scanner

et MODIS-Aqua. Leurs principales caractéristiques techniques sont récapitulées en table 3.1. Un exemple de la représentation de la chlorophylle de surface au $1/12^\circ$ pouvant être obtenue sur l’Atlantique Nord à partir des données de ces 3 capteurs est présenté en figure 3.1. Les informations concernant ces missions d’observation et l’accès aux données de la chlorophylle de surface sont disponibles à l’adresse <http://oceancolor.gsfc.nasa.gov>.

	SeaWIFS	MERIS	MODIS
Satellite	SeaStar-OrbView2 (USA)	Envisat (Europe)	Aqua (USA)
Période d’activité	1997-2010	2002-2012	2002-aujourd’hui
Bande spectrale	412-865 nm	412.5-900 nm	412-865 nm
Nombre de capteurs	8	15	9
Altitude d’orbite	705 Km	799.8 Km	705 Km
Période orbitale	99 min	100.6 min	98.4 min
Largeur de fauchée	2800 Km	1150 Km	2330 Km
Résolution au nadir	1130 m	300 m	1000 m

TABLE 3.1 – Principales caractéristiques techniques des capteurs d’observation de la couleur de l’océan SeaWIFS, MERIS et MODIS

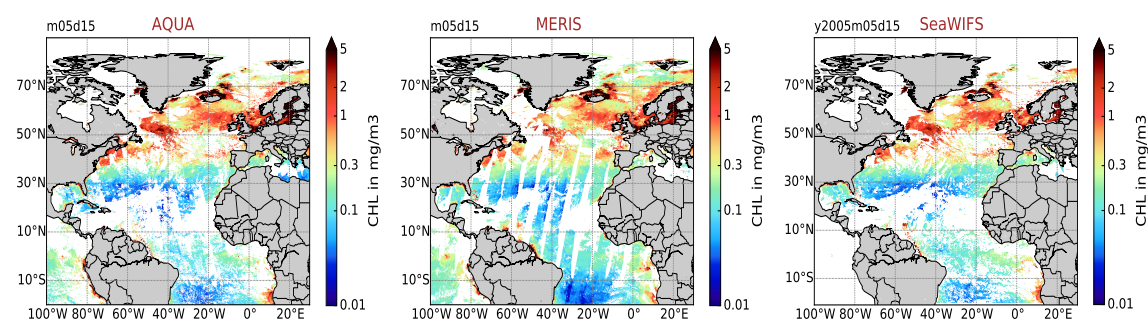


FIGURE 3.1 – Images satellites au $1/12^\circ$ de la chlorophylle de surface de l’Atlantique Nord obtenues à partir des observations MODIS-Aqua (gauche), MERIS-ENVISAT (centre) et SeaWIFS-SeaStar (droite). Ces observations sont présentées en moyenne de 5 jours pour le 15 mai 2005.

Si l’observation de la chlorophylle faite par ces 3 capteurs paraît relativement similaire, (cf. figure 3.1), on constate que leurs couvertures nuageuses diffèrent significativement. Pour éviter la présence de “trous” dans les observations, ces données peuvent être combinées. Dans le cadre du projet OC-CCI (Ocean Colour Climate Change Initiative) de l’ESA⁴, les données MERIS et MODIS ont ainsi été basculées sur les longueurs d’ondes de SeaWIFS pour combler une partie des données manquantes. Le résultat propose, en quelque sorte, une version améliorée des observations SeaWIFS. Ces données font partie des produits du projet MyOcean-Copernicus accessibles à l’adresse <http://marine.copernicus.eu>. La figure 3.2 présente des cartes de chlorophylle de surface obtenues à partir de ces données de haute résolution ($1/24^\circ$). Par abus de langage, ces observations seront renommées SeaWIFS-HR.

4. Agence Spatiale Européenne

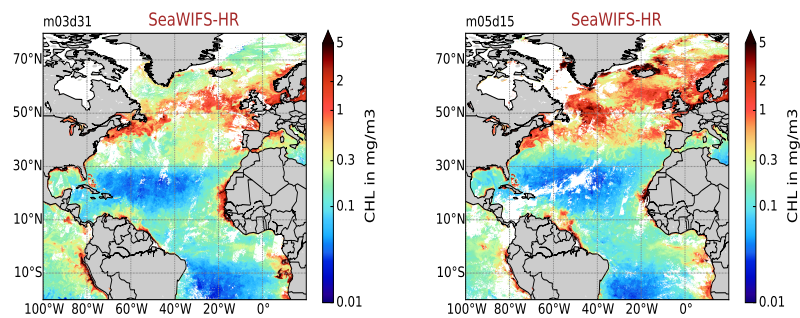


FIGURE 3.2 – Images satellites au $1/24^\circ$ de la chlorophylle de surface de l’Atlantique Nord obtenues à partir des données d’observations SEAWIFS-HR. Ces observations sont présentées en moyenne de 5 jours pour le 15 mai 2005.

Dans cette thèse, les observations SeaWIFS seront utilisées pour ajuster les paramètres biogéochimiques de la simulation d’ensemble et valider les résultats des simulations du modèle PISCES. Afin de s’affranchir, en partie, des difficultés liées à l’extrapolation horizontale des données (dus à la présence de “trous” dans les observations), les données au $1/24^\circ$ SeaWIFS-HR seront utilisées pour réaliser les expériences d’analyse (correction de l’état prédit par le modèle). Les performances de ces analyses seront évaluées avec le jeu d’observation MERIS. Les données du satellite AQUA ne seront qu’indirectement utilisées par l’intermédiaire des observations SeaWIFS-HR.

3.1.3 État de l’art en assimilation de données couleur de l’eau

De manière générale, la comparaison directe entre les modèles biogéochimiques et les observations de la couleur de l’eau pose de sérieuses complications. En effet, les modèles ne représentent qu’une partie des processus à l’origine des distributions de chlorophylle observées par les satellites. Si ce point est une motivation forte pour l’assimilation de données de la couleur de l’eau, il en augmente dans le même temps la difficulté.

En assimilation de données de la couleur de l’eau, c’est l’estimation de paramètres, via des méthodes variationnelles, qui a reçu le plus d’attention. Ces travaux ont principalement été motivés par le programme JGOFS (Joint Global Ocean Flux Study). L’objectif de cette approche est d’améliorer la performance des modèles en réajustant la valeur de leurs paramètres (p.ex. Fasham et al., 1995; Schartau et al., 1998; Fennel et al., 2001). La difficulté pour mettre en oeuvre les méthodes variationnelles a cependant contraint les études à des configurations de faible dimension spatiale (en général 1D). Elles ont toutefois permis, entre autres, de souligner le niveau élevé d’incertitude des paramètres biogéochimiques des modèles. L’article de Gregg et al. (2009) propose une revue complète de ces études.

En biogéochimie, l’assimilation séquentielle de données ne s’est développée que plus tardivement (à partir des années 2000). À l’exception de quelques travaux (p.ex. Hoteit et al., 2003, 2004; Magri et al., 2005), les études se sont essentiellement concentrées sur l’estimation d’état. Le tableau 3.2, inspiré par Gregg et al. (2009) et complété pour ce manuscrit, présente une liste non exhaustive des publications traitant de l’assimilation de données de la couleur de l’eau dans une configuration d’océan réaliste. Le simple fait de pouvoir ainsi énumérer ces travaux souligne à lui seul leur faible quantité.

Auteurs	Méthode	Dimension	Zone d’étude	Observation
Ishizaka, 1990	Nudging	3D	N.Atlantic	CZCS
Armstrong et al., 1995	Nudging	3D	N.Atlantic	CZCS
Semovski and Wozniak, 1995	KF & variationnelle	3D	N.Atl-Baltic	CZCS
Friedrichs, 2001	Variationnelle	1D	Pac-Equatorial	SeaWIFS
Garcia-Gorritz et al., 2003	Variationnelle	3D	Adriatic Sea	SeaWIFS
Hemmings et al., 2003	Variationnelle	0D	N.Atlantic	SeaWIFS
Natvik and Evensen, 2003a,b	Séquentielle EnKF	3D	N.Atlantic	SeaWIFS
Hemmings et al., 2004	Variationnelle	0D	N.Atlantic	SeaWIFS
Losa et al., 2004	Séquentielle SIRV	0D	N.Atlantic	CZCS
Triantafyllou et al., 2007	Séquentielle SEEK	3D	Méditerranée	SeaWIFS
Nerger and Gregg, 2007	Séquentielle SEIK	3D	global	SeaWIFS
Gregg, 2008	Séquentielle CRAM	3D	global	SeaWIFS
Nerger and Gregg, 2008	Séquentielle SEIK	3D	global	SeaWIFS
Ciavatta et al., 2011	Séquentielle EnKF	3D	English channel	MODIS
Ford et al., 2012	Séquentielle	3D	global	GlobColour
Hu et al., 2012	Séquentielle EnKF	3D	côte Est US	SeaWIFS & MODIS
Mattern et al., 2013	Séquentielle SIR	3D	côte Est US	SeaWIFS
Xiao and Friedrichs, 2014	Variationnelle	1D	côte Est US	SeaWIFS
Fontana et al., 2013	Séquentielle SEEK	3D	N.Atlantic	SeaWIFS
Doron et al., 2013	Séquentielle SEEK	3D	N.Atlantic	GlobColour

TABLE 3.2 – Tableau récapitulatif (non-exhaustif) des principaux efforts en assimilation de données satellites de la couleur de l’eau dans une configuration réaliste d’océan.

L’étude de Ishizaka (1990) fait office de précurseur en montrant la première que l’assimilation d’observations de la couleur de l’eau pouvait améliorer la représentation de la chlorophylle dans un modèle 3D. Les données satellites étaient pourtant disponibles depuis de nombreuses années. Le tableau 3.2 met en évidence la prédominance de techniques séquentielles d’assimilation, en particulier le SEEK. On constate également que seulement 3 études proposent une démarche d’ensemble. Parmi elles, les travaux de Natvik and Evensen (2003a,b) sont considérés comme la première expérience démontrant la faisabilité de l’assimilation d’ensemble de données de la couleur de l’eau dans une configuration réaliste à l’échelle d’un bassin. Comparé à notre étude, le modèle reste relativement simple (11 variables, Drange (1996) contre 24 dans PISCES) et de basse résolution ($\approx 1^\circ$ contre $1/4^\circ$). Dans un contexte plus régional, les travaux de Ciavatta et al. (2011) ont mis en évidence l’importance d’utiliser un modèle biogéochimique de complexité suffisante. Ils ont également souligné l’importance de la définition des statistiques d’erreur du modèle pour les performances de l’assimilation de données. Les travaux de Hu et al. (2012) sont ceux qui se rapprochent le plus de cette thèse. Les membres de l’ensemble sont, en effet, initialisés en perturbant 5 paramètres biogéochimiques du modèle. Le modèle n’est cependant pas stochastique, les perturbations étant uniquement appliquées à l’initialisation. Cette étude prouve néanmoins que considérer l’incertitude des paramètres permet une estimation positive des corrélations d’erreur. Elle reste toutefois d’échelle régionale (côte Est américaine), et utilise un modèle biogéochimique de faible complexité (7 variables).

3.2 Le modèle couplé NEMO/PISCES

Toutes les simulations présentées dans ce document ont été réalisées à partir de la plateforme de modélisation océanique aux équations primitives NEMO (Nucleus for European Modelling of the ocean) dans sa version NEMO-3.4 (Madec et al., 2012). Le développement du système NEMO est toujours en permanente activité. Ce système est aujourd’hui majoritairement utilisé en Europe pour étudier de manière réaliste la dynamique de l’océan et ses interactions avec les différentes composantes du système climatique. Son architecture s’articule autour de 3 modules principaux :

- le module OPA (Ocean PARallélisé, Madec et al., 1998), qui modélise la dynamique de la circulation de l’océan,
- le module LIM (Louvain-la-Neuve sea Ice Model, Fichefet and Morales Maqueda, 1997), qui modélise le comportement dynamique et thermodynamique de la glace de mer,
- le module TOP (Tracer in the Ocean Paradigm), qui calcule le transport des traceurs passifs et l’évolution des variables biogéochimiques obtenues à partir des termes *SMS* (sources minus sink) des modèles PISCES ou LOBSTER.

La composante principale du système est le module OPA qui gère à la fois la mise en place de la configuration numérique de la zone d’étude et le calcul de la circulation océanique. Il fournit ainsi le cadre sur lequel sont ensuite couplés les modules LIM et TOP. Dans cette étude, les caractéristiques de la glace de mer (module LIM) sont calculées via le modèle LIM-v2. Le modèle utilisé pour le calcul des traceurs biogéochimiques (module TOP) est le modèle PISCES-v2 (Aumont et al., 2015). En plus de ces 3 compartiments, on peut noter l’existence du module AGRIF (Adaptative Grid Refinement in Fortran, Debreu et al., 2008), qui permet un raffinement interactif des mailles du modèle et du module TAM (Tangent linear and adjoint models, Vidard et al., 2015) qui correspond au linéaire tangent et adjoint pour l’assimilation de données variationnelle. Un couplage à l’atmosphère est également possible par l’intermédiaire du module OASIS (Valcke, 2006). Ces 3 derniers modules ne seront pas utilisés dans cette étude.

La représentation de la physique de l’océan est incontestablement un élément clé pour la représentation de la biogéochimie marine. Cet aspect n’étant cependant pas au cœur des objectifs de cette étude, notre démarche a été de se munir d’une représentation physique la plus réaliste possible. Pour cela, la configuration NATL025 utilisée ici, qui est un sous élément de la configuration ORCA025 (Barnier et al., 2006), est héritée du projet DRAKKAR (p.ex. Barnier et al., 2015) dont l’objectif est précisément la mise au point de configurations adaptées à la modélisation opérationnelle.

Dans la première partie de cette section, nous présenterons les caractéristiques du modèle physique OPA utilisé pour décrire la composante physique de l’océan. La seconde partie détaille le fonctionnement du modèle PISCES en se focalisant sur les équations de la production primaire. La dernière partie de cette section explicite les paramètres de couplage entre la configuration NATL025 et le modèle PISCES.

3.2.1 Le modèle de circulation océanique OPA

La circulation océanique est modélisée à partir des équations primitives du modèle tridimensionnel OPA. Il décrit l'océan comme un fluide en rotation dont le mouvement est régi par les équations de Navier Stokes (équivalent du principe fondamental de la dynamique appliqué à un fluide) et par l'équation d'état de l'eau de mer (équation 3.6) auxquelles s'ajoutent les approximations ci-dessous.

1. L'approximation de Boussinesq.
Les fluctuations de la densité ρ sont supposées négligeables mis à part dans leur contribution aux termes de flottabilité.
2. L'hypothèse d'incompressibilité.
Le volume d'un fluide reste toujours constant. Sous les approximations de Boussinesq, cette hypothèse permet d'établir la non-divergence de l'écoulement (équation 3.2) à partir de la simplification de l'équation de continuité.
3. L'approximation hydrostatique.
Les échelles de la dynamique horizontale sont supposées être bien plus grandes que les échelles de la dynamique verticale. Cette hypothèse permet de simplifier l'équation 3.1 de Navier Stokes projetée sur la verticale. Elle conduit à l'équation 3.3 qui traduit l'équilibre hydrostatique entre les termes de flottabilité et le gradient vertical de pression P . Il est à noter que cette approximation supprime des processus convectifs et les écoulements gravitaires prévus par les équations du mouvement. Ils doivent par conséquent être paramétrés avec des schémas spécifiques.
4. L'approximation de couche mince.
La profondeur de l'océan est supposée faible devant le rayon terrestre. Les effets de courbure verticale sont négligés dans le système de coordonnées sphériques.
5. L'approximation sphérique.
Les surfaces géopotentielle sont supposées être sphériques. Cette hypothèse permet de considérer que la gravité est parfaitement parallèle au rayon terrestre.
6. L'hypothèse de fermeture turbulente.
Les flux turbulents des échelles sous-maille (non résolues) sont paramétrés à partir des gradients de plus grande échelle.

L'ensemble de ces hypothèses permet d'établir le système d'équations primitives (équations 3.1 à 3.6) qui décrit l'évolution temporelle de la circulation océanique à partir de 4 variables pronostiques (directement résolues par les équations du modèle) : la température potentielle T , la salinité S et le champ horizontal des vitesses $\vec{U}_h = (u, v)$. Le modèle étant hydrostatique, il est important de mentionner que la vitesse verticale ne fait pas partie des variables pronostiques. Elle est calculée à partir de l'équation 3.2 et se déduit donc de la conservation de la masse dans le cas d'un écoulement non divergent. Je précise que l'incertitude sur l'estimation de l'advection verticale, qui joue un rôle important sur le mélange océanique, est un élément non négligeable pour la biogéochimie.

Dans les équations présentées à la page suivante, \vec{U} représente le champ total des vitesses (u, v, w) . Le paramètre de Coriolis f est fonction de la latitude ϕ et de la rotation

de la terre $\Omega : f = 2\Omega \sin\phi$. Le terme \vec{F}^U est un terme source de quantité de mouvement. Les termes \vec{D}^{U_h} , D^T et D^S sont des paramétrisations sous-maille dont les expressions seront explicitées un peu après dans le document. C'est dans ces termes que sont inclus les forçages atmosphériques (cf. section 3.3). Ils gouvernent en grande partie les échanges air/mer (par exemple thermiques et solaires) et la dynamique des échanges verticaux, dont la représentation est elle aussi cruciale pour la biogéochimie.

$$\left(\frac{\partial}{\partial t} + \vec{U} \cdot \vec{\nabla}\right) \vec{U}_h = -\frac{1}{\rho_0} \vec{\nabla}_h P - f \vec{k} \wedge \vec{U}_h + \vec{F}^U + \vec{D}^U \quad (3.1)$$

$$\vec{\nabla} \cdot \vec{U} = 0 \quad (3.2)$$

$$\frac{\partial P}{\partial z} = -\rho g \quad (3.3)$$

$$\frac{\partial T}{\partial t} = -\vec{\nabla} \cdot (T\vec{U}) + D^T \quad (3.4)$$

$$\frac{\partial S}{\partial t} = -\vec{\nabla} \cdot (S\vec{U}) + D^S \quad (3.5)$$

$$\rho = \rho(T, S, P) \quad (3.6)$$

Discretisation des équations

La résolution numérique de ces équations nécessite de les considérer sur une grille discrète. Pour le modèle OPA, la discrétisation spatio-temporelle est réalisée de la façon suivante :

-Spatiallement, le modèle OPA se base sur un repère en coordonnées sphériques sous les approximations de couche mince et de surface sphérique décrites précédemment. Dans ce repère, la discrétisation des variables se fait sur une grille Arakawa de type C (Arakawa and Lamb, 1980, figure 3.3 (a)). Les grandeurs scalaires sont calculées au centre de la maille (point T), les grandeurs vectorielles au centre des faces de la maille (point U,V,W) et le rotationnel des grandeurs vectorielles est calculé au centre des arêtes de la maille (point f). Les coordonnées verticales sont définies sur un nombre de niveaux constant, dont l'épaisseur augmente de la surface vers le fond. La figure 3.3 (b) illustre ce point dans le cas de 46 niveaux verticaux (résolution verticale utilisée dans cette étude). Les équations sont discrétisées en « partial step », ce qui signifie que le dernier niveau vertical n'est pas constant, mais est ajusté en fonction de la bathymétrie. En chaque point de grille, la discrétisation s'effectue selon une méthode de différences finies centrées d'ordre 2.

-Temporellement, les équations sont initialisées en suivant un schéma d'Euler, qui ne nécessite que la connaissance du pas de temps présent pour calculer le pas de temps suivant. Aux pas de temps suivants ou lorsque le modèle est initialisé à partir d'un fichier de « restart », on utilise un schéma leapfrog centré, auquel est associé un filtre de diffusion temporel d'Asselin (Asselin, 1980) qui permet d'éliminer le bruit numérique propre au schéma leapfrog. Je précise que ce schéma n'est pas adapté à la représentation des processus diffusifs,

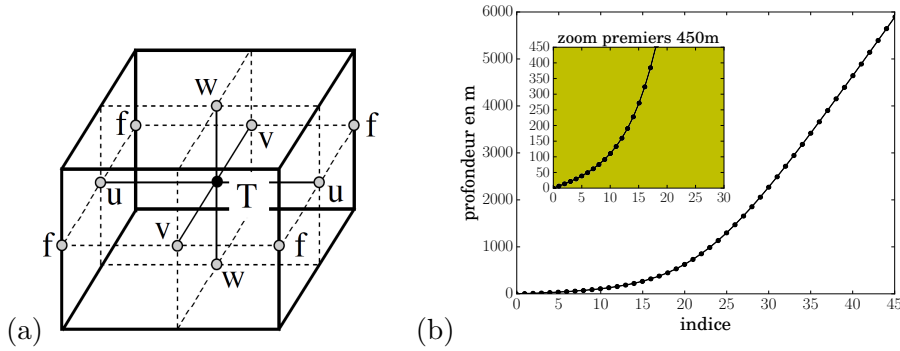


FIGURE 3.3 – (a) Schéma de la grille Arakawa C (Arakawa and Lamb, 1980). (b) Profondeur des 46 niveaux verticaux.

dont la discrétisation nécessite l'utilisation de schémas décentrés pour conserver une stabilité numérique. Il est important de préciser que ce schéma n'est pas non plus utilisé pour l'advection des traceurs biogéochimiques qui suivent un schéma d'Euler décentré avant.

Paramétrisations sous-maille

À cause de la discrétisation numérique des équations, certains processus dont les échelles spatio-temporelles caractéristiques sont plus fines que la résolution du modèle ne peuvent pas être explicitement représentés. Une quantité non négligeable d'énergie turbulente qui devrait être déplacée vers les petites échelles reste confinée à l'échelle de la grille du modèle, pouvant provoquer d'importantes instabilités numériques. L'impact de ces processus sous-maille sur la plus grande échelle est exprimée par les termes de diffusion $D_{U_h}^h$, D_T^h et D_S^h (cf. équations 3.1, 3.4 et 3.5) en suivant l'hypothèse de fermeture turbulente explicitée précédemment.

-Sur l'horizontale, les flux de diffusion latérale D_T^h et D_S^h des traceurs actifs T et S sont exprimés localement, le long des surfaces de densité potentielle constante, à partir d'un opérateur laplacien. La dissipation latérale de quantité de mouvement $D_{U_h}^h$ est exprimée le long des surfaces géopotentiels à partir d'un opérateur bilaplacien.

$$D_x^h = A_x^h \Delta x \quad \text{avec } x = (T, S) \quad (3.7)$$

$$D_{\vec{u}_h}^h = \vec{\Delta}(A_{U_h}^h \vec{\Delta} \vec{U}) \quad (3.8)$$

-Sur la verticale, les flux de diffusion turbulente, à l'origine du mélange vertical, sont paramétrés par analogie avec la diffusion moléculaire selon l'équation 3.9 .

$$D_X^v = \frac{\partial}{\partial z} (A_X^v \frac{\partial X}{\partial z}) \quad \text{avec } X = \vec{u}_h, T, S \quad (3.9)$$

Les termes de diffusivité turbulente A_T^v et A_S^v et le terme de viscosité turbulente $A_{\vec{u}_h}^v$ sont ensuite exprimés en suivant le schéma de fermeture turbulente TKE (Blanke and Delecluse, 1993) qui s'appuie sur le calcul de l'équation pronostique de l'énergie cinétique.

Une paramétrisation d’ajustement convectif (enhanced vertical diffusion), autorisant des valeurs élevées aux termes de diffusivité et de viscosité turbulente dans les régions où la stratification est instable, est également associée au schéma TKE. Elle permet de simuler le mélange généré par dissipation d’instabilités baroclines induites par la convection hivernale (\approx advection verticale) qui ne peut être pris en compte dans un modèle hydrostatique.

La représentation du mélange océanique, en particulier celle du mélange vertical est fondamentale pour la biogéochimie. Elle conditionne entre autres l’estimation de la couche de mélange océanique (couche de surface homogène en température) dont les variations saisonnières d’épaisseur régulent en grande partie les apports de nutriments vers la surface. L’estimation de ce mélange par le modèle est donc primordiale pour pouvoir modéliser de manière pertinente les processus biogéochimiques (Berline et al., 2007; Béal et al., 2010).

Advection

L’advection exprime le transport des propriétés d’une particule fluide par le champ vectoriel des vitesses d’écoulement. Il est exprimé par l’opérateur $\vec{U} \cdot \vec{\nabla}$ et s’applique à la quantité de mouvement (équation 3.1) et aux traceurs actifs de température et de salinité (équations 3.4 et 3.5). Comme le modèle est hydrostatique, l’advection verticale ne peut pas être directement explicitée. Elle est donc déterminée à partir de l’intégration verticale de l’équation de continuité. À noter que l’advection issue de processus non résolus est paramétrée par la diffusion turbulente (mélange). Pour les traceurs passifs, le terme d’advection sera explicité en section 3.30. Dans notre étude, la discrétisation de l’advection des traceurs est réalisée à partir de la formulation TVD (Total Variance Dissipation, LeVeque, 1978) qui combine un schéma de discrétisation temporelle centré d’ordre 2 avec un schéma explicite. Cette méthode évite d’avoir des concentrations négatives comme ce peut être le cas avec des schémas centrés d’ordre 2.

L’advection horizontale fait partie des facteurs essentiels à la représentation biogéochimique de l’océan. Elle joue notamment un rôle prépondérant dans la distribution spatiale des propriétés biogéochimiques à l’origine de la variabilité des concentrations en chlorophylle (Killworth et al., 2004). À l’équateur, l’advection verticale favorise la remontée de nutriments vers la surface. La pertinence de la formulation TVD en tant que schéma d’advection pour la biogéochimie a été démontrée dans Dufour et al. (2013).

Conditions limites et forçages

-Dans cette étude, les frontières latérales du modèle seront des frontières fermées. On distingue des frontières réelles, d’interface entre l’océan et les surfaces continentales (zones côtières) et des frontières “virtuelles”, d’interface océan/océan, délimitant le domaine. Ces limites sont soumises à la condition « free slip » de glissement libre sans frottement. À ces interfaces, la vitesse normale est nulle. La vitesse tangentielle est égale à la vitesse « off shore » la plus proche, de sorte que la dérivée tangentielle de la vitesse et la vorticit é s’annulent. Au niveau des frontières terrestres, un apport d’eau douce par les rivières (runoffs) est introduit par l’intermédiaire de fichiers de forçage. En plein océan, les frontières fermées possèdent une zone de relaxation (tampon) dans laquelle les grandeurs physique T et S subissent un rappel climatologique (données T_{clim} et S_{clim} , voir section 3.3, table 3.6).

-En surface, pour le modèle OPA, les conditions atmosphériques sont spécifiées par des fichiers de forçage qui seront explicités en section 3.3.2. En plus de la condition limite de surface libre, on utilise une formulation « bulk » pour calculer les échanges de flux de quantité de mouvement, de chaleur et d'eau douce à l'interface air/mer. Le terme « bulk » exprime l'estimation de quantités locales à partir de quantités globales. Un rappel en sel (sea surface restoring) est également appliqué pour assurer la conservation de la masse.

-Au fond de l'océan, les vitesses normales sont considérées nulles. Aucun flux de chaleur géothermique n'est considéré. Des échanges de quantité de mouvement sont pris en compte par un terme de frottement non linéaire exprimé sous forme quadratique.

3.2.2 Le modèle biogéochimique PISCES

Le modèle de circulation océanique OPA présenté dans la partie précédente est couplé avec le modèle de biogéochimie marine PISCES-v2 (Aumont et al., 2015). Son développement a été initié dans les années 2000, à partir du modèle biogéochimique HAMOCC⁵ (Six and Maier-Reimer, 1996), dans le but de décrire les principaux cycles biogéochimiques du carbone de l'océan dans le contexte de modification climatique (Aumont et al., 2003). Il a depuis été utilisé pour de nombreuses études (p.ex. Gehlen et al., 2006; Rodgers et al., 2008; Brasseur et al., 2009; Steinacher et al., 2010) et fait partie des modèles du CMIP5⁶ utilisés pour les activités du Groupe d'Experts intergouvernemental du Climat (GIEC).

3.2.2.1 Complexité biogéochimique

Le modèle PISCES a été élaboré afin de posséder un niveau de complexité lui permettant de simuler les comportements biologiques de grande échelle des premiers niveaux trophiques (plancton) des écosystèmes marins (p.ex. Bopp et al., 2005). L'une de ses particularités est qu'il permet de s'affranchir de l'utilisation permanente de rapports stœchiométriques constants en conciliant l'approche dite des modèles de Monod (Monod, 1942) et celle dite des quotas (Droop, 1983). Cette particularité sera précisée dans la suite. Son architecture s'appuie sur 24 variables interdépendantes dont les principales interactions sont représentées schématiquement en figure 3.4.

Nous pouvons distinguer 4 grands compartiments :

-Le compartiment **PHY** regroupe les organismes planctoniques végétaux photosynthétiques (phytoplancton). Ces organismes sont autotrophes, c'est-à-dire qu'ils synthétisent leur matière organique (carbone) à partir d'énergie lumineuse et d'éléments nutritifs puisés dans le milieu (compartiment NUT). Leur habitat se limite donc à la couche euphotique de l'océan, qui s'étend de la surface jusqu'à la profondeur à laquelle l'intensité lumineuse correspond à 1% de celle à l'interface air/mer. La classification du phytoplancton est faite en fonction de la taille des individus. PISCES distingue 2 différentes classes, illustrées en image par les figures 3.5 (a) et (b) :

5. HAMburg model of Oceanic Carbon Cycle

6. Coupled model Intercomparison Project Phase 5

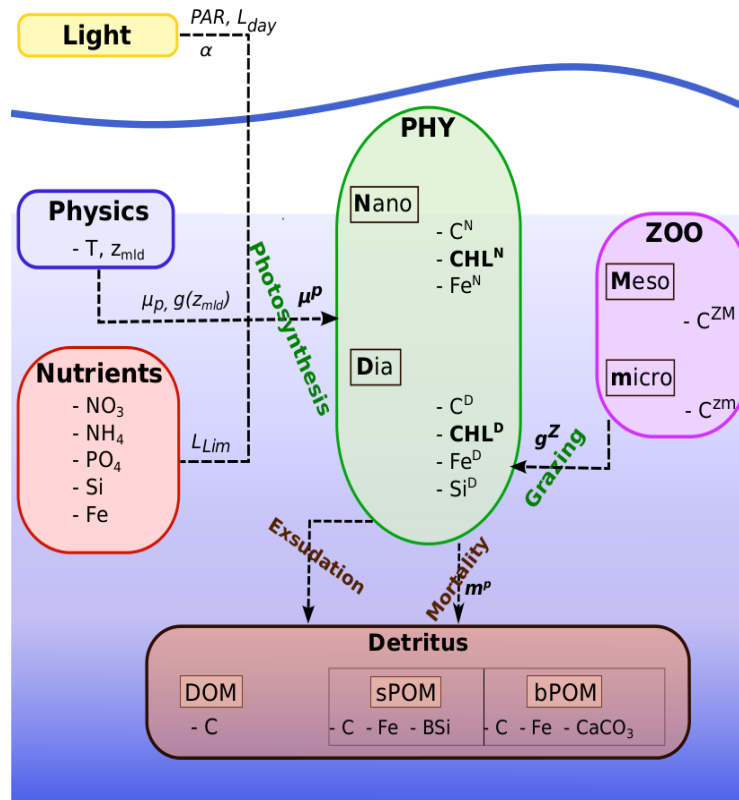


FIGURE 3.4 – Représentation schématique des 4 principaux compartiments de PISCES. À l'exception du carbone inorganique, de l'alcalinité et de l'oxygène, toutes les variables du modèle sont indiquées. Pour conserver une représentation simple, seules les interactions avec le compartiment **PHY** sont représentées. Quelques-uns des paramètres clés sont également indiqués.

- le nanophytoplancton (N), organismes de petite taille (jusqu'à une vingtaine de microns).
- les diatomées (D), organismes dont la taille peut aller jusqu'au millimètre.

Outre la taille des individus, la différence majeure entre ces 2 groupes réside dans la présence d'un squelette externe silicieux (frustule) chez les diatomées. Le modèle caractérise ces 2 classes d'espèces par leur concentration interne en carbone (C^N et C^D), fer (Fe^N et Fe^D) et chlorophylle (CHL^N et CHL^D), auxquelles s'ajoute, uniquement pour les diatomées, la concentration en silice (Si^D). Les rapports Fe/C , CHL/C et Si/C sont donc entièrement prédits par le modèle (approche des quotas) alors que les autres rapports $C/N/P$ sont supposés constants en suivant les valeurs de Takashi et al. (1985).

-Le compartiment **ZOO** rassemble les organismes planctoniques animaux (zooplancton). Contrairement au plancton végétal, le zooplancton est hétérotrophe. Il synthétise sa matière organique en consommant (broutage) le phytoplancton, les débris organiques et le zooplancton de plus petite taille. Comme pour le compartiment **PHY**, la classification du zooplancton est faite en fonction de la taille des individus. PISCES distingue 2 classes illustrées en image en figures 3.5 (c) et (d) :

- le microzooplancton (Zm), organismes de petite taille (jusqu'à environ $200 \mu m$).
- le mesozoplancton (ZM), organismes dont la taille peut aller jusqu'à plusieurs millimètres.

Dans les équations du modèle, ces 2 groupes de zooplancton se différencient principalement par leur capacité de broutage. Le mesozoplancton peut en effet consommer du phytoplancton, du microzooplancton et toutes les tailles de matériel particulaire alors que le microzooplancton ne peut se nourrir que de phytoplancton et de petites particules organiques.

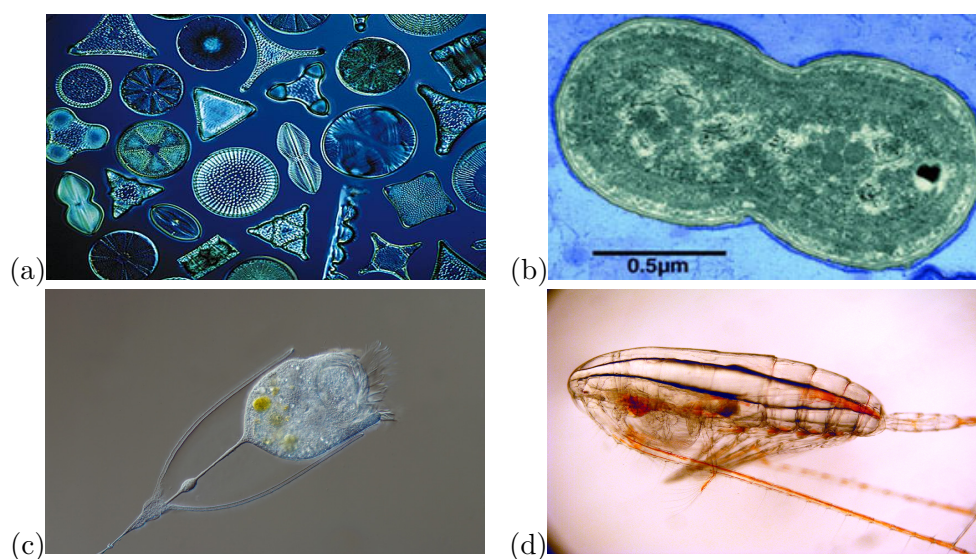


FIGURE 3.5 – (a) diatomées. (b) synecococcus (espèce de nanophytoplancton) source John B Waterbury <http://genome.jgi.doe.gov>, (c) microzooplancton nauplii (d) mesozoplancton calanidus source <http://arthropoda.southernfriedscience.com>.

-Le Compartiment **NUT** est constitué de 5 nutriments limitant la croissance du phytoplancton : les nitrates (NO_3), l'ammonium (NH_4), le phosphate (PO_4), le fer (Fe) et le silicate (Si). Les constituants azotés (nitrates et ammonium) et le phosphate ne sont pas totalement indépendants et sont reliés via les rapports stœchiométriques de Redfield. L'activité microbienne n'est pas explicitement simulée par le modèle mais les processus de conversion de l'ammonium en nitrate (nitrification) et de reminéralisation de la matière organique (dénitrification) sont toutefois paramétrés. Les apports en nutriments par les rivières, les dépôts des poussières atmosphériques et la sédimentation de fond, sont retranscrits dans le modèle via des fichiers de forçage extérieur (cf. section 3.3.2).

-Le compartiment **DETRITUS** regroupe l'ensemble de la matière détritique. Il est constitué de la matière organique dissoute (DOM) et de la matière organique particulaire (POM). La matière organique particulaire est ensuite subdivisée en 2 classes en fonction de la taille des particules : les petites particules dont la taille est inférieure à $100 \mu m$ ($sPOM$) et les grosses particules ($bPOM$) dont la taille est comprise entre 100 et $500 \mu m$. Ces 2 types de particules se distinguent principalement par leur vitesse verticale de chute

(export vers l'océan profond). Dans PISCES, le matériel dissous est uniquement constitué de carbone. Le matériel particulaire détritique est caractérisé par une fraction organique constituée de carbone (C) et une fraction inorganique représentée par les concentrations en silicate ($B*Si*$), fer (Fe) et calcite ($CaCO_3$). Les autres rapports $C/N/P$ sont imposés par les rapports de Redfield.

En plus de ces 21 variables, reliées à la représentation des écosystèmes marins, le modèle PISCES inclut 3 variables supplémentaires en interaction avec l'atmosphère.

-Le carbone inorganique dissous (DIC) et l'alcalinité (Alk). Ces variables définissent le système des carbonates (équilibre chimique entre le CO_2 , le HCO_3^- et les ions carbonates CO_3^{2-}) dans lequel la biogéochimie intervient par la capacité de certains organismes marins à utiliser du carbone pour la construction de carapaces calcaires (principalement $CaCO_3$). Potentiellement, elles permettent de quantifier le degré d'acidification de l'océan, qui est un des impacts environnementaux majeurs issu de la modification du climat.

-L'oxygène (O_2). La présence de l' O_2 est indispensable au couplage avec les niveaux trophiques océaniques supérieurs (p.ex. couplage au modèle de pêche SEAPODYM, Lehodey et al., 2008) et aux études biogéochimiques environnementales en région côtière (détection de zones anoxiques).

3.2.2.2 Équations de la production primaire

Dans cette thèse, on se focalise sur la capacité de PISCES à estimer la production de matière organique issue de la photosynthèse : la production primaire. Elle est évaluée en termes de concentration totale en chlorophylle (somme des variables CHL^N et CHL^D). Par conséquent, seules les formulations du modèle caractérisant les interactions avec le compartiment PHY (cf. figure 3.4) seront explicitées dans cette section. Une description des autres équations pourra être trouvée dans la littérature (Aumont et al., 2003, 2015). Afin de faciliter la lecture des équations, les notations utilisées sont précisées en table 3.3. La valeur ainsi que le descriptif des paramètres biogéochimiques intervenant dans les formulations du modèle sont récapitulés en table 5.1. Pour une description approfondie des paramètres du modèle, le lecteur peut également se référer à l'annexe B.

Biomasse du phytoplancton

Au premier ordre, la production de la biomasse du phytoplancton est gouvernée par l'équation 3.10 qui traduit l'évolution de leur concentration en carbone C^P . Elle traduit l'équilibre entre la croissance et la mortalité du phytoplancton et le broutage par le zooplancton. L'exposant P fait référence au nanophytoplancton (N) et aux diatomées (D).

$$\frac{\partial C^P}{\partial t} = \overbrace{(1 - \delta^P)\mu^P C^P}^{\text{croissance}} - \overbrace{m^P \left(\frac{C^P}{K_m + C^P} \right) C^P - sh \times w^P (C^P)^2}^{\text{mortalité}} - \overbrace{\sum_Z g^Z(P) C^Z}^{\text{broutage}} \quad (3.10)$$

Notation	Description	Notation	Description
N	nanophytoplancton	D	diatomées
P	phytoplancton, $P = (N, D)$	C^P	carbone du phytoplancton
L_{day}	durée du jour	CHL^P	chlorophylle du phytoplancton
T	température	sh	taux de cisaillement
K_{nut}^P	paramètre de demi saturation de limitation des nutriments	ρ^P	rapport photosynthétique énergie assimilée/énergie absorbée
$\theta_{min}^{Fe,P}$	rapport minimum Fer/C	L_{Lim}^P	limitation par les nutriments
Zm	microzooplancton	ZM	mesozooplancton
Z	zooplancton, $Z = (Zm, ZM)$	C^Z	carbone du zooplancton
z_{eu}	profondeur de la couche euphotique	z_{mld}	profondeur de la couche de mélange
I_{Zm}	éléments nutritifs du Zm $I_{Zm} = (P, sPOC)$	I_{ZM}	éléments nutritifs du ZM $I_{ZM} = (P, Zm, sPOC, bPOC)$
w^P	mortalité quadratique	F	seuil de ressources nutritives minimum

TABLE 3.3 – Descriptif des paramètres biogéochimiques qui interviennent dans les équations du modèle PISCES présentées dans la section 3.2.2.2.

Terme de croissance

Dans PISCES, la croissance du phytoplancton est régie par le calcul d'un taux de croissance μ^P caractérisé par l'équation 3.11 ci-dessous.

$$\mu^P = \mu_P f(L_{day}) g(z_{mld}) \left[1 - \exp \left(\frac{-\alpha^P \left(\frac{CHL^P}{C^P} \right) PAR^P}{L_{day} (\mu_{ref} + b_{resp})} \right) \right] L_{Lim}^P \quad (3.11)$$

Comparée à la première version de PISCES (Aumont et al., 2003), cette formulation se base sur une amélioration de PISCES-v2 (option *ln_newprod*). Elle ajoute un facteur de respiration b_{resp} , attestant d'une perte de carbone due à une respiration minimale nécessaire au maintien des fonctions vitales en l'absence de photosynthèse (respiration basale). La formulation du taux de croissance contient une fonction d'atténuation $g(z_{mld})$ qui témoigne via l'équation 3.13 de la dépendance du phytoplancton au rapport de profondeur entre la couche de mélange Z_{mld} et la couche euphotique Z_{eu} (Lévy et al., 1998). L'impact de la durée du jour est caractérisé par la fonction $f(L_{day})$ (équation 3.12). L'intensité lumineuse disponible pour la photosynthèse PAR (Photosynthetic Available Radiation) est calculée à partir d'une version simplifiée du modèle à 3 bandes (bleu, vert, rouge) proposé par Morel (1988) (équation 3.15). Le terme L_{Lim}^P est un facteur de limitation par les nutriments défini en fonction de l'élément disponible en quantité minimale (équation 3.16). Enfin, la dépendance de la croissance à la température T est définie par μ_P en suivant le modèle d'Eppley (Eppley, 1972, équation 3.17). Nous pouvons remarquer qu'elle est identique pour les 2 classes de phytoplancton.

$$f(L_{day}) = f_{day}^P \left(\frac{L_{day}}{0.5 + L_{day}} \right) \quad (3.12)$$

$$g(Z_{mld}) = 1 - \frac{T_{dark}}{T_{dark} + T_{dark}^P} \quad (3.13)$$

$$T_{dark} = [max(0, Z_{mld} - Z_{eu})]^2 / 86400 \quad (3.14)$$

$$PAR^P(z) = \beta_1^P PAR_1(z) + \beta_2^P PAR_2(z) + \beta_3^P PAR_3(z) \quad (3.15)$$

$$L_{Lim}^P = min(L_{PO_4}^P, (L_{NO_3}^P + L_{NH_4}^P), L_{Fe}^P, L_{Si}^D) \quad (3.16)$$

$$\mu_P = \mu_{max}^0 \exp(bT) \quad (3.17)$$

La limitation de la croissance du phytoplancton par les différents nutriments est explicitée par les équations 3.18 à 3.22. On constate que l'expression pour la limitation par le fer est différente. Le modèle PISCES suit en effet, pour le fer, une approche des quotas (Droop, 1983), basée sur un quota optimal $\theta_{opt}^{Fe,P}$ (Buitenhuis and Geider, 2010) et sur une demande minimale $\theta_{min}^{Fe,P}$ (Flynn and Hipkin, 1999) nécessaire aux activités vitales. Pour exprimer l'existence de différents comportements biologiques à partir de 2 classes de phytoplancton, les paramètres de demi saturation K_{NUT}^P pour la limitation par les différents nutriments, sont définis par le modèle. Ils sont calculés selon la taille des espèces prédominantes du milieu, les espèces de plus grande taille possédant un rapport surface/volume plus faible.

$$L_{PO_4}^P = \frac{PO_4}{PO_4 + K_{PO_4}^P} \quad (3.18)$$

$$L_{NH_4}^P = \frac{K_{NO_3}^P NH_4}{K_{NH_4}^P K_{NO_3}^P + K_{NH_4}^P NO_3 + K_{NO_3}^P NH_4} \quad (3.19)$$

$$L_{NO_3}^P = \frac{K_{NH_4}^P PO_3}{K_{NH_4}^P K_{NO_3}^P + K_{NH_4}^P NO_3 + K_{NO_3}^P NH_4} \quad (3.20)$$

$$L_{Si}^D = \frac{Si}{Si + K_{Si}^D} \quad (3.21)$$

$$L_{Fe}^P = min(1, max(0, \frac{Fe^P}{C^P} - \theta_{min}^{Fe,P} / \theta_{opt}^{Fe,P})) \quad (3.22)$$

Terme de broutage

Le broutage du phytoplancton par le zooplancton est défini par l'équation 3.23. Il est caractérisé par un terme de limitation (*préférence*) qui exprime la préférence du zooplancton aux différents éléments nutritifs. Ce terme est construit de manière analogue à celui de la limitation de la croissance du phytoplancton par les nutriments (sous l'approche de Monod). L'indice I fait référence aux différents éléments nutritifs dont la classe de zooplancton considérée (Zm ou ZM) peut se nourrir (cf. section 3.2.2.1). Le terme I_{thresh}^Z définit un seuil de concentration minimale en dessous duquel les éléments nutritifs ne sont plus consommés. Similairement, le terme de limitation F représente le seuil de ressources

nutritives minimum, calculé par le modèle, en dessous duquel il n'y a plus de broutage. La dépendance du broutage à la température g_m^Z est caractérisée par l'équation 3.24, selon un modèle équivalent au modèle d'Eppley pour la croissance du phytoplancton (équation 3.17). Contrairement au terme μ_{max}^0 , la dépendance du broutage à la température $g_{max}^{0,Z}$ peut cette fois être différente pour les 2 classes de zooplancton.

$$g^Z(I) = g_m^Z \overbrace{\left(\frac{p_J^I \max(0, C^I - I_{thresh}^Z)}{K_G^Z + \sum_I (p_I^Z C^I)} \right)}^{\text{préférence}} F \quad (3.23)$$

$$g_m^Z = g_{max}^{0,Z} \exp(b'T) \quad (3.24)$$

Terme de mortalité

La mortalité (cf. équation 3.10) se compose d'un 1^{er} terme dit de mortalité naturelle. Il est calculé à partir d'un coefficient de mortalité m^P et d'une constante de demi saturation K_m . Le second terme est une mortalité quadratique relative à l'aggrégation et à la sédimentation de certaines cellules, causées par la turbulence. Elle est définie par un taux de cisaillement sh renforcé dans la couche de mélange (équation 3.25) et par un coefficient d'aggrégation w^P . L'expression de ce coefficient (équation 3.26) permet de prendre en compte l'augmentation de l'aggrégation des diatomées, en cas de limitation par les nutriments, due à une sécrétion de mucus.

$$sh = \begin{cases} 1.0 & \text{si } z \leq z_{mld} \\ 0.01 & \text{si } z > z_{mld} \end{cases} \quad (3.25)$$

$$w^P = \begin{cases} w_{min}^P & \text{si } P = N \\ w_{min}^P + w_{max}^D(1 - L_{lim}^D) & \text{si } P = D \end{cases} \quad (3.26)$$

Chlorophylle du phytoplancton

La quantité de chlorophylle du phytoplancton est calculée selon l'équation 3.27. Excepté pour l'expression du terme de croissance, elle est logiquement très similaire à l'équation 3.10 de la production de biomasse du phytoplancton.

$$\begin{aligned} \frac{\partial CHL^P}{\partial t} = & \overbrace{\left((1 - \delta^P) \mu^P C^P (12\theta_{min}^{chl} + (\theta_{max}^{chl,P} - \theta_{min}^{chl}) \rho^P) \right)}^{\text{Croissance}} - \overbrace{\sum_Z \frac{CHL^P}{C^P} g^Z(P) C^Z}^{\text{Broutage}} \\ & - \underbrace{m^P \left(\frac{C^P}{K_m + C^P} \right) CHL^P - sh \times w^P C^P CHL^P}_{\text{Mortalité}} \end{aligned} \quad (3.27)$$

Paramètre	Description	Valeur
Phytoplancton		
μ_{max}^0	taux de croissance maximum à 0°C	0.6 j ⁻¹
b	sensibilité à la température	0.064
$\theta_{opt}^{Fe,P}$	quota optimal de Fer	7 $\mu\text{molFe/molC}$
μ_{ref}	taux de croissance de référence	1.0 j ⁻¹
b_{resp}	respiration basale	0.033 j ⁻¹
δ^P	exsudation de carbone organique dissous	0.05
m^P	taux de mortalité	0.01 j ⁻¹
K_m	constante de 1/2 saturation de mortalité	0.2 $\mu\text{molC}^{-1}\text{L}^{-1}$
w_{max}^D	aggrégation maximale des diatomées	0.03 $\text{molC}^{-1}\text{j}^{-1}$
w_{min}^P	aggrégation minimale	0.01 $\text{molC}^{-1}\text{j}^{-1}$
α^P	efficacité photosynthétique (pislope)	2 $\text{Wm}^{-2}\text{j}^{-1}$
f_{day}^P	facteur de dépendance à la durée du jour	1.5
T_{dark}^P	temps de résidence moyen dans la MLD	(N; D)=(3;4) j
β_i^P	coefficient d'absorption dans le spectre bleu	(N; D)= (2.1;1.6)
β_2^P	coefficient d'absorption dans le spectre vert	(N; D)= (0.42;0.69)
β_3^P	coefficient d'absorption dans le spectre rouge	(N; D)= (0.4;0.7)
$\theta_{max}^{Chl,P}$	rapport chlorophylle sur carbone maximal	(N; D)=(0.033;0.05) mgChl/mgC
θ_{min}^{Chl}	rapport chlorophylle sur carbone minimal	0.0033 mgChl/mgC
Zooplancton		
b'_Z	sensibilité du broutage à la température	0.076
$g_{max}^{0,Z}$	taux de broutage maximum à 0°C	(Zm; ZM)=(3;0.7) j ⁻¹
p_N^Z	préférence pour le nanophytoplancton	(Zm; ZM)=(1.0;0.3)
p_D^Z	préférence pour les diatomées	(Zm; ZM)=(0.5;1.0)
p_{POC}^Z	préférence du zooplancton pour le POC	(Zm; ZM)=(0.1;0.3)
p_{Zm}^{ZM}	préférence pour le microzooplancton	1.0
K_G^Z	constante de 1/2 saturation du broutage	20.0 $\mu\text{molC.L}^{-1}$
I_{thresh}^Z	seuil minimum de nourriture	0.001 $\mu\text{molC.L}^{-1}$

TABLE 3.4 – Valeurs des paramètres biogéochimiques utilisés dans les équations de la section 3.2.2.2. Pour une description approfondie de la valeur de l'ensemble des paramètres du modèle PISCES, le lecteur peut se référer à la « namelist » présentée en annexe B.

La quantité de chlorophylle issue de la croissance du phytoplancton est paramétrée à partir du modèle de Geider (Geider et al., 1997) selon l’équation 3.28. Tel que défini, le rapport CHL/C peut ainsi varier spatio-temporellement en fonction de la durée du jour et de l’intensité lumineuse. Ce point est un facteur clé pour notre étude qui vise à représenter la distribution de la chlorophylle dans un domaine aux conditions d’ensoleillement particulièrement variées. On remarquera que l’expression de la croissance du phytoplancton μ^{IP} à partir de laquelle est déduite ρ^P n’est pas exactement identique à celle de μ^P utilisée dans l’équation 3.27 et le calcul de la biomasse du phytoplancton.

$$\rho^P = \frac{144\mu^{IP}C^P L_{day}}{\alpha^P CHL^P PAR^P} \quad (3.28)$$

$$\mu^{IP} = \mu_P g(z_{mld}) \left[1 - \exp\left(\frac{-\alpha^P (\frac{CHL^P}{C^P}) PAR^P}{L_{day}(\mu_P L_{Lim}^P)}\right) \right] L_{Lim}^P \quad (3.29)$$

3.3 Mise en place du modèle couplé au 1/4° sur l’Atlantique Nord

Dans cette section, nous présentons les caractéristiques du couplage entre le modèle de circulation océanique OPA et le modèle biogéochimique PISCES. Afin de s’assurer d’une représentation physique réaliste, le modèle PISCES a été greffé sur la configuration de l’Atlantique Nord, NATL025, dont nous détaillons ici les principaux paramètres. Je précise que cette configuration a servi de base aux travaux du « large Benchmark » du projet SANGOMA, dont cette thèse fait partie. Toutes les informations sur les valeurs des paramètres du modèle couplé sont présentées dans les annexes.

3.3.1 Caractéristiques du couplage

Le couplage entre le modèle de circulation océanique OPA et le modèle PISCES a été réalisé par l’intermédiaire du module TOP de NEMO. Il possède la particularité d’être effectué en direct ou « on-line », ce qui signifie que les variables biogéochimiques et physiques sont calculées conjointement, à chaque pas de temps. L’intérêt est d’éviter les approximations dues à l’utilisation de champs physique moyens. Les travaux de cette thèse figurent ainsi, avec ceux de Resplandy et al. (2012) sur la mer d’Arabie et de Dufour et al. (2013) dans l’océan Austral, parmi les premiers à proposer des simulations couplées « on-line » du modèle PISCES à une résolution « eddy-permitting ».

Dans ce couplage, aucune rétroaction sur la physique n’est considérée. Les variables biogéochimiques sont considérées comme des traceurs passifs dont l’évolution est déterminée par l’équation 3.30 d’advection-diffusion.

$$\frac{\partial C}{\partial t} = \overbrace{-\nabla \cdot (uC)}^A - \overbrace{A_C^h \nabla_h^2 C}^{D^h} + \overbrace{\frac{\partial}{\partial z} (A_C^v \frac{\partial C}{\partial z})}^{D^v} + SMS(C) \quad (3.30)$$

Le vecteur d'état C inclut les 24 variables du modèle PISCES. Elles sont calculées par le terme SMS (sources moins puits) selon une discrétisation temporelle par schéma explicite arrière. Les termes d'advection A , de diffusion horizontale D^h et de diffusion verticale D^v sont calculés par le modèle physique. La table 3.5 récapitule les principales caractéristiques du couplage. Pour plus de détails concernant les caractéristiques numériques du modèle, le lecteur peut se référer aux « namelists » des annexes A et C. Les comparaisons entre simulations présentent, en effet, régulièrement des écarts dus à seulement quelques différences dans leurs paramètres. Présenter ces « namelists » permet ainsi d'assurer la reproductibilité des résultats et d'identifier les éventuelles différences entre les configurations. Contrairement aux traceurs T et S, qui suivent un schéma d'advection TVD (voir section 3.2.1), l'advection des traceurs biogéochimiques est paramétrée selon le schéma MUSCL (Monotone Upstream Scheme for Conservative Laws, LeVeque, 2002). Cette choix provient de l'incompatibilité du schéma d'advection TVD avec une discrétisation temporelle décentrée. Les paramétrisations des termes de diffusion sont identiques à celles des traceurs passifs (T et S) décrites en section 3.2.1.

3.3.2 Configuration du modèle

Le modèle couplé au $1/4^\circ$ que nous utilisons est hérité de la configuration physique NATL025 mise en place dans le cadre du projet DRAKKAR (cf. section précédente). Cette configuration est issue d'un découpage de la configuration globale ORCA025 (Barnier et al., 2006) adapté pour l'Atlantique Nord. Elle repose sur une grille non régulière obtenue par projection MERCATOR. La zone d'étude s'étend de 20°S à 80°N en latitude (530 points de grille) et de 23°E à 98°W en longitude (486 points de grille). À noter que le prolongement de l'Atlantique Nord jusqu'à 20°S s'explique par la nécessité d'effectuer le découpage dans une zone où la dynamique de l'océan est minimale. Sur la verticale, la résolution est de 46 niveaux (cf. section 3.2.1). La résolution temporelle du modèle est de un pas de temps toutes les 40 minutes, ce qui correspond à 36 itérations journalières. Afin de s'accorder avec les travaux de Fontana et al. (2013) et de disposer des observations SeaWIFS, la période d'étude qui nous intéressera est la période 2005-2010.

Paramètres principaux du couplage NEMO/PISCES	
Schéma d'advection T,S	TVD
Schéma d'advection C	MUSCL
Diffusivité verticale turbulente	$A_X^v = 1.10^{-5} \text{ m}^2/\text{s}$ avec $X = T, S, C$
Viscosité verticale turbulente	$A_{\vec{U}}^v = 1.10^{-4} \text{ m}^2/\text{s}$
Diffusion horizontale laplacienne	$A_{T,S}^h = 300 \text{ m}^2/\text{s}$ et $A_C^h = 520 \text{ m}^2/\text{s}$
Diffusion horizontale bilaplacienne	$A_{\vec{U}}^h = 1.5.10^{-11} \text{ m}^2/\text{s}$

TABLE 3.5 – Principales caractéristiques du couplage. Les notations sont identiques à celles utilisées pour les équations 3.7 à 3.9 de la section 3.2.1 et à celles de l'équation 3.30. L'indice C fait référence aux traceurs biogéochimiques du modèle PISCES. T , S et \vec{U} représentent respectivement la température, la salinité et le vecteur vitesse.

Initialisation

Afin d'assurer un état initial réaliste, les variables physiques et biogéochimiques du modèle couplé sont initialisées au 1^{er} janvier 2002 à partir de simulations préalablement validées. Une période d'initialisation ou « spin-up de contrôle », d'une durée de 3 ans, est ensuite réalisé entre janvier 2002 et décembre 2004 afin de garantir la stabilité du système.

- Variables biogéochimiques : les 24 variables biogéochimiques sont initialisées à partir des sorties de la simulation globale couplée au 1/4° BIOMER du système de modélisation de MERCATOR-Océan. Cette simulation couplée NEMO/PISCES est basée sur la même grille horizontale (ORCA025). La grille verticale comporte, par contre, 56 niveaux. Une interpolation linéaire a ainsi dû être effectuée pour adapter les résultats de cette simulation aux 46 niveaux de la grille verticale utilisée ici. La simulation BIOMER a été réalisée avec un couplage indirect, ou « off-line », et n'inclut pas d'assimilation de données. Tous les détails sur cette simulation sont disponibles sur le site de MERCATOR-Océan : <http://www.mercator-ocean.fr>.
- Variables physiques : les variables physiques utilisées pour l'initialisation proviennent d'une simulation de 13 ans (1989-2002) du modèle OPA, dont les variables T et S ont été initialisées à partir de la climatologie Levitus (Levitus et al., 1998). Cette simulation repose sur la même configuration physique (NATL025). Ses résultats ont été validés par l'étude de Candille et al. (2015).

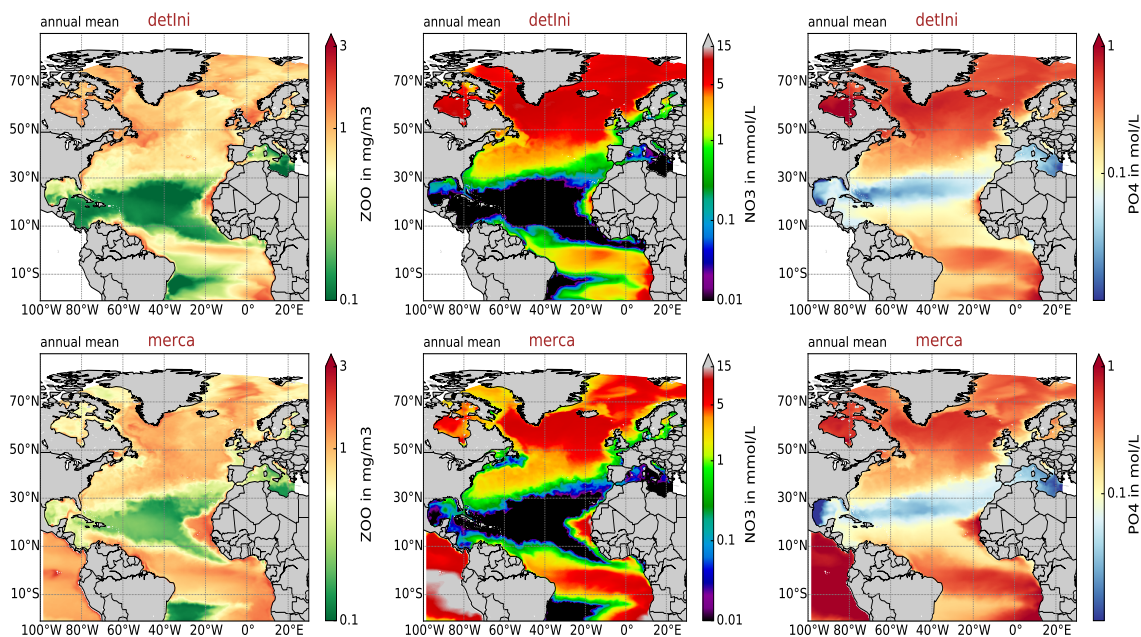


FIGURE 3.6 – Cartes de surface de la simulation déterministe *detIni*, présentée dans le chapitre suivant (ligne du haut) et de la simulation Mercator BIOMER, utilisée pour initialiser les variables de PISCES (ligne du bas). Les cartes affichent les moyennes annuelles de l'année 2005 des concentrations en microzooplancton (1^{ère} colonne), nitrates (2^{ème} colonne) et phosphates (3^{ème} colonne).

La figure 3.6 compare l'état biogéochimique du modèle couplé NATL025/PISCES à celui du modèle couplé BIOMER de MERCATOR-Océan en 2005, juste après les 3 années du spin-up « de contrôle ». La correspondance entre les 2 représentations prouve que les écarts entre les 2 configurations ont peu d'impacts sur la biogéochimie. Ce résultat justifie donc l'utilisation des sorties du modèle couplé BIOMER pour initialiser les variables biogéochimiques. Je précise que les légers écarts entre les 2 simulations peuvent être mis sur le compte des effets « off-line » versus « online », de la discrétisation verticale ou encore de différences entre les « namelists ».

Conditions limites et forçages

Les conditions aux limites sont imposées par le modèle physique (cf. section 3.2.1). Nous présentons ici les différentes données de forçage utilisées dans la configuration couplée NATL025/PISCES. Un récapitulatif est donné en table 3.6.

- Forçages atmosphériques :

Les champs atmosphériques sont utilisés comme condition limite de surface à l'interface air/mer. Ils proviennent des réanalyses atmosphériques⁷ ERA-Interim (Berrisford et al., 2011) interpolées sur la grille NATL025. Via les équations 3.31 à 3.33, les champs atmosphériques fournissent les conditions aux limites des équations de conservation de la salinité, de la température et de la quantité de mouvement (cf. précédente, équations 3.4, 3.5 et équations 3.7 à 3.9).

$$A_{\vec{u}_h}^v \frac{\partial \vec{u}_h}{\partial z} \Big|_{z=0} = \frac{\vec{\tau}}{\rho_s} \quad (3.31)$$

$$A_S^v \frac{\partial S}{\partial z} \Big|_{z=0} = - \frac{F_w S_{z=0}}{\rho_s} \quad (3.32)$$

$$A_T^v \frac{\partial T}{\partial z} \Big|_{z=0} = \frac{Q_{net}}{\rho_s C_{ps}} \quad (3.33)$$

Les termes $S_{z=0}$ et ρ_s représentent respectivement la salinité et la densité à la surface de l'océan. C_{ps} est la capacité calorifique de l'eau de mer. Les termes $\vec{\tau}$, Q_{net} et F_w sont définis à partir des variables prescrites par les fichiers de forçage.

La tension exercée par le vent $\vec{\tau}$ à la surface de l'océan est calculée à partir des composantes u et v du vent à 10 mètres \vec{u}_{atm} , selon l'équation 3.34. Le terme de vent relatif \vec{u}_{rel} , est obtenu en retranchant la contribution des courants de surface de l'océan \vec{u}_s , de sorte que $\vec{u}_{rel} = \vec{u}_{atm} - \vec{u}_s$.

Les flux d'eau douce à l'interface air/mer F_w , sont définis par l'équation 3.35. Ils sont calculés à partir de l'équation de la surface libre et de la somme entre les apports par les rivières R_{ofs} (uniquement appliqués en surface) et le flux net de précipitation moins

7. Reconstruction de l'état de l'atmosphère sur une longue période à partir d'un modèle de circulation dont les sorties ont été corrigées par assimilation de données

évaporation. Un rappel en salinité S_{clim} est également introduit pour limiter la dérive en salinité du modèle, induite par les incertitudes sur les flux de précipitation à la surface de l'océan. Le terme q_{atm} représente l'humidité spécifique de l'air et le terme $q_{sat}(T_s)$ l'humidité spécifique de l'air à saturation pour la température de surface T_s , calculée par le modèle.

Les flux d'échange de chaleur entre l'océan et l'atmosphère Q_{net} , sont définis par l'équation 3.36. Ils sont calculés comme la somme des flux de chaleur radiatifs (grandes et courtes longueurs d'onde) et des flux de chaleur turbulents. Les flux de chaleur radiatifs sont caractérisés à partir des données de forçages extérieurs rad_{sw} et rad_{lw} , de l'albédo α et de la constante de stephan-Boltzmann σ . Le flux de chaleur latente est le résultat de l'évaporation E et le flux de chaleur sensible de la température de l'air T_{atm} . Le terme L_{vap} représente la chaleur latente d'évaporation.

$$\vec{\tau} = \rho_{atm} C_D \|\vec{u}_{rel}\| \vec{u}_{rel} \quad (3.34)$$

$$F_w = P + R_{offs} + \overbrace{\rho_{atm} C_E \|\vec{u}_{atm}\| (q_{atm} - q_{sat}(T_s))}^{\text{Évaporation } (= -E)} - \overbrace{S_{clim}}^{\text{damping}} \quad (3.35)$$

$$Q_{net} = \overbrace{(1 - \alpha) rad_{sw}}^{Q_{sw}} + \overbrace{rad_{lw} - \sigma T_s^4}^{Q_{lw}} - \overbrace{\frac{Q_{latente}}{L_{vap} E}}^{Q_{latente}} + \overbrace{\rho_{atm} C_{pa} C_H \|\vec{u}_{atm}\| (T_{atm} - T_s)}^{Q_{sensible}} \quad (3.36)$$

Dans ces équations, ρ_{atm} représente la densité de l'air et C_{pa} sa capacité calorifique. Les termes C_D , C_E et C_H sont des coefficients de transfert turbulent, évalués à partir des champs de surface fournis par les forçages et le modèle. Ils sont calculés en suivant la formulation de Large and Yeager (2004).

-Forçages biogéochimiques :

En plus des forçages atmosphériques qui fournissent notamment les conditions limites de la diffusivité, paramètre essentiel pour la biogéochimie, les formulations de PISCES peuvent inclure des apports extérieurs en nutriments. À l'exception du fer, ces apports sont systématiquement compensés par un enfouissement sédimentaire équivalent. On distingue 3 différentes sources extérieures de nutriments :

- Les apports de carbone organique et inorganique dissous par les rivières. Ces données sont des produits des modèles d'érosion GLOBAL-NEWS2 (Mayorga et al., 2010) et GEM (Global Erosion Model, Ludwig et al. (1996)).
- Les apports de fer par sédimentation de fond. Ces apports sont le résultat de la reminéralisation du carbone organique par les bactéries. Ils dépendent de la surface topographique en contact avec l'eau de mer. Ils sont basés sur la représentation bathymétrique haute résolution ETOPO5 en considérant un flux constant de $1 \mu mol.m^{-2}.jour^{-1}$.

- Les apports en fer, en silicate et en phosphate par les poussières atmosphériques. Les données utilisées sont basées sur une climatologie du modèle de Tegen and Fung (1995). Ces 3 apports sont interdépendants et sont introduits dans le modèle à partir d'un fichier unique. La quantité de fer par les poussières atmosphériques est estimée en considérant que les aérosols sont composés à 3.5% de fer. L'apport de phosphore repose sur un contenu total en phosphore de 750 ppm (Mahowald et al., 2008). Le contenu en silicate est paramétré suivant les estimations de Moore et al. (2001).

Variable	Origine	Fréquence	Description
q_{atm}	ERA-Interim	3h	humidité spécifique de l'air à 2m
\vec{u}_{atm}	ERA-Interim	3h	composantes u,v du vent à 10m
T_{atm}	ERA-Interim	3h	température de l'air à 2m
rad_{lw}	ERA-Interim	24h	radiations solaires "long wave"
rad_{sw}	ERA-Interim	24h	radiations solaires "short wave"
P	ERA-Interim	24h	précipitations : pluie et neige
R_{offs}	DRAKKAR	1 mois	apports côtiers d'eau douce
S_{clim}, T_{clim}	Levitus	1 mois	rappel en salinité de surface
DOC	Mayorga et al. (2010)	1 an	apports des rivières en carbone organique dissous
DIC	Ludwig et al. (1996)	1 an	apports des rivières en carbone inorganique dissous
Fer_{sed}	ETOPO5	1 an	apports de fer par sédimentation
Fer_{atm}, Si, PO_4	Tegen and Fung (1995)	1 mois	apports atmosphériques en nutriments

TABLE 3.6 – Forçages utilisés dans la configuration couplée NATL025/PISCES . Les apports côtiers d'eau douce sont une compilation climatologique de données, mise au point dans le cadre du projet DRAKKAR (Bourdallé-Badie and Treguier, 2006). Les rappels en température et en salinité de surface aux interfaces sont réalisés à partir de la climatologie Levitus (Levitus et al., 1998). Le carbone organique dissous (DOC) provient du jeu de données GLOBAL-NEWS2 décrit dans (Mayorga et al., 2010). Les données en carbone inorganique dissous (DIC) sont issues du modèle de prévision Global Erosion décrit dans Ludwig et al. (1996).

3.4 Conclusion

Dans ce chapitre, nous avons présenté le modèle couplé NEMO/PISCES et les observations satellites que nous utiliserons par la suite. Nous pouvons retenir les éléments suivants :

- Cette étude s'appuie en grande partie sur les données satellites de la couleur de l'eau SeaWIFS.
- Les observations SeaWIFS seront utilisées pour vérifier la cohérence de la représentation de la chlorophylle de surface du modèle PISCES. Les observations SeaWIFS-HR seront utilisées pour l'étape d'analyse. Les données MERIS permettront de vérifier la cohérence des analyses.
- Pour garantir une représentation physique réaliste, le modèle PISCES est couplé à la configuration au $1/4^\circ$ de l'océan Atlantique Nord NATL025, mise en place dans le cadre du projet DRAKKAR.
- Le modèle PISCES possède 24 variables biogéochimiques, dont 2 classes de phytoplancton, pour estimer les distributions de la chlorophylle.

Toutes les simulations qui seront présentées ont été réalisées sur les machines de calcul parallèle de l'IDRIS (Institut du Développement et des Ressources en Informatique Scientifique). Bien que peu mise en valeur dans ce manuscrit, il est important de se rendre compte que la partie « développement technique » représente environ $1/4$ du temps de travail de cette thèse. Pour mieux le percevoir, nous pouvons conclure ce chapitre en évoquant quelques chiffres.

Environ 1000 heures de calcul, réparties sur 160 processeurs, sont nécessaires pour réaliser une année de simulation (déterministe) du modèle couplé NEMO/PISCES au $1/4^\circ$ sur l'Atlantique Nord. La simulation qui sera décrite dans le chapitre suivant est d'une durée de 6 ans. Lorsqu'il est question d'une simulation d'ensemble, le coût informatique est démultiplié en fonction du nombre de membres. Dans cette thèse, la simulation d'ensemble comportera 60 membres. Le calcul est rapide. Il nous a fallu environ 60 000 heures de calcul, réparties cette fois sur 1024 processeurs, pour réaliser cet ensemble (cf. chapitre 5). Au total, en comptabilisant les diverses expériences faites pour mettre en place les paramétrisations stochastiques et le schéma d'assimilation de données, cette thèse aura donc nécessité au minimum 100 000 heures de calcul.

Dans le chapitre suivant, nous présenterons la simulation déterministe du modèle couplé NATL025/PISCES, réalisée sur 6 ans (entre 2005 et 2010).

Chapitre 4

Une simulation déterministe du modèle couplé NEMO/PISCES dans l'océan Atlantique Nord

Sommaire

4.1	La production primaire dans l'océan Atlantique Nord	62
4.2	De LOBSTER à PISCES	65
4.2.1	Chlorophylle de surface	65
4.2.2	Distribution verticale de la chlorophylle	68
4.3	Variabilité de la simulation déterministe	71
4.3.1	Variabilité interannuelle de la chlorophylle	71
4.3.2	Comparaison avec les observations SeaWIFS	77
4.3.3	Diversité des espèces	81
4.4	Conclusion	84

Dans ce chapitre, nous présentons les résultats d'une simulation au $1/4^\circ$ du modèle couplé NEMO-PISCES, décrit au chapitre précédent, en suivant une approche classique, déterministe, de modélisation. Afin d'être en mesure d'analyser ces résultats, la première section détaille quelques-uns des principaux processus à l'origine la production primaire dans l'océan Atlantique Nord. La seconde partie de ce chapitre aura pour objectif de démontrer les dispositions du modèle à représenter les caractéristiques grande échelle de la chlorophylle, identifiables dans les images de la couleur de l'eau. Ce point permettra de justifier l'utilisation du modèle PISCES, à la place de LOBSTER, pour estimer l'état biogéochimique de l'océan à l'échelle d'un bassin. La dernière partie de ce chapitre se concentrera sur une analyse « climatologique » des distributions de la chlorophylle, basée sur les 6 années de simulation (2005-2010) du modèle PISCES. Plus particulièrement, on se focalisera sur la variabilité interannuelle du modèle et sur des comparaisons avec les observations SeaWiFS.

4.1 La production primaire dans l'océan Atlantique Nord

Dans cette première partie, nous dressons un état des lieux, non exhaustif, des principaux processus biogéochimiques qui contrôlent la production primaire dans l'océan Atlantique Nord. Cette thèse n'ayant pas pour objectif d'analyser le détail de ces processus, on se restreindra à une vision sommaire, mais non moins importante pour comprendre les résultats qui seront énoncés dans la suite.

La présentation des principales caractéristiques biogéochimiques se déclinera en suivant le découpage¹ proposé par Longhurst (1995) et Sathyendranath et al. (1995) (cf. figure 4.1 b). Cette partition de l'océan en provinces biogéochimiques a été établie à partir des images de la couleur de l'eau du capteur CZCS (voir section 3.1) et de la dynamique des masses d'eau. Elle a depuis prouvé sa pertinence en étant utilisée comme support de délimitation dans un grand nombre de travaux (p.ex. Li and Harrison, 2001; Uitz et al., 2008; Doron et al., 2013). Dans cette section, nous détaillerons la dynamique de la production primaire des 4 grands domaines écologiques que sont : les régions équatoriales (trade-winds), les régions des latitudes moyennes et subtropicales (westerlies), les régions subpolaires (polar) et les régions côtières (coastal). Dans la suite de ce manuscrit, ce sera toutefois la subdivision de l'Atlantique Nord en 18 sous-provinces (nommées provinces de Longhurst) qui sera utilisée.

Les régions équatoriales

Aux abords de l'équateur, la dynamique de l'océan est le résultat d'un rayonnement solaire maximum et du changement de signe de la force de Coriolis. À cause de la diminution de la vorticit  planétaire, l'équilibre géostrophique n'est plus prédominant. La vorticit  relative favorise une circulation océanique complexe, majoritairement régie par les variations des vents de surface et la présence d'ondes d'instabilité tropicale. L'action des alizés induit une circulation des couches de surface principalement zonale et dirigée vers l'ouest : le courant sud équatorial (SEC). L'accumulation de masses d'eau chaudes qui en

1. L'idée d'un découpage de l'océan en provinces biogéochimiques a été initiée par le programme JGOFS (Joint Global Ocean Flux Study) dans le courant des années 90.

résulte fait apparaître un fort gradient de pression (pente est-ouest de la thermocline) qui est compensé par un puissant jet de subsurface dirigé vers l’est : le sous-courant équatorial (EUC).

La divergence équatoriale du transport d’Ekman, due aux alizés, favorise l’existence d’« upwellings »² qui permet l’apport de nutriments vers la surface. L’énergie lumineuse ne limitant pas ici la croissance du phytoplancton, la distribution de la chlorophylle est en grande partie conditionnée par ces apports de nutriments (Voituriez et al., 1982; Longhurst, 1993). La production primaire de la zone équatoriale est donc relativement importante et sa variabilité est en grande partie rythmée par les anomalies d’intensité des vents de surface (Grotsky et al., 2008).

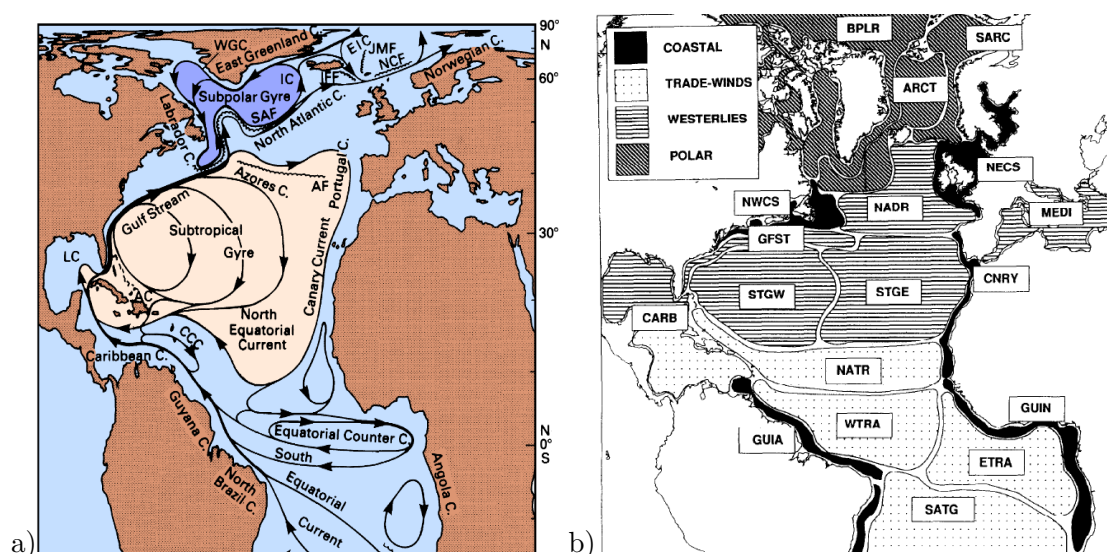


FIGURE 4.1 – a) Principaux courants océaniques de l’océan Atlantique Nord (source : Tomczak and Godfrey (2013)). b) Découpage de l’océan Atlantique Nord en provinces biogéochimiques tel que proposé par Sathyendranath et al. (1995).

Les régions des latitudes moyennes et subtropicales

La circulation océanique des latitudes moyennes et subtropicales est marquée par la présence du gyre subtropical s’étendant entre 10°N et 40°N. Le courant de bord ouest de cette circulation géostrophique anticyclonique n’est autre que le Gulf Stream. Ce puissant courant de surface se sépare au large de Terre-Neuve en 3 branches : les courants des Canaries et des Açores, qui constituent la partie Est du gyre subtropical, et la dérive Nord Atlantique qui déplace les eaux chaudes des tropiques vers les régions subpolaires. Le Gulf Stream possède la spécificité d’être le siège d’une activité tourbillonnaire turbulente particulièrement intense. Cette variabilité de méso- et de subméso-échelle a un impact fort sur le transport et le mélange des masses d’eau dont dépendent fortement les processus biogéochimiques (p.ex. Hitchcock et al., 1993; McGillicuddy et al., 2007; Williams, 2011). Il en résulte des événements localisés d’intense activité phytoplanctonique qui peuvent ensuite avoir un impact sur la plus grande échelle (Brentnall et al., 2003; Lévy et al., 2011). La

2. Les upwellings sont des remontées d’eaux profondes vers la surface

résolution actuelle des modèles rend le plus souvent incertaine la représentation de ces phénomènes.

À l'intérieur du gyre subtropical, la production primaire est soumise à des conditions oligotrophes (production limitée par les nutriments). La convergence des courants vers le centre du gyre, par transport d'Ekman, favorise en effet la présence de « downwellings » qui inhibe les flux verticaux vers la surface. La croissance du phytoplancton est ainsi limitée par la disponibilité en nutriments. La production primaire reste généralement faible. Elle est dominée par le phytoplancton de petite taille (Maranon et al., 1999), qui présente le meilleur rapport surface sur volume, et se caractérise par la présence d'un maximum de chlorophylle en subsurface (DCM, Pérez et al., 2006). La profondeur de ce phénomène dépend de plusieurs mécanismes et résulte d'un compromis entre intensité lumineuse et disponibilité des nutriments.

La dynamique saisonnière du phytoplancton est marquée par l'existence d'événements de fortes concentrations en chlorophylle nommés « blooms ». En hiver, la couche de mélange est bien plus profonde que la couche euphotique. Ce fort mélange a tendance à inhiber la production primaire. Au printemps, la restratification de la colonne d'eau permet l'apport de nutriments vers la surface, à l'origine de cette floraison printanière. La consommation rapide des nutriments entraîne un retour à des conditions oligotrophes dès le début de l'été. Il s'ensuit une diminution de la concentration en chlorophylle. En automne, la lente déstratification de la colonne d'eau peut parfois donner lieu à un autre bloom, de plus faible intensité. Pour un descriptif détaillé du phénomène, le lecteur peut se référer à Dutkiewicz et al. (2001) et Lévy et al. (2005). Précisons cependant que si la ventilation de la couche de mélange semble être une condition sine qua non de l'existence de blooms, leurs intensités et leurs conditions de déclenchement restent soumises à plusieurs hypothèses (p.ex. Behrenfeld and Boss, 2014; Llort, 2015)

Les régions subpolaires

La circulation océanique des hautes latitudes de l'Atlantique Nord est caractérisée par la présence du gyre subpolaire. Le courant Nord Atlantique (NAC), issu de la bifurcation du Gulf Stream, constitue la branche nord de ce gyre. Les eaux chaudes et salées qu'il apporte sont ensuite transportées vers le Groënland via le courant d'Irvinger. Durant leur trajet, ces eaux échangent d'importantes quantités de chaleur avec l'atmosphère. Leur lent refroidissement aboutit à la formation d'eaux denses qui plongent au niveau de la mer du Labrador (Schmitz and McCartney, 1993), emportant avec elles d'importantes quantités de carbone dans l'océan profond. Ce mécanisme de séquestration, associé à la circulation thermohaline, est un processus fondamental dans le contexte actuel de compréhension du cycle du carbone et de modification climatique.

Contrairement au gyre subtropical, les forts vents d'ouest des moyennes latitudes génèrent une circulation cyclonique des courants de surface et une divergence du transport d'Ekman. Les eaux sont donc généralement bien mélangées et riches en nutriments. La croissance du phytoplancton est alors limitée par la durée du jour et l'intensité lumineuse. La production primaire reste ainsi relativement élevée toute l'année, avec des valeurs maximales dans les couches de surface. Elle est soumise à un cycle saisonnier caractérisé par un bloom printanier qui se déclenche un peu après celui des latitudes moyennes et subtropicales. Si les conditions de son déclenchement sont assez comparables, sa variabilité

spatio-temporelle semble plus complexe (Lacour et al., 2015). Elle dépend de multiples facteurs tel l’Oscillation Nord Atlantique (NAO, Henson et al., 2009). Les changements climatiques, notamment la fonte des glaces (Kahru et al., 2011), semblent également avoir un impact important sur les distributions en chlorophylle de la région subpolaire.

Les régions côtières

Les caractéristiques biogéochimiques des régions côtières seront peu détaillées dans ce manuscrit. En effet, la configuration au $1/4^\circ$ du modèle PISCES, utilisée ici, ne permet pas de correctement les représenter. Leur proximité avec le littoral et l’influence des marées en font en effet le siège de processus physiques et biogéochimiques bien distincts de ceux de l’océan ouvert. Ce sont des zones riches en nutriments, dans lesquelles l’activité biologique est particulièrement prononcée. Les concentrations en chlorophylle sont donc toujours très élevées, ce qui explique qu’elles représentent un enjeu environnemental fort. En Afrique de l’Ouest (Mauritanie, Côte d’Ivoire et Gabon), on notera la présence d’intenses upwellings due au déplacement des masses d’eau vers l’ouest par les Alizés.

4.2 De LOBSTER à PISCES

Comme énoncé en introduction, la première étape de cette thèse est de justifier le passage d’un modèle biogéochimique de faible complexité, le modèle LOBSTER, à un modèle de complexité intermédiaire, le modèle PISCES. Sans faire ici une présentation approfondie du modèle LOBSTER, on peut préciser qu’il a été conçu à l’origine pour des applications locales (au sein du Programme Océan Multidisciplinaire Méso-Echelle, Paci et al., 2007; Lévy et al., 2005) et qu’il se compose de 6 variables pronostiques : 2 nutriments (nitrate, ammonium), une variable détritique, une classe de phytoplancton, une classe de zooplancton et une variable de carbone organique dissous. Pour plus d’informations, le lecteur peut se référer à la documentation de Lévy et al. (2004).

À partir de la configuration couplée NATL025/PISCES, dont les caractéristiques ont été présentées dans la partie précédente, on réalise une simulation déterministe entre janvier 2005 et décembre 2010 que l’on nommera *detIni*. Dans cette section, l’aptitude de cette simulation pour évaluer la production de chlorophylle dans l’Atlantique Nord est comparée avec celle d’une simulation du modèle LOBSTER, que l’on nommera *detLob*, établie pour les expériences d’assimilation de Fontana et al. (2013). L’intérêt de cette comparaison vient du fait que ces 2 simulations (*detIni* et *detLob*) possèdent exactement la même configuration physique (NATL025). La configuration couplée NATL025/LOBSTER ayant été validée par plusieurs travaux (Doron et al., 2011, 2013; Fontana et al., 2013), la pertinence d’utiliser le modèle PISCES est évaluée en démontrant qu’il permet d’améliorer la représentation de la chlorophylle, variable corrigée lors de l’assimilation de données de la couleur de l’eau.

4.2.1 Chlorophylle de surface

En premier lieu, on compare la faculté des modèles PISCES et LOBSTER à représenter les caractéristiques spatiales et temporelles de répartition de la chlorophylle de surface identifiables dans les observations satellites. Pour cela, on s’appuie sur la figure 4.2 qui

présente des moyennes mensuelles de la chlorophylle de surface des données SeaWIFS et des simulations *detIni* et *detLob*. Leur analyse va permettre de constater que :

1. la simulation *detIni* montre des écarts significatifs avec les observations SeaWIFS.
2. la représentation de la chlorophylle de surface du modèle PISCES affiche toutefois d'importantes améliorations par rapport à celle du modèle LOBSTER.

Concernant le point 1., on constate qu'en dépit d'une répartition spatio-temporelle de grande échelle relativement satisfaisante (différenciation nord-sud et été-hiver), PISCES peine à simuler convenablement l'intensité de l'activité biologique et à délimiter spatialement les différents types de régimes biogéochimiques. La simulation *detIni* a, par exemple, tendance à fortement surestimer les quantités de chlorophylle de surface dans la région du Gulf Stream (partie ouest du domaine, entre 30°N et 50°N). Au mois de mai, la présence d'une large zone dans laquelle les concentrations en chlorophylle sont particulièrement élevées n'apparaît jamais dans les observations. Cette exagération de la production primaire (principalement pendant le bloom de printemps) est associée à des gradients nord-sud disproportionnés autour de 30°N. Une grande partie des quantités de chlorophylle semble ainsi être déplacée vers l'est, plutôt que de longer la côte Est américaine comme on peut l'observer dans les images SeaWIFS. Un comportement comparable peut également être observé aux abords du courant équatorial. De manière générale, la persistance de forts gradients en chlorophylle contraste avec des observations satellites aux gradients nord-sud de grande échelle majoritairement faibles. En conséquence, l'étendue des conditions oligotrophes du gyre subtropical nord de la simulation *detIni* apparaît trop développée.

Dans la partie nord du domaine (au-dessus de 50°N), le modèle PISCES a du mal à reproduire les fortes variabilités spatiales que l'on peut observer dans les données satellites. Les concentrations en chlorophylle semblent de plus régulièrement sous-estimées.

Malgré ces imperfections, la représentation de la chlorophylle de surface du modèle PISCES présente tout de même d'importantes améliorations par rapport à celle du modèle LOBSTER (point 2.). Dans les hautes latitudes (au-dessus de 50°N), la flagrante sous-estimation hivernale des quantités de chlorophylle de la simulation *detLob* est clairement en désaccord avec les observations SeaWIFS. Au moment du déclenchement du bloom de printemps, les très faibles concentrations laissent ensuite rapidement place à de très fortes concentrations qui conduisent immédiatement à une surestimation de la chlorophylle. L'importante hétérogénéité spatiale de petite échelle des images SeaWIFS est également absente de la représentation du modèle LOBSTER.

En dépit d'une surestimation de la chlorophylle par le modèle PISCES, la propagation nordique de la production primaire du Gulf Stream suit mieux la ligne de côte Est-Américaine. Dès la fin de la période estivale (septembre), le modèle LOBSTER simule de faibles quantités de chlorophylle de manière quasiment uniforme sur l'ensemble du domaine. La persistance hivernale du modèle PISCES est donc beaucoup plus satisfaisante. Dans le gyre subtropical, l'influence des conditions oligotrophes est largement sous-estimée dans la simulation *detLob*. Ce constat peut s'expliquer, entre autres, par une prédominance d'organismes de petite taille qui ne peut s'exprimer en tenant compte d'une seule classe de phytoplancton.

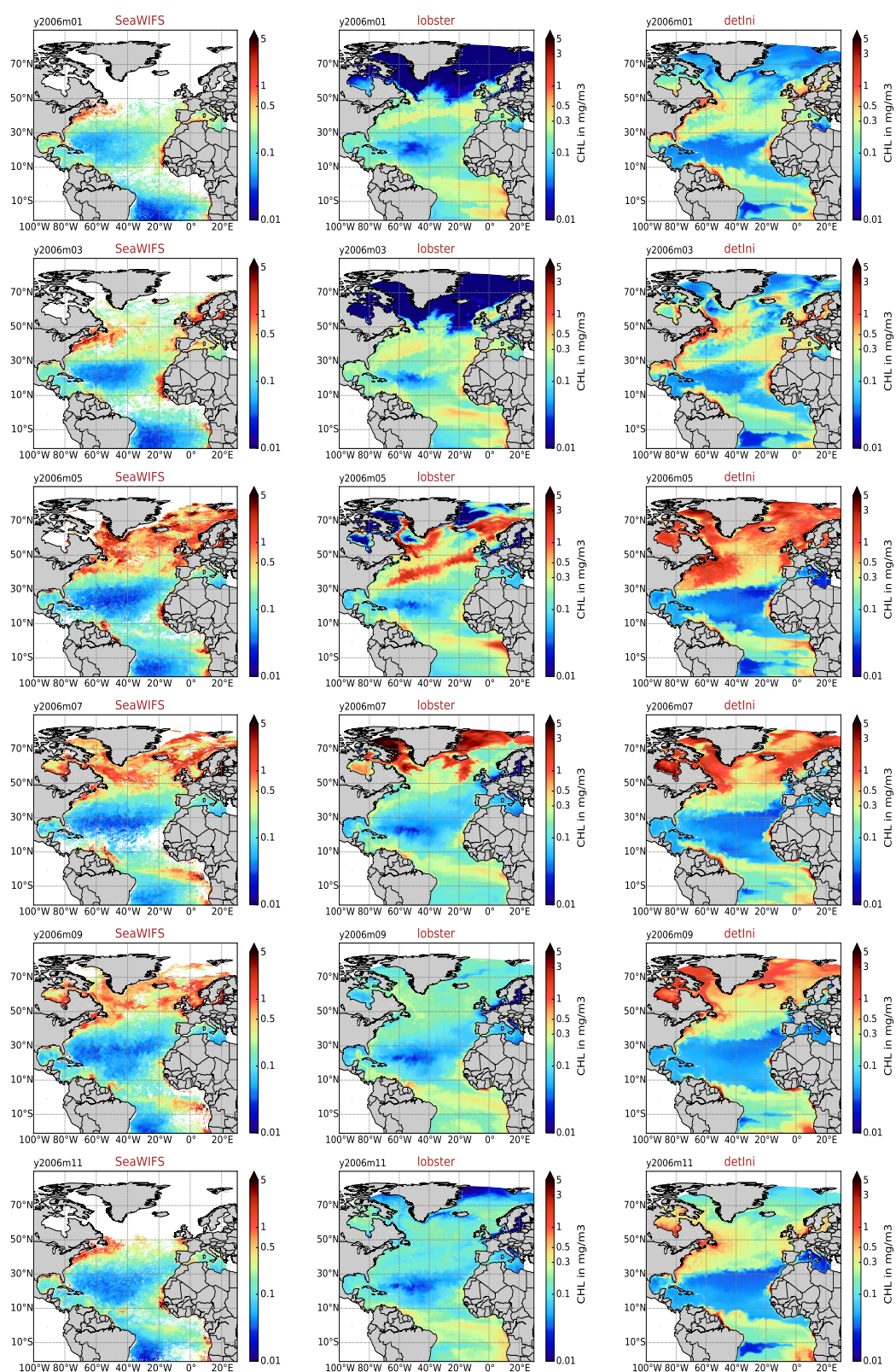


FIGURE 4.2 – Cartes de surface des moyennes mensuelles de la chlorophylle des données SeaWiFS (1^{ère} colonne), du modèle LOBSTER (simulation *detLob*, 2^{ème} colonne) et du modèle PISCES (simulation *detIni*, 3^{ème} colonne). En partant de la ligne du haut, les résultats sont présentés pour successivement les mois de janvier, mars, mai, juillet, septembre et novembre de l'année 2006.

Au niveau des zones côtières, la simulation *detIni* représente convenablement les fortes concentrations en chlorophylle visibles dans les images de couleur de l'eau. Contrairement à PISCES, le modèle LOBSTER ne permet pas de stimuler la production phytoplanctonique par des apports nutritifs provenant des rivières, des sédiments ou des poussières aériennes (cf. section 3.3.2). Je précise qu'une simulation "test" du modèle PISCES, réalisée sans ces apports extérieurs, a révélé des concentrations de chlorophylle bien plus réduites au niveau des régions côtières. Par ailleurs, considérer uniquement l'azote comme élément nutritif peut expliquer en partie l'existence de « déserts phytoplanctoniques » dans la simulation *detLob*.

Pour résumer, les distributions spatiales et temporelles de la chlorophylle de surface de la simulation *detIni* sont nettement plus en accord avec les observations SeaWiFS que celles de la simulation *detLob*. Malgré des écarts aux observations qui restent significatifs (résultant également en partie des incertitudes sur la physique), l'utilisation du modèle PISCES permet donc d'améliorer incontestablement l'estimation de la chlorophylle de surface dans l'océan Atlantique Nord.

4.2.2 Distribution verticale de la chlorophylle

L'une des difficultés majeures de l'assimilation de données de la couleur de l'eau provient d'observations uniquement disponibles en surface. L'extrapolation des valeurs de chlorophylle sur la verticale est donc une donnée essentielle. Entre autres, la transition vers une plate-forme biogéochimique de complexité supérieure s'explique par le besoin de fournir une contrainte dynamique verticale pertinente pour l'assimilation de données. Dans cette section, l'objectif est de s'assurer que la complexité biogéochimique du modèle PISCES améliore également la représentation verticale de la chlorophylle.

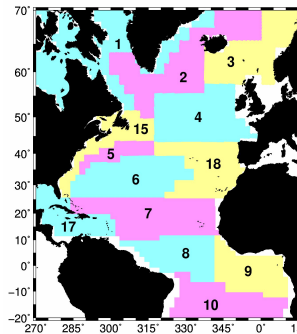


FIGURE 4.3 – Provinces biogéochimiques de Longhurst (Longhurst, 1995), telles que présentées dans Doron et al. (2011). Les zones côtières sont exclues de cette représentation.

Sur la verticale, les observations disponibles sont en général éparées et de faible résolution temporelle. Pour comparer les structures verticales de la simulation *detIni* avec celles de la simulation *detLob*, on utilise une base de données climatologiques récemment établie par Sauzede et al. (2015). Cette climatologie de chlorophylle est dérivée de données in-situ de fluorescence regroupées par province de Longhurst dont on rappelle les délimitations spatiales en figure 4.3. Ces données climatologiques possédant un niveau d'incertitude non

négligeable, on considérera ici l’allure des profils verticaux plutôt que la valeur stricto-sensu de l’observation à un niveau de profondeur donné. Les profils verticaux des données climatologiques sont comparés en figure 4.4 avec des moyennes mensuelles et spatiales des simulations des modèles LOBSTER et PISCES pendant l’année 2006. Afin d’enrichir la comparaison, des profils mensuels de la simulation *detIni* sont calculés à partir de moyennes spatio-temporelles sur les 6 années de simulation. Cette représentation déterministe peut être considérée comme une estimation “climatologique” du modèle (notée ici *detIni-clim*). Par province, nous pouvons observer les comportements suivants.

- Pour la province 7 (gyre subtropical), les profils verticaux des simulations *detIni* et *detIni-clim* sont en bon accord avec la climatologie (mis à part en mars mais la climatologie contient très peu de données). Au mois d’octobre, même si les maximums de subsurface (DCM) sont localisés légèrement trop en profondeur, leurs étendues verticales et leurs intensités sont correctement reproduites. Le modèle LOBSTER simule lui aussi bien l’existence des DCM. Il sous-estime toutefois leurs profondeurs et leurs amplitudes verticales.
- Dans les plus hautes latitudes (province 4), les distributions verticales des 2 modèles sont qualitativement assez comparables. Les difficultés pour reproduire des profils verticaux en accord avec les observations peuvent ainsi être reliées à la tendance à sous-estimer la chlorophylle dans les couches de surface.
- Dans la province 6 (région du Gulf Stream), les structures verticales des 2 modèles sont très différentes de celles de la climatologie pour le mois de mai. Ces écarts s’accompagnent de fortes surestimations des concentrations en chlorophylle de surface. Pour les 2 autres dates, le modèle PISCES reproduit plutôt convenablement l’allure des profils verticaux. Les maximums de subsurface de la simulation *detLob* sont, par contre, assez différents des observations en ce qui concerne leurs localisations verticales et leurs intensités.
- Dans la partie sud du domaine (province 9), les comportements sont comparables à ceux de la province 6. Pendant le bloom de printemps, aucune simulation ne reproduit l’intense maximum de subsurface présent dans les données climatologiques. En mars, seuls les profils verticaux de la simulation *detIni* sont cohérents avec la climatologie. En octobre, les observations sont constituées de trop peu de données.

De manière générale, les profils verticaux simulés par le modèle PISCES sont plus proches des observations que ceux du modèle LOBSTER. Les profils “climatologiques” (*detIni-clim*) sont presque toujours identiques à ceux de la simulation *detIni*, signe d’une variabilité interannuelle relativement faible (cf. section suivante).

On peut également remarquer que la cohérence des structures verticales semble dépendre de l’estimation de la chlorophylle de surface. Dans la province 6, les simulations ont par exemple tendance à surestimer la production en chlorophylle pendant la période du bloom de printemps (mois de mai). Il s’ensuit une diminution des concentrations avec la profondeur, alors que les observations indiquent un bloom de subsurface. En octobre, lorsque la représentation de la chlorophylle de surface est proche des observations, PISCES simule convenablement l’existence d’un maximum de subsurface. Le lien entre chlorophylle de surface et production verticale a d’ailleurs été préalablement évoqué par les travaux

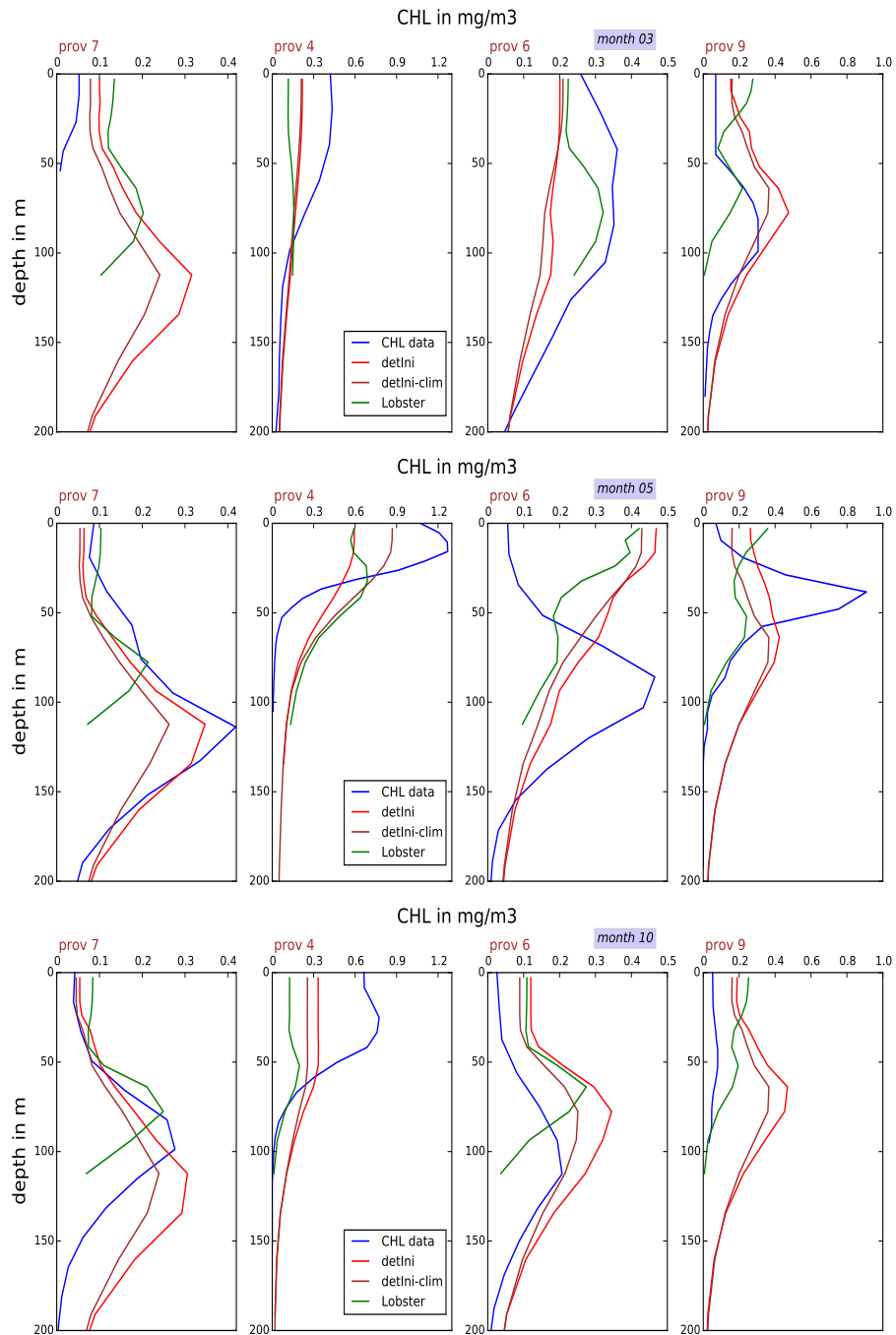


FIGURE 4.4 – Profils verticaux de la chlorophylle présentés en moyenne spatiale pour les provinces de Longhurst 4, 6, 7 et 9. La courbe bleue représente une climatologie établie par Sauzède et al. (2015) (notée CHL data), la courbe marron est une climatologie sur 6 ans de la simulation *detIni* (notée ici *detIni-clim*). Les courbes vertes et rouges sont des moyennes mensuelles des modèles LOBSTER (simulation *detLob*, notée ici Lobster) et PISCES (simulation *detIni*), présentées pour les mois de mars (1^{ère} ligne), mai (2^{ème} ligne) et novembre (3^{ème} ligne) de l'année 2006.

de Uitz et al. (2006). L'amélioration de la chlorophylle de surface sera justement l'une des préoccupations du chapitre 5. Nous présenterons pour cela une méthode stochastique permettant de simuler certaines sources d'incertitude biogéochimiques du modèle PISCES.

Conclusion

Pour conclure, l'ensemble des résultats présentés dans cette section montrent que le passage LOBSTER → PISCES permet globalement d'améliorer l'estimation de la chlorophylle dans l'océan Atlantique Nord. Puisque le modèle physique et les forçages utilisés sont identiques pour les 2 simulations, cet effet ne peut provenir que des formulations biogéochimiques, c'est-à-dire du niveau de complexité du modèle PISCES. Les travaux de Mattern et al. (2015) qui comparent du point de vue de l'assimilation de données, les performances de modèles biogéochimiques possédant différents niveaux de complexité, laissent présager que ces améliorations pourront ensuite se répercuter positivement lors de l'assimilation de données de la couleur de l'eau.

4.3 Variabilité de la simulation déterministe

Dans la section précédente, la pertinence d'utiliser le modèle PISCES à la place du modèle LOBSTER a pu être démontrée. Avant de mettre en place une version probabiliste du modèle PISCES (ce sera l'objet du chapitre suivant), nous présentons certaines caractéristiques de la chlorophylle extraites des 6 années de la simulation *detIni*.

Dans la première partie de cette section, on s'intéresse à la variabilité interannuelle. Cela nous permettra d'examiner la dépendance de la représentation de la chlorophylle aux paramètres extérieurs (p.ex. forçages atmosphériques). La seconde partie a pour but d'approfondir la comparaison initiée dans la section 4.2.1 entre l'estimation de la chlorophylle du modèle PISCES et les observations SeaWIFS. On s'attardera en particulier sur l'analyse des variabilités spatiales et temporelles. La dernière partie de cette section aborde la notion de diversité des espèces. Nous regarderons, de manière assez succincte, si la distinction du phytoplancton en 2 classes peut être suffisante pour représenter les principaux comportements biogéochimiques de l'océan Atlantique Nord.

4.3.1 Variabilité interannuelle de la chlorophylle

Pour caractériser la variabilité interannuelle de la simulation (que l'on qualifiera parfois par abus de langage de variabilité "climatologique"), on calcule en chaque point de grille du modèle et à chaque pas de temps t du modèle (tous les 5 jours) l'écart-type interannuel sd^t des concentrations de chlorophylle des 6 années de simulation tel que :

$$sd^t = \sqrt{\frac{1}{6} \sum_{a=2005}^{2010} (x^{ta} - \overline{x^{ta}})^2} \quad \text{avec} \quad \overline{x^{ta}} = \frac{1}{6} \sum_{a=2006}^{2010} x^{ta} \quad (4.1)$$

où x^{ta} correspond ici à la concentration en chlorophylle au pas de temps t de l'année de simulation a (de 2005 et 2010). l'écart-type interannuel sd^t représente donc la racine carrée de l'écart quadratique moyen entre la moyenne "climatologique" du modèle et les concentrations en chlorophylle des différentes années de la simulation. Plus la valeur de cet

écart-type sera élevée, plus la variabilité interannuelle de la simulation *detIni* (renotée sur les figures *detIni-clim*) sera importante.

Variabilité de surface

Pour le 31 mars, le 15 mai et le 27 octobre, on représente en figure 4.5 des cartes de l'écart-type interannuel de surface sd^t (noté *sd* sur les figures). Afin de déconvoluer l'écart-type de la valeur des concentrations en chlorophylle (difficulté inhérente au mélange statistique de valeurs aussi hétérogènes que celles rencontrées en biogéochimie), on représente également les cartes de surface de l'écart-type relatif (noté *relative sd* sur les figures), calculé comme le rapport sd^t/x^{t_a} .

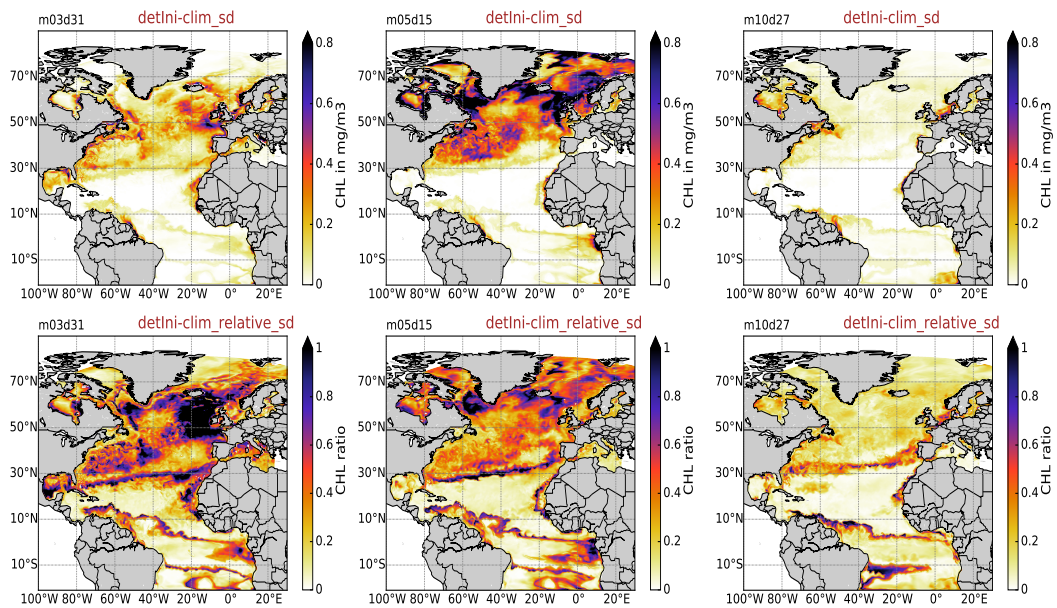


FIGURE 4.5 – Cartes de surface des écarts-types (*sd*, 1^{ère} ligne) et des écarts-types relatifs (*relative sd*, 2^{ème} ligne) interannuels de la chlorophylle, calculés selon l'équation 4.1 à partir des 6 années de simulation. Les résultats sont présentés pour le 31 mars (1^{ère} colonne), le 15 mai (2^{ème} colonne) et le 27 octobre (3^{ème} colonne).

À partir des cartes présentées en figure 4.5, on constate que la variabilité interannuelle de la simulation *detIni* ne prend des valeurs significatives qu'à partir de la période du bloom de printemps. Elle se restreint à la partie nord du domaine (au-dessus de 30°N) et ne dépasse les $0.8 \text{ mg chlorophylle}/\text{m}^3$ que dans certaines zones côtières, localisées dans les hautes latitudes. En dessous de 30°N , les concentrations en chlorophylle ne varient quasiment pas d'une année sur l'autre, quelle que soit la période considérée.

À la fin du mois de mars, l'effet du bloom de printemps ne se fait pas encore ressentir sur la productivité totale. Les concentrations en chlorophylle sont donc encore relativement faibles sur l'ensemble du domaine. Les cartes des écarts-types relatifs affichent alors de fortes valeurs, régulièrement supérieures à 40% de la moyenne climatologique, dans la partie nord du domaine. Compte tenu des faibles concentrations de chlorophylle, des différences significatives existent donc entre les années. Cette importante variabilité in-

terannuelle, diagnostiquée par l'écart-type relatif, peut en partie s'expliquer par une date d'initiation du bloom de printemps pouvant varier de quelques jours selon les années. À l'ouest des côtes de l'Europe du Nord, l'existence d'une zone de forte variabilité peut également être observée. Cette zone est en effet reconnue pour subir régulièrement, durant l'hiver, d'intenses événements climatiques.

Dans la partie sud du domaine, la faible variabilité interannuelle, visible sur les cartes des écarts-types (relatifs et absolus), n'est pas cette fois compensée par de faibles concentrations en chlorophylle. Au mois d'octobre, l'absence de variabilité interannuelle est confirmée sur l'ensemble du domaine par les cartes des écarts-types absolus. Sur les cartes des écarts-types relatifs, on remarque néanmoins des lignes de forte variabilité permanente (même en octobre), situées au niveau des fronts du gyre subtropical et du courant sud-équatorial. Ces zones avaient déjà été identifiées en section 4.2.1 comme le siège de brusques gradients, contrastant avec les images satellites. Si l'existence de ces forts gradients est systématique, leur localisation exacte peut légèrement varier d'une année sur l'autre, expliquant la variabilité observée. Ce point met également en avant la stabilité des structures de la simulation *detIni*.

Afin d'analyser plus précisément la variabilité climatologique du modèle PISCES, on s'intéresse à l'évolution temporelle de l'écart-type et de l'écart-type relatif interannuels. Pour éviter de se restreindre à un seul point de grille tout en conservant une cohérence biogéochimique, les écarts-types sont moyennés par province de Longhurst (cf. figure 4.3). En figure 4.6, la variabilité interannuelle de la simulation *detIni* est présentée pour les provinces 4, 6, 7 et 9.

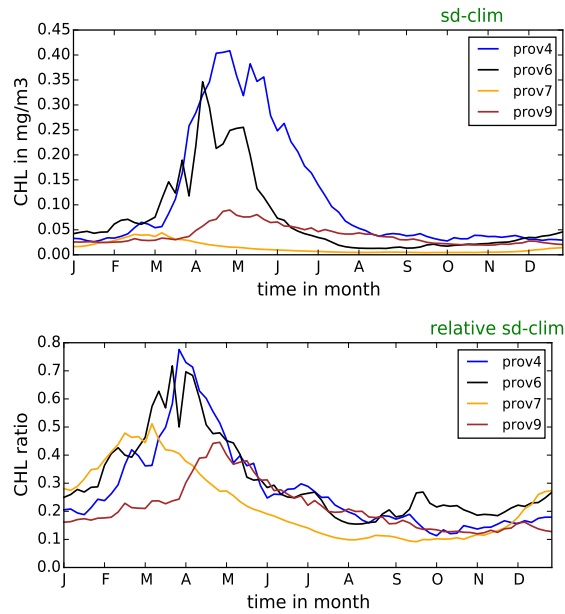


FIGURE 4.6 – Séries temporelles des écarts-types (haut) et des écarts-types relatifs (bas) interannuels de la chlorophylle de surface, calculés selon l'équation 4.1 à partir des 6 années de simulation. Les résultats sont présentés en moyenne spatiale pour les provinces de Longhurst 4, 6, 7 et 9.

Globalement, l'analyse des séries temporelles de la figure 4.6 confirme l'essentiel des tendances observées précédemment sur la figure 4.5.

- La variabilité interannuelle de la simulation se restreint à la partie nord du domaine (provinces 4 et 6). À noter que dans les provinces 7 et 9, les valeurs élevées d'écart-type relatif sont principalement dues à la variabilité des fronts du gyre subtropical et du courant sud-équatorial. Elles ne reflètent donc pas complètement la faible variabilité de ces 2 provinces.
- La variabilité interannuelle est maximale en période de forte dynamique biogéochimique (bloom de printemps).
- L'intensité de la variabilité interannuelle, diagnostiquée par l'écart-type absolu, va toujours de paire avec les concentrations en chlorophylle. Elle augmente au moment de l'initiation du bloom de printemps (mois de mars), atteint son maximum entre avril et juin, avant de rediminuer et conserver des valeurs faibles ($\leq 0.05 \text{ mg/m}^3$) le restant de l'année.
- Par rapport aux concentrations en chlorophylle, la variabilité est maximale au début du bloom de printemps (de fin mars à début mai) et stagne ensuite autour de 20% sur l'ensemble du domaine.

Variabilité verticale

Sur la figure 4.4, nous avons constaté que les profils verticaux de la simulation *detIni*

(pour l'année 2006) et de la moyenne climatologique (profils *detIni-clim*) étaient très semblables. Afin de caractériser plus précisément la variabilité interannuelle sur la verticale, la figure 4.7 représente les écarts-types absolus et relatifs le long de coupes longitudinales à 30°W. Les résultats sont présentés pour le 31 mars, le 15 mai et le 27 octobre.

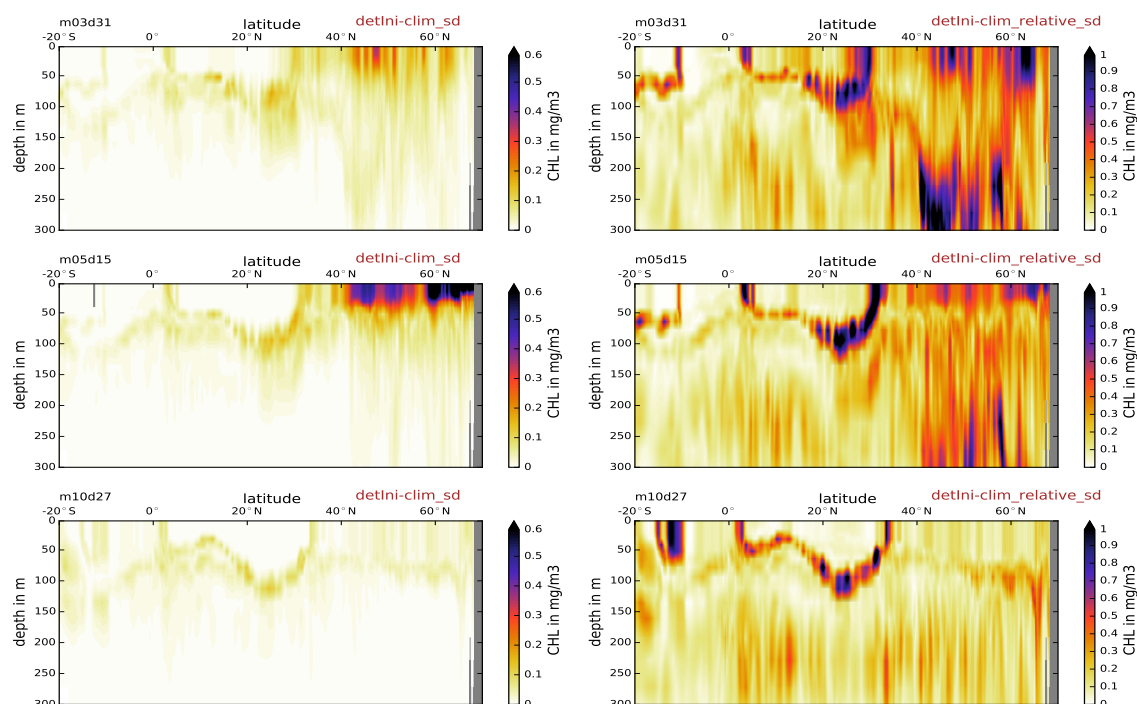


FIGURE 4.7 – Structures verticales des écarts-types (sd, 1^{ère} colonne) et des écarts-types relatifs interannuels (relative sd, 2^{ème} colonne) de la chlorophylle le long d'une coupe longitudinale à 30°W. Les résultats sont calculés selon l'équation 4.1 à partir des 6 années de simulation. Ils sont présentés pour le 31 mars (1^{ère} ligne), le 15 mai (2^{ème} ligne) et le 27 octobre (3^{ème} ligne).

Pour les 3 dates considérées, les écarts-types sont généralement faibles dans l'ensemble de la couche euphotique. La variabilité interannuelle ne dépasse 0.2 mg/m³ que sur les premiers niveaux verticaux et seulement au-dessus de 40°N. Des valeurs significatives d'écart-type (autour de 0.1 mg/m³) peuvent également être distinguées aux niveaux des maximums de subsurface (DCM). Mis à part ces profondeurs, la variabilité interannuelle est presque partout proche de zéro en dehors de la période du bloom de printemps.

Compte tenu des faibles concentrations en chlorophylle dans la colonne d'eau, les maximums de subsurface présentent une forte variabilité interannuelle (autour de 80%). À l'exception de leurs niveaux de profondeur, les valeurs d'écart-type relatif sont quasiment nulles dans toute la partie sud du domaine. Dans la partie nord, les fortes valeurs d'écart-type absolu du bloom de printemps sont atténuées par des concentrations en chlorophylle plus importantes. Dans les 50 premiers mètres, la variabilité interannuelle reste toutefois significative. En octobre, la variabilité relative est partout très faible en dehors des profondeurs de localisation des maximums de subsurface.

En profondeur (autour de 250-300 mètres), on constate que l'écart-type relatif peut

prendre des valeurs élevées, en particulier dans la partie nord du domaine (légèrement perceptibles fin mars sur la coupe des écarts-types absolus). Logiquement, en dessous de la couche euphotique (en général une centaine de mètres), l'intensité lumineuse est nulle. Il est donc peu réaliste d'imaginer la présence de chlorophylle et encore moins l'existence d'une variabilité significative. Les concentrations en chlorophylle étant très faibles, cette variabilité reste ici toutefois un artefact puisque toute variance, même infime, peut conduire à d'importantes valeurs.

Bilan

Pour récapituler, l'essentiel de la variabilité interannuelle de la couche euphotique se produit aux profondeurs des maximums de subsurface et dans les 50 premiers mètres de la partie nord du domaine. Dans la partie nord, nous pouvons associer cette variabilité à celle de la chlorophylle de surface. En ce qui concerne les maximums de subsurface, la variabilité est certainement associée à la représentation de la couche de mélange.

De manière générale, la variabilité interannuelle reste très localisée. Elle se décèle principalement aux profondeurs des maximums de subsurface et agit donc peu sur leur localisation verticale. Par conséquent, l'allure des profils verticaux est peu modifiée d'une année sur l'autre. En figure 4.4, nous avons observé des différences manifestes entre les profondeurs des maximums de subsurface de la simulation *detIni* et celles des données climatologiques. Ces résultats laissent ainsi suggérer que la variabilité climatologique du modèle ne pourra expliquer que partiellement ces écarts aux observations.

Conclusion partielle

L'existence d'une variabilité interannuelle majoritairement présente dans la partie nord du domaine suggère que l'estimation de la production primaire est sensible aux variations saisonnières des forçages atmosphériques. L'absence de rythme saisonnier et d'événements climatiques prononcés dans la partie sud du domaine peut ainsi expliquer, en partie, l'absence de variabilité au-dessous de 30°N. Évidemment une analyse plus poussée serait nécessaire pour déterminer précisément les processus qui pilotent cette variabilité interannuelle. Cela reste toutefois en dehors du cadre de cette étude. De plus, se baser sur seulement 6 années de simulation paraît insuffisant pour pouvoir véritablement caractériser une variabilité climatologique. On se restreindra donc aux implications potentielles de ces résultats pour l'assimilation de données. Ils semblent, en effet, indiquer que les incertitudes des paramètres extérieurs ne pourront qu'en partie expliquer les écarts aux observations SeaWIFS.

En assimilation de données, les modes dominants de propagation d'erreur sont généralement estimés à partir de variabilités climatologiques. L'estimation de cette incertitude peut ainsi être comparée à la variabilité interannuelle présentée dans cette section. L'erreur a priori du modèle, ainsi calculée, serait donc vraisemblablement insuffisante pour caractériser pleinement les incertitudes du modèle. Dans la partie suivante, on examine de manière plus précise les circonstances dans lesquelles cette variabilité interannuelle peut expliquer les écarts aux observations SeaWIFS.

4.3.2 Comparaison avec les observations SeaWIFS

Écarts moyens

Dans la section 4.2, on a mis en évidence les principales différences entre les observations SeaWIFS et la simulation *detIni* sur l'année 2006. On ne réitérera pas ici cette comparaison pour les 6 années de simulation car les conclusions seraient sensiblement identiques. Néanmoins, il est possible de généraliser les écarts entre le modèle et les observations en calculant la racine carrée de l'erreur quadratique moyenne (RMSE). La RMSE peut être considérée comme équivalente à l'écart-type calculé précédemment en remplaçant la moyenne inter-annuelle $\overline{x^{ta}}$ par la valeur de l'observation y^{ta} (qui varie cette fois d'une année sur l'autre). Elle peut s'exprimer de la façon suivante.

$$RMSE^t = \sqrt{\frac{1}{6} \sum_{a=2005}^{2010} (x^{ta} - y^{ta})^2} \quad (4.2)$$

À chaque pas de temps, nous pouvons calculer ainsi la racine carrée de la moyenne (des 6 années de simulation) de l'écart quadratique entre la chlorophylle de surface x^{ta} , estimée par le modèle, et les observations y^{ta} . Le passage de la grille du modèle vers celle de SeaWIFS (au $1/12^\circ$) est fait par interpolation bilinéaire. Je précise que d'une année sur l'autre, la couverture nuageuse peut être différente et par là même la quantité d'observations disponibles. Le nombre de données, utilisé pour calculer la moyenne, peut donc varier entre 2 points de grille des observations. Au minimum, 3 valeurs sont disponibles entre 2005 et 2010. Pour réduire l'impact de la variation du nombre de données, les RMSE sont moyennées spatialement par province de Longhurst. La valeur relative de cette RMSE (absolue) est ensuite obtenue en normalisant par la concentration climatologique en chlorophylle $\overline{x^{ta}}$ (moyennée par province de Longhurst). Pour les provinces 4, 6, 7 et 9, les séries temporelles des RMSE absolues et relatives sont présentées en figure 4.8.

En tenant compte des résultats obtenus précédemment lors des comparaisons avec les images SeaWIFS (section 4.2.1), plusieurs points peuvent être soulignés.

- À cause de conditions oligotrophes, les ordres de grandeur des concentrations en chlorophylle de la province 7 sont globalement faibles. Ce constat explique que les RMSE absolues soient relativement constantes et inférieures à celles des autres provinces. Par rapport à ces faibles concentrations, on observe une forte variabilité de la RMSE relative au cours de l'année. Ces fluctuations temporelles peuvent être perçues comme le reflet d'une étendue spatiale du gyre subtropical différente dans le modèle et les observations.
- Dans la province 9, les concentrations en chlorophylle ont peu d'impact sur l'amplitude des écarts aux observations. Les courbes présentant l'évolution temporelle de la RMSE et de la RMSE relative ont des allures similaires, marquées par une faible saisonnalité. Les maximums, observés entre juillet à fin septembre, peuvent s'expliquer par la tendance du modèle à sous-estimer la chlorophylle dans partie côtière du courant sud-équatorial et au niveau de l'upwelling du Benguela.

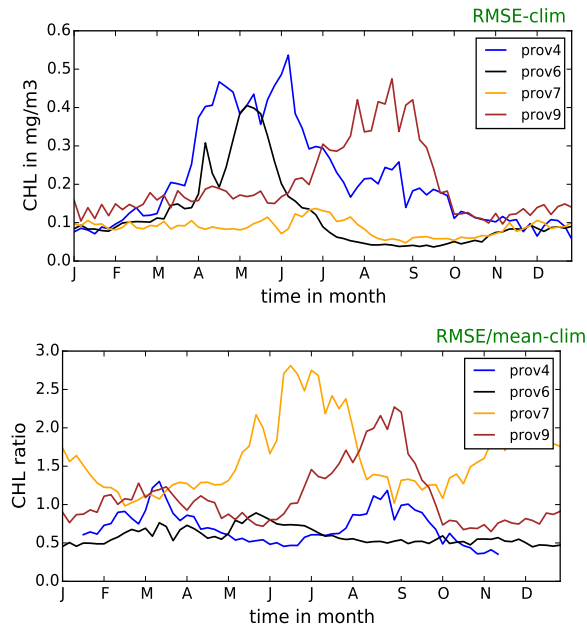


FIGURE 4.8 – Séries temporelles de la RMSE (haut) et de la RMSE relative (bas) moyennées spatialement dans les provinces de Longhurst 4,6,7 et 9. La RMSE est calculée selon l'équation 4.2 et la RMSE relative est calculée comme le rapport $(RMSE^t / \overline{x^{t_a}})$ entre la RMSE (absolue) et la concentration climatologique de la chlorophylle $\overline{x^{t_a}}$.

- Plus au nord (province 4), les fortes valeurs de RMSE, en période de bloom de printemps, sont principalement à mettre sur le compte de concentrations plus élevées en chlorophylle. Dans la région du Gulf Stream (province 6), l'évolution temporelle des écarts (absolus) aux observations à une allure similaire à celle de la province 4, avec un maximum prononcé pendant le bloom de printemps. Les valeurs de la RMSE relative sont relativement constantes (autour de 50%). Ceci peut manifester la présence d'un biais dans le modèle allant, d'après la figure 4.2, dans le sens d'une surestimation de la chlorophylle. La régularité des valeurs de RMSE dans les provinces 6 et 4 montre que les variations saisonnières de la chlorophylle sont, en moyenne, respectées dans ces 2 régions.

Récapitulatif

En résumé, ces analyses confirment, pour les 6 années de simulation, l'essentiel des caractéristiques spatiales et temporelles de répartition de la chlorophylle de surface déduites de la figure 4.2. À savoir :

- une surestimation permanente de la chlorophylle de surface autour de la région du Gulf Stream.
- des écarts aux observations maximum en période de bloom de printemps dans la partie nord du domaine.
- une différenciation nord-sud et une saisonnalité qui respectent les gammes de concen-

tration en chlorophylle identifiables dans les observations.

Dans la section 4.3.1, la variabilité interannuelle du modèle a été décrite à partir de l'écart quadratique moyen sd^t autour de la valeur climatologique moyenne du modèle $\overline{x^{ta}}$. Pour évaluer l'écart moyen aux observations ($RMSE^t$), le calcul de la RMSE considère non pas l'écart entre le modèle et l'observation "climatologique" moyenne (qui serait l'équivalent rigoureux de la valeur climatologique moyenne du modèle $\overline{x^{ta}}$), mais l'écart entre le modèle et la valeur y^{ta} de l'observation considérée pour chaque année. Néanmoins, nous pouvons estimer que l'ordre de grandeur du rapport entre ces 2 quantités est quantitativement représentatif de la cohérence entre la variabilité interannuelle du modèle (principalement due aux conditions atmosphériques) et l'amplitude des écarts aux observations. En d'autres termes, le rapport $RMSE^t/sd^t$ peut apporter des éléments de réponse à la question : dans quelles mesures est-ce que l'incertitude du forçage physique permet expliquer les écarts aux observations ?

En figure 4.9, nous présentons pour cela les séries temporelles du rapport $RMSE^t/sd^t$ entre l'écart moyen avec les observations SeaWIFS ($RMSE^t$) et la variabilité interannuelle de la simulation *detIni* (sd^t), préalablement moyennés par province de Longhurst. L'égalité entre la RMSE et l'écart-type est signalée par la ligne en pointillés verts. Lorsque l'on se situe au-dessus de cette ligne, ce diagnostic laisse suggérer que le niveau d'incertitude sur le forçage physique³ est globalement inférieure à l'amplitude des écarts aux observations.

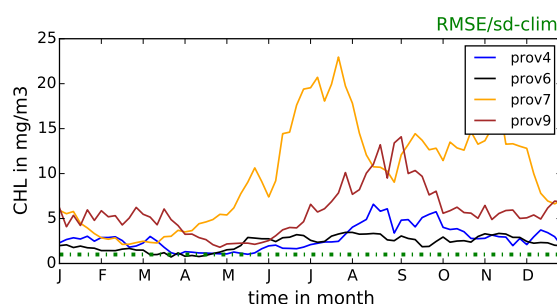


FIGURE 4.9 – Séries temporelles de la RMSE moyennée spatialement puis normalisées par l'écart-type sd^t moyen calculé précédemment selon l'équation 4.1. Les résultats sont présentés pour les provinces de Longhurst 4,6,7 et 9. La ligne en pointillés verts exprime l'égalité entre la RMSE et l'écart-type interannuel.

Dans la partie sud du domaine, les rapports très élevés s'expliquent par une variabilité interannuelle du modèle presque toujours proche de 0 (cf section 4.3.1). Au nord, malgré la forte variabilité interannuelle, les valeurs de l'écart-type et de la RMSE ne sont comparables qu'au début du bloom de printemps (période de variabilité interannuelle relative maximale). Au niveau du Gulf Stream (province 6), le rapport est proche de 1. Le modèle présentant un fort biais positif dans cette région, l'écart-type interannuel est estimé dans une gamme de concentrations très supérieure. Le rapport entre ces 2 quantités ne peut donc pas s'avérer représentatif puisque le modèle surestime en permanence la chlorophylle.

3. La variabilité intrinsèque est négligée

Bilan

De manière générale, les écarts moyens aux observations, diagnostiqués par la RMSE, sont toujours très supérieurs à l'écart-type interannuel de la simulation *detIni*. Ce résultat souligne une nouvelle fois la difficulté de ces incertitudes "climatologiques" à expliquer l'amplitude des écarts aux observations.

Variabilité spatiale

Dans la section 4.2.1, l'analyse de la figure 4.2 a montré que la simulation *detIni* présente, en général, des structures spatiales relativement lisses (des concentrations assez homogènes), qui contrastent avec la forte variabilité spatiale de petite échelle des images SeaWIFS. Les comportements biogéochimiques étant très non linéaires, ces fluctuations de petite échelle ne sont généralement pas de moyenne nulle. Elles peuvent donc avoir un impact sur la production totale. La RMSE exprimant un écart moyen (entre le modèle et les observations), la variabilité spatiale n'était pas incluse dans ce diagnostic. Une partie non négligeable de l'amplitude des écarts avec les observations peut néanmoins s'expliquer par des différences d'hétérogénéité spatiale. Afin de comparer la variabilité spatiale de la chlorophylle de la simulation *detIni* avec celle des observations SeaWIFS, on calcule, pour chaque année à chaque pas de temps du modèle, l'écart-type des concentrations en chlorophylle dans les provinces de Longhurst 4, 6, 7 et 9. Pour les 6 années de simulation, les résultats sont présentés sur la figure 4.10.

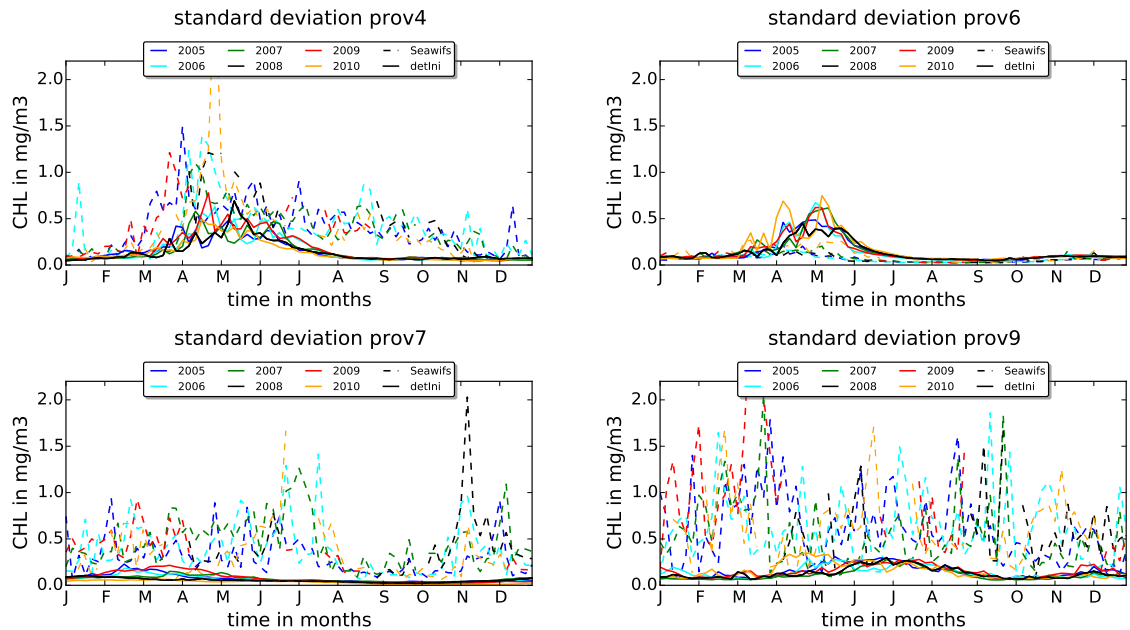


FIGURE 4.10 – Séries temporelles des écarts-types de la chlorophylle de surface calculés dans les provinces de Longhurst 4, 6, 7 et 9 pour chacune des 6 années de simulation. Les lignes en pointillés présentent des résultats pour les observations SeaWIFS et les lignes pleines pour la simulation *detIni*.

Mis à part dans la province 6, où la persistance de faibles concentrations avait été observée dans les images satellites, la figure 4.10 montre que la variabilité spatiale des données SeaWiFS est toujours supérieure à celle de la simulation *detIni*. Que ce soit pour les observations où le modèle PISCES, les ordres de grandeur des écarts-types sont relativement constants d'une année sur l'autre. La majeure partie des différences, entre les années, de la variabilité spatiale des observations peut s'expliquer par la disponibilité des observations considérées pour le calcul des écarts-types.

Dans la simulation *detIni*, l'évolution temporelle des écarts-types suit globalement celle des quantités de chlorophylle. L'hétérogénéité spatiale est maximale pendant le bloom de printemps et dans la moitié nord du domaine.

Si on compare les valeurs d'écart-type des provinces 4 et 7, nous pouvons constater que les observations affichent, au contraire, des variations spatiales plus élevées lorsque les concentrations en chlorophylle sont faibles. À noter que la très forte hétérogénéité que l'on peut observer dans la province 9 est principalement due à l'upwelling de Benguela. Dans la partie nord du domaine (province 4), la variabilité spatiale des observations est maximale lors de l'initiation du bloom de printemps, lorsque les concentrations moyennes sont encore faibles. Dans la province 6, la variabilité spatiale des données SeaWiFS est toujours faible, inférieure à celle de la simulation *detIni*. L'influence du bloom de printemps se fait à peine ressentir. Si les variabilités spatiales du modèle et des observations sont comparables, nous avons constaté (cf. section 4.2.1) que le modèle surestime fortement, dans cette zone, les concentrations en chlorophylle.

Bilan

De manière générale, l'importante variabilité spatiale des observations peut être reliée à l'existence de blooms de petite échelle, non simulés par le modèle. Une telle hétérogénéité non simulée par le modèle a forcément un effet, à la fois sur la représentation locale (différence en un point de grille) et sur une plus large échelle (moyenne non nulle des fluctuations), pouvant expliquer certains écarts aux observations. Dans cette variabilité, une part provient de la différence entre la résolution du modèle et celle des observations. Puisque le modèle biogéochimique est non linéaire, les fluctuations des petites échelles, non résolues, peuvent entraîner un biais sur l'estimation moyenne des quantités de chlorophylle. Ce point sera réabordé dans le chapitre suivant (section 5.1.2.2) lorsque l'on présentera une paramétrisation stochastique permettant de prendre en compte cette source d'incertitude.

4.3.3 Diversité des espèces

Dans les sections précédentes, la variabilité spatiale et temporelle des distributions de la chlorophylle nous a contraint à fractionner notre domaine d'étude. Cette régionalisation s'est faite selon les provinces de Longhurst, dont les délimitations ont été établies en fonction des différences de comportements biogéochimiques. Dans PISCES, la chlorophylle totale est calculée comme la somme des contributions respectives des diatomées et du nanophytoplancton. Les différents régimes biogéochimiques ne peuvent donc s'exprimer qu'à partir des comportements de ces 2 classes de phytoplancton. Dans cette section, l'objectif est d'examiner l'importance relative de la variabilité de ces 2 catégories d'espèces sur la

répartition spatio-temporelle de la chlorophylle totale.

En utilisant les 6 années de la simulation *detIni*, nous pouvons, comme précédemment, calculer en chaque point de grille et à chaque pas de temps la moyenne climatologique (des 6 années de simulation) et l'écart-type interannuel (équation 4.1) des concentrations en chlorophylle des diatomées CHL^D et du nanophytoplancton CHL^N . Sur la figure 4.11, nous présentons les séries temporelles du rapport CHL^D/CHL^N entre les moyennes climatologiques des 2 types de phytoplancton (figure du haut), moyennées spatialement dans les provinces de Longhurst 4, 6, 7 et 9. Les séries temporelles de la chlorophylle totale (en moyenne des 6 années de simulation) sont également représentées (figure du bas). L'objectif de cette figure 4.11 est d'évaluer de manière globale, l'impact de la prédominance relative de ces 2 classes de phytoplancton sur la concentration totale en chlorophylle.

Dans l'ensemble, les allures des courbes présentant l'évolution temporelle de la concentration totale en chlorophylle suivent celles du rapport entre les 2 classes de phytoplancton. Cela montre que l'essentiel de la variabilité en chlorophylle du modèle résulte de la variabilité des concentrations en diatomées. Les variations des concentrations en chlorophylle dépendent donc en grande partie du degré de prédominance des diatomées. Dans les provinces 7 et 9, la persistance de faibles concentrations en chlorophylle de surface (totale) peut ainsi s'expliquer par la prédominance permanente du nanophytoplancton. À l'inverse, les fortes productions en chlorophylle qui apparaissent lors du bloom de printemps, vont de pair avec une augmentation des concentrations en diatomées. En hiver, c'est le nanophytoplancton qui est majoritairement prédominant. Les concentrations en chlorophylle de surface sont alors relativement faibles sur l'ensemble du domaine.

Afin d'analyser la distribution de la variabilité interannuelle de la chlorophylle totale entre les 2 classes de phytoplancton, la figure 4.12 affiche cette fois les séries temporelles du rapport $sd_{CHL^D}^t/sd_{CHL^N}^t$ entre les écarts-types interannuels des diatomées et du nanophytoplancton. Cette figure permet d'appuyer le constat précédent. Elle montre en effet que l'essentiel de la variabilité interannuelle du modèle résulte de la variabilité des diatomées. Ce point peut ainsi expliquer en partie que la variabilité interannuelle du modèle PISCES soit limitée aux régions des hautes latitudes.

Récapitulatif

En définitive, nous avons constaté que les distributions spatiales, comme la variabilité interannuelle du modèle, sont principalement le résultat des variations spatiales et temporelles des concentrations en diatomées. Dans le modèle PISCES, les régimes biogéochimiques se différencient donc principalement par le niveau de prédominance des diatomées. Ce résultat s'explique en fait surtout par la différence de taille entre ces 2 classes de phytoplancton. Les diatomées étant plus grosses, elles contiennent plus de chlorophylle. Leur variabilité a donc plus d'impact sur la concentration totale en chlorophylle. Les imperfections sur la répartition spatiale des diatomées peuvent donc expliquer une grande partie des écarts aux observations. Cela peut également être la raison à l'existence de forts contrastes entre les régions. On peut toutefois se questionner de la légitimité d'une représentation dépendant presque exclusivement du degré de prédominance des diatomées. L'insuffisante hétérogénéité spatiale de la simulation *detIni* que nous avons observée au paragraphe précédent pourrait, en effet, être due à une trop faible variabilité du nanophytoplancton.

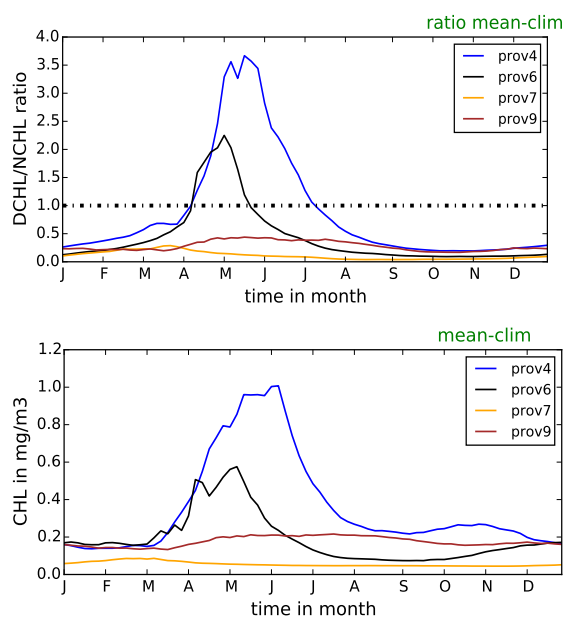


FIGURE 4.11 – Figure du haut : séries temporelles du rapport CHL^D/CHL^N entre les concentrations moyennes de la chlorophylle des diatomées et celles du nanophytoplancton obtenues à partir des 6 années de simulation. Les séries temporelles présentées en moyenne spatiale dans les provinces 4, 6, 7 et 9. La droite en pointillés noirs indique des concentrations identiques pour ces 2 catégories d'espèces. Pour ces mêmes provinces, la figure du bas présente les mêmes séries temporelles pour la chlorophylle totale $CHL^D + CHL^N$.

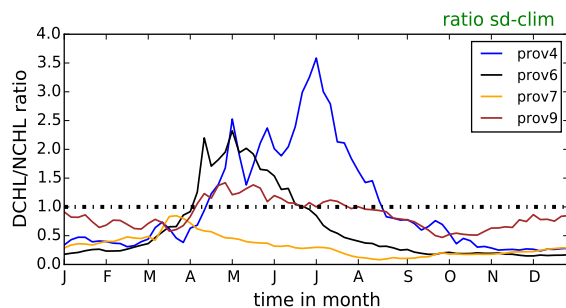


FIGURE 4.12 – Séries temporelles du rapport $sd^t_{CHL^D}/sd^t_{CHL^N}$, entre les écarts-types interannuels des concentrations en chlorophylle des diatomées et celles du nanophytoplancton, moyennées spatialement dans les provinces 4, 6, 7 et 9. La droite en pointillés noirs indique une variabilité identique pour les 2 espèces.

Comparaison avec les données PHYSAT

En complément de la concentration en chlorophylle (totale), les données satellites peuvent également apporter des précisions sur le type d'espèce dont est issue la chlorophylle observée. À partir des données SeaWiFS, il est en effet possible de détecter la signature de différentes espèces de phytoplancton en décomposant le spectre d'absorption

lumineux. En fonction des différents pigments utilisés pour la photosynthèse, la méthode PHYSAT (Alvain et al., 2005) permet ainsi de répartir la chlorophylle de surface parmi 6 espèces : les nanoeucaryotes, les nanoprochlorococcus, les synechococcus, les diatomées, les phaeocystis et les cocolithophoridés. Je précise que déconvoluer des anomalies de réflectance du signal total est un processus difficile. Ces résultats sont donc à prendre de manière plus informative que qualitative. Leur utilisation restera ici très sommaire.

Le modèle PISCES ne possédant que 2 classes de phytoplancton, nous pouvons, en simplifiant, suggérer que le compartiment du nanophytoplancton regroupe les 3 espèces de plus petite taille des données PHYSAT et que celui des diatomées regroupe les 3 autres. Sur la figure 4.13, les cartes de surface de la prédominance des 2 classes de phytoplancton de PISCES sont comparées avec celles des observations SeaWiFS, déduites de la méthode PHYSAT. Les niveaux de prédominance du prochlorococcus et des diatomées sont également présentés.

Cette figure permet de confirmer que la surestimation des quantités de chlorophylle par le modèle, au niveau du Gulf Stream, est principalement due à un bloom de diatomées non existant dans les observations. Dans les données PHYSAT, les diatomées ont, en effet, plutôt tendance à se répartir dans la partie la plus au nord du domaine. Dans cette zone, la variabilité spatiale des observations est associée à une répartition particulièrement « patchy », non simulée par le modèle. L'existence de ces amas de phytoplancton (qui est souvent appelée la *patchiness*) peut s'expliquer autrement que par la turbulence de la circulation océanique. En effet, il a depuis longtemps été montré qu'une telle répartition peut exister dans un océan homogène à cause des interactions et de la compétition entre les différentes espèces (Levin and Segel, 1976).

Sur la figure 4.13, les cartes des données PHYSAT montrent que les organismes de grosse taille sont en majeure partie constitués de diatomées (comparaison entre la carte de la 1^{ère} ligne, 1^{ère} colonne et la carte de la 1^{ère} ligne, 2^{ème} colonne). Au contraire, la carte de prédominance du nanoprochlorococcus (2^{ème} ligne, 1^{ère} colonne) montre que le type d'espèce composant le nanophytoplancton peut varier significativement selon la région considérée. Avec un seul compartiment pour représenter le comportement du phytoplancton de petite taille (la classe nanophytoplancton), le modèle PISCES ne peut évidemment pas faire cette distinction. Chaque espèce possédant un fonctionnement biogéochimique propre, il ne pourra donc pas discerner l'existence de différents comportements issus de la prédominance de plusieurs types d'espèces au sein d'une même classe de phytoplancton. Par exemple, chaque espèce de phytoplancton possède un comportement vertical particulier, résultant notamment de besoins propres en lumière et/ou en nutriments. Les écarts entre les profils verticaux de la simulation *detIni* et ceux des observations (cf. section 4.2.2) pourraient donc, en partie, s'expliquer par la présence de seulement 2 classes de phytoplancton pour simuler la multiplicité des fonctionnements biogéochimiques. Dans les chapitres suivants, nous montrerons que considérer l'incertitude des paramètres biogéochimiques du modèle permet de simuler des comportements pouvant être assimilés à l'existence de différentes espèces de phytoplancton dans une même classe de phytoplancton.

4.4 Conclusion

Dans ce chapitre, nous avons présenté le modèle biogéochimique PISCES en suivant une démarche classique, déterministe, de modélisation biogéochimique. L'estimation de la

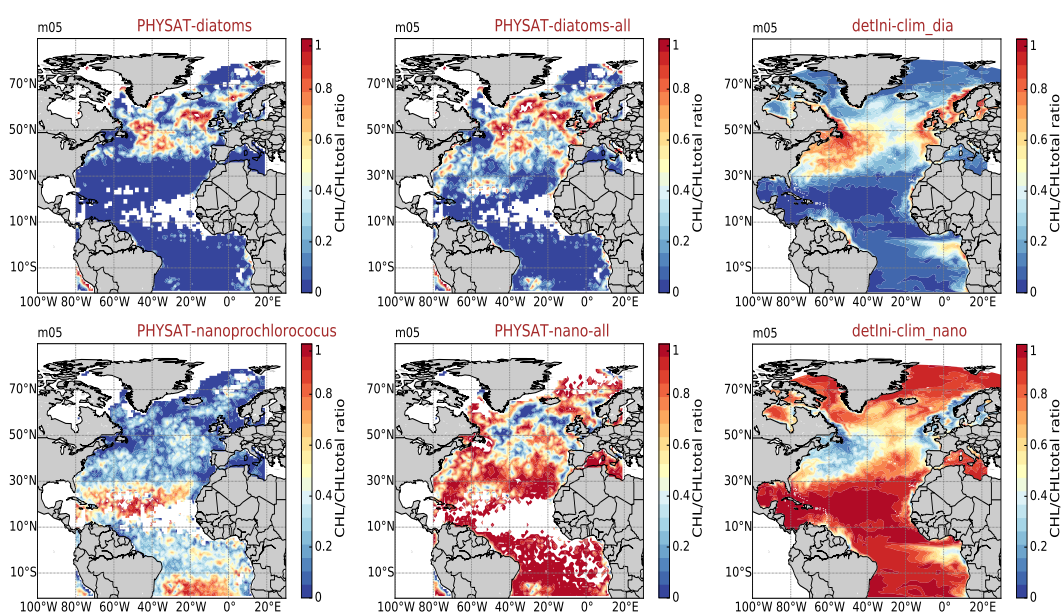


FIGURE 4.13 – Cartes de prédominance des espèces de phytoplancton, présentées au mois de mai 2005 comme le rapport entre la quantité de chlorophylle de l'espèce considérée et la concentration totale de chlorophylle. L'identification des espèces des observations satellites SeaWiFS est faite en 6 espèces selon la méthode PHYSAT. Les organismes représentés sont les diatomées (1^{ère} ligne, 1^{ère} colonne), la somme des 3 espèces de plus grande taille (1^{ère} ligne, 2^{ème} colonne), les nanoeucaryotes (2^{ème} ligne, 1^{ère} colonne) et la somme des 3 espèces de plus petite taille (2^{ème} ligne, 2^{ème} colonne). La contribution relative des diatomées (1^{ère} ligne, 3^{ème} colonne) et du nanophytoplancton (2^{ème} ligne, 3^{ème} colonne) de la simulation *detIni* est également présentée.

production primaire (chlorophylle) affichée ici est équivalente à ce qui peut être aujourd'hui produit en prévision opérationnelle, notamment par le système de prévision européenne MERCATOR-Océan. En raison du peu d'observations disponibles sur la verticale et de leur faible résolution spatio-temporelle, les caractéristiques verticales de cette simulation seront considérées, dans le chapitre suivant, comme des structures de référence que nous chercherons à conserver dans la simulation d'ensemble.

On rappelle les principaux résultats de ce chapitre.

- La transition LOBSTER → PISCES a été justifiée en section 4.2. Dans notre contexte d'étude, nous avons montré que le niveau de complexité du modèle PISCES permet une meilleure représentation des distributions en chlorophylle.
- La variabilité interannuelle de la simulation *detIni*, résultant principalement des incertitudes sur les forçages extérieurs, semble insuffisante pour expliquer les écarts entre la simulation déterministe et les observations SeaWiFS.
- La variabilité spatiale de la chlorophylle de surface est très inférieure à celle des observations SeaWiFS.
- La surestimation de la chlorophylle dans la région du Gulf Stream est le résultat d'un

bloom de diatomées qui n'apparaît pas dans les observations PHYSAT.

- La présence de 2 classes de phytoplancton paraît insuffisante pour représenter la diversité des comportements biologiques de l'océan Atlantique Nord.

Dans le prochain chapitre, nous présenterons une méthode, basée sur l'utilisation de processus stochastiques, permettant de faire évoluer le modèle PISCES vers une version probabiliste. Nous décrirons également les paramétrisations stochastiques que nous avons mises en place pour simuler certaines sources d'incertitude biogéochimiques du modèle PISCES.

Chapitre 5

Mise en place d'une version probabiliste du modèle PISCES

Sommaire

5.1 Simulation explicite d'incertitudes biogéochimiques	89
5.1.1 Une approche stochastique	89
5.1.1.1 Démarche et objectifs	89
5.1.1.2 Génération des perturbations	91
5.1.2 Pour quelles incertitudes ?	93
5.1.2.1 Incertitudes sur les paramètres biologiques	93
5.1.2.2 Incertitudes des échelles non résolues	95
5.2 Ajustement conjugué des paramètres stochastiques et déterministes du modèle	97
5.2.1 Des paramétrisations indissociables du modèle...	98
5.2.2 ...impliquant un réajustement des paramètres déterministes	100
5.2.3 Conclusion partielle	103
5.3 Impacts des paramétrisations stochastiques	103
5.3.1 Répartition de la chlorophylle de surface	104
5.3.2 Structures verticales de la chlorophylle	106
5.4 Conclusion	109

Dans le chapitre précédent, nous avons montré que PISCES était capable de représenter les principaux éléments de la dynamique de grande échelle observables dans les images de la couleur de l'eau SeaWIFS. La modélisation biogéochimique restant basée sur d'importantes approximations (cf. chapitre 2), nous avons toutefois constaté que les distributions de la chlorophylle demeurent imparfaites. Par exemple, un élément évoqué au chapitre précédent pour expliquer certains écarts aux observations est la diversité des espèces. Le modèle PISCES permet en effet uniquement de simuler l'existence de 2 classes de phytoplancton. Ces 2 compartiments sont donc censés englober la multiplicité des comportements biologiques de l'Atlantique Nord, ce qui ne peut évidemment pas être le cas.

Dans ce chapitre, nous montrerons qu'il est possible de tenir compte des incertitudes du modèle PISCES en utilisant une approche probabiliste. Pour cela, on propose de simuler de manière explicite certaines incertitudes biogéochimiques par l'intermédiaire de paramétrisations stochastiques. L'impact de ces incertitudes pourra ensuite être quantifié par une distribution de probabilité, construite à partir des membres d'une simulation d'ensemble. Pour notre étude, l'objectif initial de la prise en compte d'incertitudes est d'améliorer la définition des statistiques d'erreur a priori (contexte d'assimilation de données, cf. chapitre 2). Nous verrons que l'évolution vers un modèle probabiliste permet d'aller plus loin. En plus de pouvoir assigner un niveau d'incertitude aux prévisions, on montrera que cette approche permet en effet d'améliorer la représentation de la chlorophylle de surface par rapport à la simulation déterministe *detIni* (cf. chapitre 4).

La première section de ce chapitre présente la méthode utilisée pour générer des processus aléatoires qui seront ensuite introduits dans le modèle pour simuler 2 sources d'incertitudes biogéochimiques : les incertitudes sur les paramètres biogéochimiques du modèle et les incertitudes dues aux effets des échelles sous-maille non résolues. Pour que les processus aléatoires puissent simuler ces incertitudes, ils doivent être intégrés dans le modèle de manière spécifique. Pour chacune des 2 sources d'incertitude, des paramétrisations stochastiques distinctes ont ainsi été mises en place. Elles seront précisément décrites dans cette section.

Le modèle PISCES ayant été initialement calibré pour une approche déterministe, on montre dans la seconde section de ce chapitre que la transition vers une version probabiliste n'est pas immédiate. Ainsi, avant de réaliser une simulation d'ensemble capable de fournir une représentation de la chlorophylle qualitativement comparable avec les observations SeaWIFS, un réajustement du modèle déterministe a été nécessaire. On suggère pour ce faire en section 5.2.2 de modifier la valeur de certains paramètres du modèle PISCES. Avec ce nouveau jeu de paramètres, on réalise une simulation d'ensemble de 60 membres dont les résultats sur la chlorophylle seront évalués dans la dernière section. On montrera en particulier que la moyenne d'ensemble de cette simulation fournit une meilleure estimation de la chlorophylle de surface que la simulation déterministe *detIni*.

Pour terminer cette introduction, je précise que l'évaluation rigoureuse de la simulation d'ensemble n'est pas l'objet de ce chapitre. L'objectif principal est ici de présenter et de valider la méthode utilisée pour inclure des processus aléatoires dans le modèle PISCES. Une analyse statistique de la simulation d'ensemble fera l'objet du chapitre suivant. Je tiens également à souligner que les résultats de ce chapitre ont été publiés (Garnier et al., 2016). Cet article est consultable en annexe D. En parallèle, la méthode permettant de

générer des processus aléatoires pouvant être inclus dans les équations du modèle a été implémentée dans NEMO. Cette procédure a elle aussi fait l’objet d’une publication dont je suis co-auteur (Brankart et al., 2015).

*I believe that ultimate climatic model will be stochastic,
ie random numbers will appear somewhere in time derivatives
Edward Lorenz (1975)*

5.1 Simulation explicite d’incertitudes biogéochimiques

Dans cette première partie, on détaille la méthode utilisée pour simuler des incertitudes dans le modèle PISCES. La démarche adoptée, le choix des différentes incertitudes simulées, et la description des processus aléatoires qui permettent de générer des perturbations stochastiques sont présentés en section 5.1.1. La manière dont les processus aléatoires sont introduits dans le modèle (i.e les paramétrisations stochastiques) pour simuler les 2 différentes classes d’incertitude est explicitée en section 5.1.2.

5.1.1 Une approche stochastique

5.1.1.1 Démarche et objectifs

L’objectif de cette première partie est de préciser le pourquoi et le comment de la démarche probabiliste présentée dans ce chapitre. Tout d’abord, il me semble important de garder à l’esprit que le cadre de cette thèse reste l’assimilation de données de la couleur de l’eau. Le but initial de cette étude était donc de réaliser une simulation d’ensemble capable de générer de la dispersion sur la chlorophylle, représentative de l’incertitude des prévisions du modèle. Les caractéristiques statistiques de cette dispersion permettent ensuite de définir la matrice de covariances d’erreur pour l’assimilation de données. Dans ce contexte, nous avons porté notre attention sur les moyens de générer une dispersion en accord avec les observations que l’on souhaite intégrer dans le modèle. La simulation d’ensemble sera donc considérée comme cohérente si son niveau de dispersion permet d’expliquer (au moins en partie) les écarts aux observations. En d’autres termes, si la simulation d’ensemble “capture” la majorité des observations. Au départ de ce travail, la méthode à utiliser pour produire cette simulation d’ensemble restait cependant une question ouverte. La problématique à laquelle ce chapitre tentera de donner des éléments de réponse peut ainsi s’écrire : quelles sont les sources d’incertitudes à prendre en compte et comment les représenter afin de réaliser une simulation d’ensemble dont la dispersion permettra de définir des statistiques d’erreur cohérentes avec les observations de la couleur de l’eau SeaWIFS ?

Jusqu’ici, de nombreuses études se sont penchées sur l’impact de la physique sur les processus biogéochimiques. Il est en effet fondamental. Nous avons d’ailleurs évoqué à plusieurs reprises l’importance de la couche de mélange ou de la circulation océanique sur la production primaire. Nous avons également montré au chapitre 3 qu’un effort conséquent avait été fait pour se munir d’une contrainte physique réaliste. Toutefois, sans remettre en cause le rôle des incertitudes de la partie physique du modèle, notre choix s’est porté uniquement sur l’impact des incertitudes des formulations biogéochimiques du modèle PISCES. L’amplitude des écarts entre les distributions de la chlorophylle simulées par les modèles

LOBSTER et PISCES, ainsi que l'analyse de la variabilité interannuelle de la simulation *detIni*, nous ont, du reste, encouragés à suivre cette démarche (cf. chapitre 4). Pour justifier notre approche, nous devons donc vérifier que les incertitudes, sur la formulation des principaux processus à l'origine de la production primaire, permettent de produire une dispersion aux propriétés satisfaisantes pour l'assimilation de données.

Comme énoncé au chapitre 1, le choix des incertitudes considérées et d'une approche stochastique pour les simuler s'appuie en grande partie sur les travaux antérieurs de l'équipe MEOM. En Atlantique Nord, les travaux de Doron et al. (2011) ont mis en évidence la variabilité spatiale (et donc l'incertitude) de quelques paramètres biologiques clés comme le taux de croissance et le broutage. Disposer de différentes valeurs de paramètres permettrait donc d'améliorer la représentation de la chlorophylle. En parallèle, Brankart (2013) a montré que la non-linéarité des équations (dans ce cas, l'équation d'état de l'eau de mer) avait un impact non négligeable sur l'estimation grande échelle des grandeurs physiques (dans ce cas, la densité). Les équations qui régissent les processus biogéochimiques étant elles aussi fortement non linéaires, ces effets dus aux échelles non résolues doivent également impacter l'estimation de la chlorophylle du modèle PISCES. Je précise que cette source d'incertitude doit être distinguée des impacts de la non-représentation des processus de petite échelle (turbulence) sur la biogéochimie. Malgré cette distinction, nous sommes conscient que l'existence d'échelles non résolues en biogéochimie est en partie due à la physique. En réalité, nous sommes incapable de discriminer les raisons de la non-résolution des échelles biogéochimiques. Il est cependant important de garder à l'esprit que l'on s'intéressera ici uniquement aux conséquences de la non-linéarité des équations sur l'estimation des variables biogéochimiques dans un modèle au $1/4^\circ$. Cette source d'incertitude sera dans la suite désignée sous le terme d'incertitude due aux échelles biogéochimiques non résolues.

Ces 2 études nous ont ainsi orientés vers la représentation des incertitudes des paramètres biogéochimiques du modèle et de celles dues aux échelles biogéochimiques non résolues. Le choix d'une approche stochastique pour la biogéochimie marine a pour sa part été motivé par des études récentes telles que Dowd (2011) et Weir et al. (2013), qui abordent sous cet aspect le problème des incertitudes dans la modélisation des écosystèmes marins. Outre en météorologie, qui a initié l'élaboration de modèles stochastiques-dynamiques (Palmer, 2001), il faut préciser que la modélisation stochastique constitue une approche émergente dans l'ensemble de la communauté géo-scientifique. Les travaux de Juricke et al., 2013; Juricke and Jung, 2014 utilisent une approche stochastique pour simuler les incertitudes des modèles de glace. Des phénomènes climatiques comme la MAO¹ (Dijkstra et al., 2008) ou l'ENSO² (Kleeman, 2008), dont la variabilité est en général approximativement estimée par les modèles déterministes, commencent également à être modélisés selon cette démarche. Dans un contexte plus océan, les fluctuations des flux air-mer (Williams, 2012) ou l'impact des forçages extérieurs sur les variations de la MOC³ (Monahan et al., 2008; Lorenzo et al., 2012) sont désormais régulièrement considérés de manière stochastique. L'essor des travaux sur le sujet démontre qu'une approche probabiliste/stochastique, telle que celle présentée ici, constitue une démarche prometteuse aux applications diverses. C'est ce que nous essaierons également de montrer dans le cas de la biogéochimie marine.

1. Oscillation Atlantique Multidécennale

2. El Niño Southern Oscillation

3. Meridional Overtuning circulation

Nous terminerons cette section par un récapitulatif des étapes, qui ont été nécessaires avant de pouvoir réaliser une simulation d'ensemble contenant un niveau d'incertitude cohérent avec les observations SeaWIFS. Nous essaierons toujours de justifier au mieux les choix qui ont été faits. Toutefois, un certain nombre d'expériences de sensibilité ont été nécessaires pour calibrer les paramétrisations stochastiques. Par souci de lisibilité, elles ne seront pas montrées dans ce document.

Je rappelle que la philosophie générale des sections qui vont suivre est de présenter, dans sa totalité et le clairement possible, la démarche utilisée pour mettre en place une version stochastique du modèle PISCES. Pour valider cette nouvelle configuration, nous vérifierons qu'elle conserve, en ce qui concerne la répartition de la chlorophylle, la dynamique grande échelle de la simulation déterministe.

Dans les prochaines sections, nous détaillerons les étapes :

1. Mise en place de la méthode pour générer des processus aléatoires indépendants en chaque point de grille du modèle.
2. Prise en compte de corrélations spatiales. L'objectif est de considérer des structures spatiales de corrélation d'erreur.
3. Mise en place des paramétrisations stochastiques. Le but est de proposer des formulations pour intégrer les processus aléatoires de façon à ce qu'ils simulent effectivement les sources d'incertitude recherchées.
4. Ajustement du niveau d'incertitude du modèle. Cette étape est la conséquence du couplage stochastique/déterministe. La modification du niveau d'incertitude du modèle due à l'ajout de paramétrisations stochastiques nous amènera à réajuster certains paramètres du modèle déterministe (cf. section 5.2).
5. Réalisation d'une simulation d'ensemble.
6. Évaluation de la cohérence des structures moyennes de la chlorophylle de la simulation d'ensemble par comparaison avec la simulation déterministe *detIni*.

5.1.1.2 Génération des perturbations

La première étape de la mise en place de paramétrisations stochastiques consiste à générer, en chaque point de grille et à chaque pas de temps, des nombres aléatoires $\xi(t)$ qui seront ensuite introduits dans les équations du modèle (cf. section 5.1.2). Pour qu'elles aient un sens, ces perturbations aléatoires doivent reposer sur des hypothèses a priori définissant le degré d'incertitude à représenter (écart-type) ainsi que leurs structures spatiales et temporelles de corrélations. Si l'on considère par exemple l'échelle de discrétisation temporelle du modèle (1 pas de temps toutes les 40 minutes), les concentrations en chlorophylle et les incertitudes qui leur sont associées ne vont pas radicalement évoluer entre 2 pas de temps consécutifs. De la même façon, les régimes biogéochimiques ne sont pas fondamentalement différents entre 2 points de grille voisins. Pour être plus précis, les grandeurs à l'origine des incertitudes (p.ex. taux de croissance) possèdent une certaine cohérence spatiale et temporelle. Les perturbations utilisées pour simuler les incertitudes doivent donc rendre compte de ces corrélations. Cela signifie que l'on doit travailler avec des séries aléatoires corrélées. Dans notre cas, ces séries aléatoires seront en fait des processus stochastiques.

En s'inspirant des méthodes utilisées par le CEPMMT (Centre Européen pour les Prévisions Météorologiques à Moyen Terme) pour la prédiction d'ensemble météorologique (Buizza et al., 1999; Palmer et al., 2005), toutes les perturbations stochastiques $\xi(t)$ sont générées à partir de processus autorégressifs d'ordre 1 (AR(1)) définis par l'équation 5.1. Cette procédure permettant de générer des processus aléatoires est identique à celle utilisée dans Brankart (2013) ainsi que dans le projet SANGOMA (Candille et al., 2015) pour les expériences ensemblistes d'assimilation de données altimétriques. Elle a, depuis, été mise en place dans le système NEMO (Brankart et al., 2015).

$$\xi(t_k) = \varphi \xi(t_{k-1}) + \sqrt{1 - \varphi^2} w \quad (5.1)$$

$$\varphi = e^{-1/\tau} \quad (5.2)$$

Le paramètre w est un bruit blanc (nombre aléatoire) gaussien de moyenne nulle et d'écart-type σ . Puisque l'équation 5.1 est linéaire, les distributions des perturbations $\xi(t)$ seront également gaussiennes. Le paramètre $\varphi \leq 1$ rend compte selon l'équation 5.2 d'un temps de décorrélation τ qui représente le temps au bout duquel le poids (l'influence) d'une perturbation est de $1/e$. Dans cette étude, un temps de décorrélation unique $\tau = 30$ jours sera considéré. Je précise que tous les paramètres régissant les processus stochastiques seront récapitulés plus loin dans ce chapitre (table 5.1). En suivant cette procédure, on peut générer en chaque point de grille autant de processus stochastiques indépendants que l'on souhaite. Afin de conserver des structures verticales en accord avec la simulation déterministe (cf. section 5.3.2), les perturbations sont définies invariantes suivant l'axe vertical. À chaque pas de temps, n cartes horizontales de perturbations, telles que celles présentées en figure 5.1, peuvent ainsi être produites. Jusqu'ici, en un point de grille du modèle, l'intensité des perturbations ne dépend que du bruit blanc aléatoire w et de la valeur des perturbations aux pas de temps précédents en ce même point de grille. Comme expliqué précédemment, il est nécessaire d'ajouter des structures spatiales de corrélation à ces perturbations. Nous appliquons pour cela un filtre horizontal Laplacien défini par l'équation 5.3.

$$\xi'_{i,j}(t_k) = 0.5 \xi_{i,j}(t_k) + 0.125 (\xi_{i-1,j}(t_k) + \xi_{i+1,j}(t_k) + \xi_{i,j-1}(t_k) + \xi_{i,j+1}(t_k)) \quad (5.3)$$

où i, j sont les coordonnées des points de grille du modèle. À partir des $\xi(t_k)$, une nouvelle carte de perturbations $\xi'(t_k)$ peut être construite, dans laquelle l'intensité des perturbations dépend également de celle des 4 points de grille voisins. Cette opération est réalisée en chaque point de grille. Si on souhaite élargir les corrélations spatiales, l'opération peut être ainsi réitérée plusieurs fois par pas de temps. Dans cette étude, une corrélation spatiale de 5 points de grille est considérée. L'équation 5.3 sera donc appliquée 5 fois avant d'introduire les perturbations dans le modèle. L'impact de l'étendue des corrélations spatiales sur la répartition de l'intensité des perturbations est illustré en figure 5.1. Plus les corrélations spatiales sont importantes, plus on observera une répartition « en amas » de l'intensité des perturbations. Cette méthode étant assez coûteuse numériquement, elle reste plutôt adaptée à de faibles corrélations spatiales.

Toutes les perturbations utilisées dans la suite seront générées à partir d'AR(1) décrit par l'équation 5.2. Elles sont toutes issues d'une distribution gaussienne, c'est-à-dire $\xi(t) \in \mathcal{N}(\mu = 0, \sigma)$. La forme des distributions de probabilité a priori ne repose donc sur aucune hypothèse.

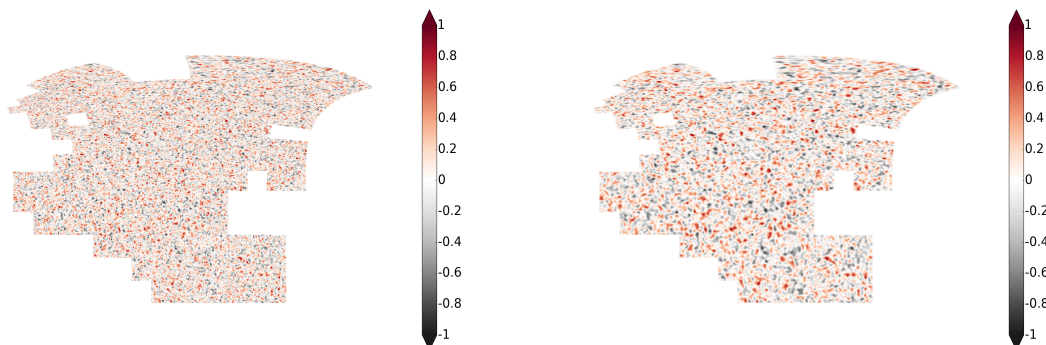


FIGURE 5.1 – Cartes de perturbations générées avec une corrélation spatiale de 5 points de grille (gauche) et de 12 points de grille (droite). Ces perturbations sont tirées dans une gaussienne de moyenne nulle et d'écart-type 0.3.

5.1.2 Pour quelles incertitudes ?

Une fois la méthode permettant de générer des processus aléatoires mise en place, il est nécessaire de définir les sources d'incertitude à simuler. Étant donné la multitude des sources potentielles d'incertitude, il était évidemment impossible de toutes les considérer. Nous avons dû faire un choix, qui reste sujet à discussion. Dans un contexte d'assimilation de données de la couleur de l'eau, l'objectif était de permettre de générer une dispersion capable d'expliquer les écarts aux observations SeaWIFS. Les sources d'incertitude impactant directement la production de la chlorophylle sont donc celles qui nous ont principalement intéressés.

Dans cette section, nous présenterons les 2 types de paramétrisations stochastiques permettant de simuler 2 sources d'incertitude biogéochimiques du modèle : les incertitudes sur les paramètres du modèle et les incertitudes dues aux échelles biogéochimiques non résolues.

5.1.2.1 Incertitudes sur les paramètres biologiques

L'une des principales difficultés pour modéliser le comportement des écosystèmes marins provient du caractère empirique des équations biogéochimiques. À partir d'un modèle, seules quelques variables et quelques paramètres sont supposés représenter des phénomènes complexes comme l'adaptation biologique ou la diversité des espèces. L'existence d'uniquement 2 différentes classes de phytoplancton dans la structure du modèle PISCES permet d'expliquer ainsi certains points de divergence entre le modèle et les observations.

Sous l'hypothèse qu'une diversité biologique peut s'exprimer à partir des équations régissant ces 2 compartiments biologiques, l'existence d'une variété de comportements biogéochimiques ne peut être simulée qu'en utilisant plusieurs jeux de paramètres (Doron

et al., 2011). Pour cela, on propose donc de considérer certains des paramètres du modèle comme incertains. Une liste de ces paramètres sera présentée à la fin de ce paragraphe. On s'intéresse tout d'abord à la démarche utilisée pour les perturber. En utilisant les processus stochastiques générés à partir de l'équation 5.1, les paramètres du modèles sont perturbés comme indiqué par les équations 5.4 et 5.5 :

$$\text{Déterministe : } \frac{\partial C}{\partial t} \Big|_{bio} = SMS(C, u, p, t) \quad (5.4)$$

$$\text{Stochastique : } \frac{\partial C}{\partial t} \Big|_{bio} = SMS(C, u, p', t) \quad \text{avec } p' = \exp[\xi(t)]p \quad (5.5)$$

où p est le vecteur contenant l'ensemble des paramètres du modèle biogéochimique (déterministe) et p' est le vecteur des paramètres perturbés. Le terme u représente les forçages extérieurs (vent, intensité lumineuse...). Avec ces notations, $\xi(t)$ est un vecteur de perturbations dont la taille correspond au nombre m de paramètres du modèle, i.e $\xi(t) = [\xi_1(t), \xi_2(t), \dots, \xi_m(t)]$ où tous les $\xi_m(t)$ sont issus de processus aléatoires indépendants. Un écart-type nul est appliqué aux perturbations agissant sur les paramètres dont on souhaite conserver une valeur unique, déterministe et un écart-type de $\sigma = 0.3$ est appliqué aux autres. Le choix d'un écart de 0.3% est le résultat de plusieurs tests de sensibilité. Il a été également légitimé par différentes études traitant de la variabilité des paramètres biogéochimiques (eg. Geider et al., 1997; Fennel et al., 2001; Mattern et al., 2008; Hu et al., 2012). Je me permets d'ajouter qu'un écart-type de 30% sous-estime très certainement l'incertitude des paramètres par comparaison aux gammes de valeurs pouvant être évoquées dans ces travaux.

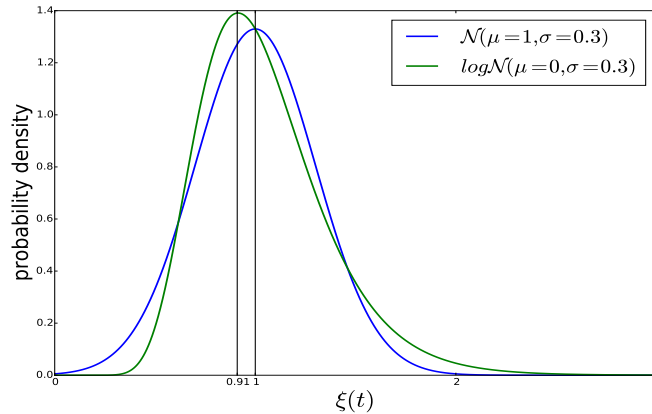


FIGURE 5.2 – Comparaison entre une distribution de probabilité log-normale $\log\mathcal{N}(\mu = 0, \sigma = 0.3)$, utilisée pour perturber les paramètres du modèle, et une distribution gaussienne $\mathcal{N}(\mu = 1, \sigma = 0.3)$.

Afin d'éviter que les paramètres puissent prendre des valeurs négatives, les perturbations gaussiennes $\xi_i(t)$ sont élevées à l'exponentielle. Cette formulation transforme donc les distributions initialement gaussiennes des perturbations en distributions log-normales. Comme on peut le voir en figure 5.2, pour un écart-type de 0.3, la forme de la distribution log-normale $\log\mathcal{N}(\mu = 0, \sigma = 0.3)$ est très proche de celle d'une distribution gaussienne $\mathcal{N}(\mu = 1, \sigma = 0.3)$. La moyenne de la distribution $\log\mathcal{N}(\mu = 0, \sigma = 0.3)$ est

$\exp(0.045) \simeq 1.04$. Par conséquent, la moyenne d'un paramètre perturbé par une perturbation aléatoire de distribution $\log\mathcal{N}(\mu = 0, \sigma = 0.3)$ est sensiblement égale à sa valeur déterministe. La forme log-normale prescrite par les paramétrisations stochastiques n'introduit donc pas de biais statistique significatif.

Pour chaque paramètre que l'on souhaite considérer comme incertain, une carte horizontale de bruit blanc log-gaussien $\xi(t)$ semblable à celles présentées en figure 5.1 est donc introduite à chaque pas de temps dans les équations du modèle. En s'appuyant sur plusieurs études dont Lima and Doney, 2004; Fennel et al., 2006; Doron et al., 2011, 2013; Hu et al., 2012, ainsi que sur des expériences de sensibilité, 7 processus stochastiques (d'écart-type non nul) indépendants sont générés pour perturber 7 paramètres biogéochimiques intervenant dans l'expression de la croissance du phytoplancton (chapitre 3, équation 3.10).

- Le taux de croissance maximal à 0°C μ_{max}^0 des 2 classes de phytoplancton,
- La sensibilité à la température b du taux de croissance maximal des 2 classes de phytoplancton,
- La sensibilité à la température du taux de broutage maximal du zooplancton b'_Z ,
- L'efficacité photosynthétique (P-I slope) α^N et α^D de chaque phytoplancton,
- La dépendance du taux de croissance à la durée du jour f_{day}^N et f_{day}^D de chaque phytoplancton.

Les paramètres b et μ_{max}^0 interviennent dans le calcul du taux de croissance maximal μ_P (équation 3.17). La formulation de ce taux de croissance étant identique pour les 2 classes de phytoplancton, les perturbations associées à ces 2 paramètres le seront également. À noter que, puisque $\mu^P \propto \mu_P L_{Lim}$, on ne peut pas mathématiquement déterminer a priori quel terme a été perturbé. Perturber le paramètre μ_{max}^0 de l'expression de $\mu_P = \mu_{max}^0 \exp(bT)$ est équivalent à perturber avec les mêmes caractéristiques le terme limitant L_{Lim} . Ainsi construites, ces paramétrisations stochastiques permettent donc de simuler une incertitude globale sur la croissance du phytoplancton. L'incertitude sur les 2 classes de phytoplancton diffère par l'intermédiaire des 2 termes $f_{day}^{P=(N,D)}$ et $\alpha^{P=(N,D)}$.

Bien évidemment d'autres jeux de paramètres auraient pu être considérés. L'approche utilisée ici doit donc être perçue comme une première étape, principalement destinée à mettre en valeur la capacité des paramétrisations stochastiques à simuler des incertitudes biogéochimiques.

5.1.2.2 Incertitudes des échelles non résolues

L'utilisation de processus stochastiques s'est avérée être une solution appropriée pour simuler les effets des échelles non résolues issus de la non-linéarité des termes d'advection (Frederiksen et al., 2012) et de l'équation d'état de l'eau de mer (Brankart, 2013) dans des modèles de circulation océanique. De manière similaire, les fluctuations sous-maille non résolues des variables biogéochimiques vont également impacter l'estimation de la production primaire de plus grande échelle.

Imaginons que la figure 5.3 soit un champ de concentrations en chlorophylle que l'on doit considérer en un point de grille représenté par un carré noir. À cause de la discrét-

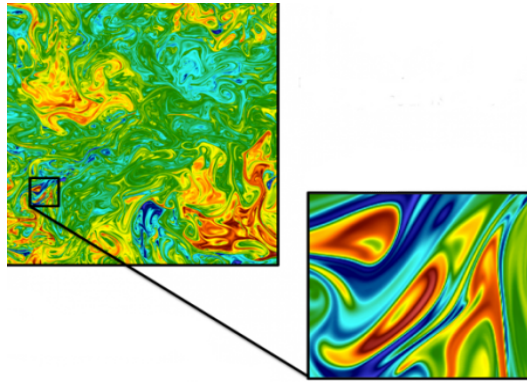


FIGURE 5.3 – Exemple de fluctuations sous-maille pouvant exister à l'intérieur d'un point de grille d'un modèle.

tisation numérique, les fluctuations à l'intérieur de ce point de grille ne peuvent pas être représentées. Les équations SMS du modèle s'appliqueront à une valeur moyenne de la chlorophylle. Mais comme ces équations sont non linéaires, $SMS(\bar{C}) \neq \overline{SMS(C)}$ avec C le vecteur de la concentration en chlorophylle. Ne pas prendre en compte les variations de la chlorophylle va induire un biais dans le calcul des termes des équations SMS et représente donc une importante source d'incertitude. À noter que les écarts d'hétérogénéité spatiale entre les observations SeaWiFS et le modèle au $1/4^\circ$ ont été illustrés au chapitre précédent. En définitive, selon le degré de concavité/convexité des équations SMS du modèle biogéochimique, la non-représentation des fluctuations sous-maille induit une sous/sur estimation du vecteur d'état. Pour simuler cet effet dû aux échelles non résolues, l'idée est de générer à partir des processus stochastiques une perturbation δC capable de représenter l'écart entre $\overline{SMS(C)}$ et $SMS(\bar{C})$. La fonction SMS est ensuite appliquée non plus à la valeur \bar{C} mais à une valeur $\bar{C} + \delta\bar{C}$ comme exprimé par les équations 5.6 et 5.7.

$$\frac{\partial C}{\partial t} \Big|_{bio} = \frac{1}{2} \left[SMS(\bar{C} + \delta\bar{C}, u, p, t) + SMS(\bar{C} - \delta\bar{C}, u, p, t) \right] \quad (5.6)$$

$$\delta\bar{C} = \xi(t)\bar{C} \quad (5.7)$$

Grâce à cette formulation, l'intensité de la réponse aux perturbations est directement proportionnelle au degré de non-linéarité du terme SMS . Pour une fonction parfaitement linéaire, l'ajout de perturbations n'aura pas d'impact sur le résultat.

La dernière étape est de définir une paramétrisation à $\delta\bar{C}$ représentant des fluctuations non résolues adaptées à la résolution du modèle. Pour définir $\delta\bar{C}$, on réalise un filtrage $1/12^\circ \rightarrow 1/4^\circ$ des données satellites SeaWiFS moyennées à 5 jours. Par mois, on obtient ainsi 6 cartes de chlorophylle de surface filtrée au $1/4^\circ$ sur la grille au $1/12^\circ$ de SeaWiFS. En chaque point de grille et à chaque pas de temps, on peut ensuite calculer l'anomalie a_t entre la concentration en chlorophylle de la haute résolution $chl_{1/12^\circ}$ et celle de la basse résolution $chl_{1/4^\circ}$ ($a_t = chl_{1/12^\circ} - chl_{1/4^\circ}$). L'écart-type mensuel de cette anomalie peut ensuite être calculé à partir de l'expression : $sd_a = \sqrt{(a_t - \bar{a}_t)^2}$. Comme la moyenne des

anomalies est nulle, sd_a représente l'écart-type entre les concentrations de chlorophylle haute et basse résolution. La figure 5.4 représente les cartes de cet écart-type normalisé par la moyenne de chlorophylle obtenue après filtrage (basse résolution).

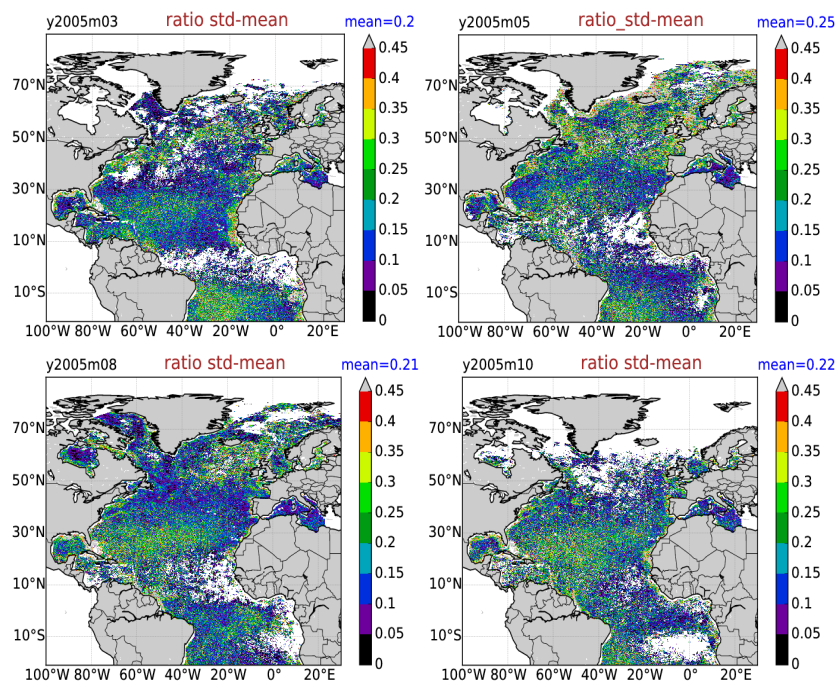


FIGURE 5.4 – Cartes du rapport $sd_a/chl_{1/4^\circ}$ entre l'écart-type sd_a des anomalies des concentrations en chlorophylle des données SeaWIFS avant et après filtrage au $1/4^\circ$ et la moyenne de chlorophylle SeaWIFS au $1/4^\circ$ après filtrage $chl_{1/4^\circ}$. La ligne du haut présente les résultats pour les mois de mars (1^{ère} colonne) et de mai (2^{ème} colonne). La ligne du bas présente les résultats pour les mois d'août (1^{ère} colonne) et d'octobre (2^{ème} colonne).

L'écart-type des fluctuations apparaît globalement proportionnel à la concentration en chlorophylle selon un facteur de 10-30%. Ce taux est relativement constant au cours du temps et présente une répartition spatiale assez homogène. On se contentera donc ici d'appliquer un écart-type constant $\sigma = 0.2$ (20%) sur l'ensemble du domaine pour définir les distributions de probabilité des perturbations stochastiques $\xi(t)$ (cf. équation 5.7).

5.2 Ajustement conjugué des paramètres stochastiques et déterministes du modèle

Comme précédemment mentionné, les paramétrisations stochastiques ont été mises en place afin de réaliser une simulation probabiliste du modèle PISCES contenant un niveau d'incertitude lui permettant d'être comparée avec les observations de la couleur de l'eau. La pertinence des paramétrisations stochastiques est donc examinée en considérant que chaque membre doit afficher une répartition de grande échelle en accord avec les images SeaWIFS. Avant d'atteindre cet objectif, nous montrerons dans la partie 5.2.1 que l'introduction de perturbations stochastiques dans le modèle, produit un fort effet moyen qui stimule

exagérément la production phytoplanctonique. Pour compenser cet effet systématique, nous proposerons en section 5.2.2 de redéfinir la valeur initiale du jeu de paramètres utilisé au chapitre précédent pour réaliser la simulation déterministe *detIni*.

5.2.1 Des paramétrisations indissociables du modèle...

À partir des perturbations stochastiques présentées en section précédente, une version probabiliste de PISCES est mise en place. Ce modèle probabiliste comprend un compartiment stochastique (cf. section 5.1) en complément à la dynamique déterministe du modèle biogéochimique (cf. chapitre précédent). En plus des paramètres déterministes initiaux, la version probabiliste de PISCES se compose donc désormais d'un jeu de paramètres élargi par ceux régissant les processus stochastiques.

Dans un premier temps, l'impact des paramétrisations stochastiques est examiné à partir d'une simulation d'ensemble de 3 membres (*stoIni*). Cette simulation est réalisée sur l'année 2005 et reste basée sur les paramètres déterministes initiaux de la simulation *detIni* (présentés au chapitre 3). Les paramétrisations stochastiques sont définies à partir des processus aléatoires $\xi(t) \in \text{Log}\mathcal{N}(\mu = 0, \sigma = 0.3)$ pour simuler les incertitudes sur les paramètres du modèle et des processus aléatoires $\xi(t) \in \mathcal{N}(\mu = 0, \sigma = 0.2)$ pour simuler les effets des échelles non résolues. Un temps de décorrélation $\tau = 1$ mois et une corrélation spatiale de 5 points de grille sont appliqués aux processus stochastiques. La valeur de ces différents paramètres est récapitulée en table 5.1. Les 3 membres sont initialisés au 01/01/2005 à partir de l'état biogéochimique de la simulation *detIni*. La figure 5.5 compare la chlorophylle de surface des 3 membres de cette simulation avec la simulation déterministe *detIni* et les données de la couleur de l'eau SeaWIFS pour le 15 mai et le 27 octobre 2005.

Par rapport à la simulation déterministe *detIni*, on constate que les paramétrisations stochastiques augmentent fortement les quantités en chlorophylle de surface. Cette augmentation s'observe principalement pendant le bloom de printemps, lorsque la dynamique biologique est la plus intense. Par conséquent, pendant cette période, la chlorophylle de surface est largement surestimée sur la quasi-totalité du domaine (voir aussi la figure 5.6). Les membres de la simulation *stoIni* ne représentent donc pas un état biogéochimique cohérent avec les données SeaWIFS. À noter que cet effet d'amplification est systématique. Il apparaît dès lors que des perturbations stochastiques sont introduites dans les équations du modèle. Des simulations additionnelles ont, en effet, toutes affichées cette caractéristique.

En section 5.1.2, on a montré que la forme de la distribution des perturbations (induite par la formulation exponentielle) restait très proche d'une gaussienne $\mathcal{N}(\mu = 1, \sigma = 0.3)$. La forte surestimation de la chlorophylle ne peut donc pas s'expliquer par l'allure des distributions de probabilité des perturbations utilisées pour simuler les incertitudes sur les paramètres biogéochimiques. En fait, cet effet est principalement le résultat des paramétrisations stochastiques des échelles non résolues. Lorsque l'on perturbe des équations non linéaires, même une distribution gaussienne parfaitement symétrique induira effectivement un effet moyen (biais), dépendant du degré de concavité/convexité des équations. La prise en compte des incertitudes des échelles non résolues produit donc un effet moyen qui conduit ici à l'augmentation des concentrations en chlorophylle.

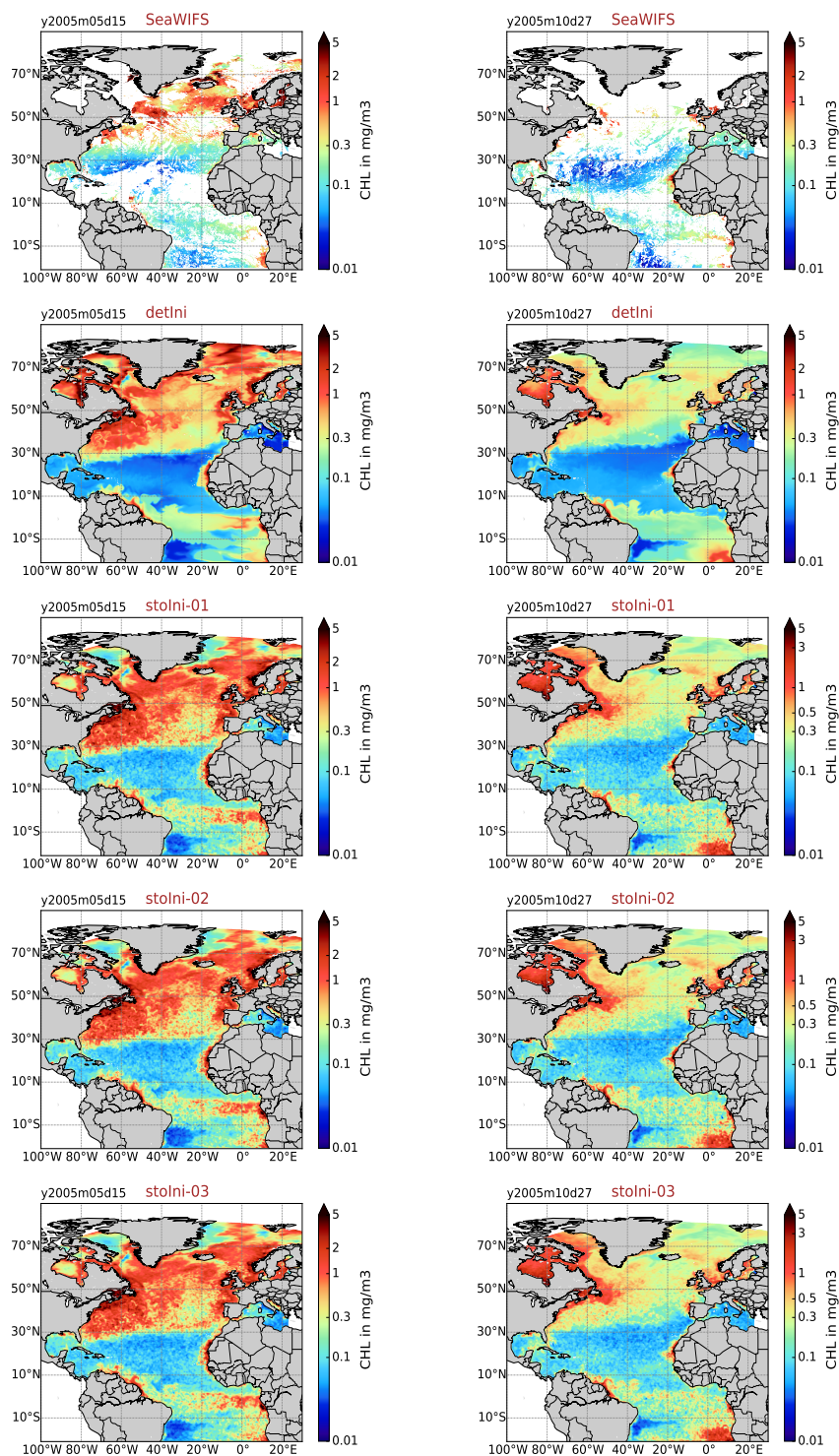


FIGURE 5.5 – Cartes de chlorophylle de surface des observations SeaWIFS (1^{ère} ligne), de la simulation déterministe *detIni* (2^{ème} ligne) et des 3 membres de la simulation d'ensemble *stoIni* (lignes 3 à 5). Les résultats sont présentés en moyenne de 5 jours pour le 15 mai (1^{ère} colonne) et le 27 octobre (2^{ème} colonne) de l'année 2005.

5.2.2 ...impliquant un réajustement des paramètres déterministes

Tel que le modèle probabiliste a été présenté, la simulation déterministe *detIni* devient une simulation stochastique dans laquelle la valeur des paramètres stochastiques est fixée à 0 (voir table 5.1). Dans le modèle déterministe, les paramètres biogéochimiques sont préalablement calibrés (parfois arbitrairement) pour obtenir un état moyen de grande échelle le plus proche possible des observations (Aumont et al., 2015). Cela explique la similitude entre les courbes jaunes et bleues de la figure 5.6.

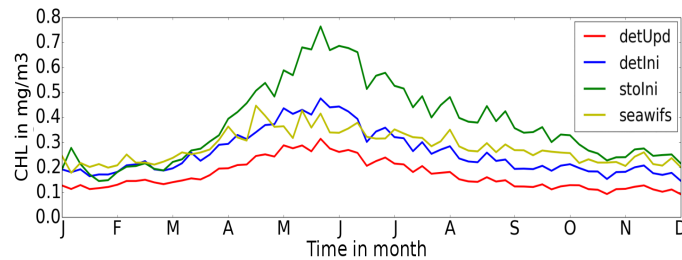


FIGURE 5.6 – Séries temporelles présentant les concentrations de la chlorophylle de surface moyennées sur l'ensemble du domaine. Les résultats sont présentés pour la simulation déterministe *detIni* (courbe bleue), les données SeaWIFS (courbe jaune), un membre de la simulation perturbée *stoIni* (courbe verte) et la simulation déterministe *detUpd* établie avec le jeu de paramètres calibré pour l'approche probabiliste (courbe rouge).

À cause du réglage des paramètres, certaines incertitudes du modèle (qui auraient pu induire de larges écarts avec les observations) peuvent être indirectement compensées (p.ex. un taux de croissance très faible pour rendre compte de la prédominance d'espèces de petite taille non simulée par le modèle). La valeur a priori des paramètres biogéochimiques (valeur déterministe) ne doit donc pas être considérée comme absolue. Elle reste le résultat d'un choix initial dépendant du niveau d'incertitude du modèle. Cela explique par exemple pourquoi 2 modèles de complexité différente peuvent posséder des taux de croissance aux valeurs très différentes.

Lorsque les paramètres stochastiques sont ensuite différents de 0 (simulation *stoIni*), certaines incertitudes du modèle deviennent explicitement simulées. Le degré d'incertitude du modèle est alors diminué. Certains paramètres initialement calibrés en fonction du niveau de complexité (d'incertitude) du modèle doivent donc être réajustés pour s'accorder à la formulation stochastique, c'est-à-dire au nouveau niveau d'incertitude. Ainsi, pour contrebalancer l'effet moyen des paramétrisations stochastiques, on propose de réajuster la valeur initiale (adaptée à l'approche déterministe) des paramètres perturbés.

Le tableau 5.1 présente les valeurs des paramètres biogéochimiques perturbés et des paramètres stochastiques. Les valeurs ajustées sont utilisées pour réaliser les simulations *detUpd* et *stoUpd* (*Upd* faisant référence à l'utilisation de paramètres « updated ») qui sont présentées dans la suite de ce paragraphe. Dans le cadre de cette thèse, recalibrer le plus finement possible les paramètres déterministes du modèle n'est pas l'objectif recherché. Ce serait de plus un procédé très coûteux numériquement. Les valeurs des paramètres proviennent ainsi majoritairement d'expériences de sensibilité et d'un compromis entre 1) des valeurs réalistes et 2) des valeurs permettant de conserver un niveau d'incertitude

suffisant. Je précise que le réajustement des paramètres respecte les gammes de valeurs de différents modèles présentées par Fennel et al. (2006). Bien évidemment, ce nouveau jeu de paramètres ne doit pas être perçu comme optimal et pourra être amélioré dans le futur.

RÉGLAGE DES PARAMÈTRES				
Paramètres déterministes du modèle PISCES	Simulations			
	detIni	stoIni	detUpd	stoUpd
Efficacité photosynthétique du nano : α^N	2	2	4	4
Efficacité photosynthétique des diatomées : α^D	2	2	4	4
Taux de croissance maximum à 0°C du phyto : μ_{max}^0	0.6	0.6	0.5	0.5
Sensibilité à la température de la croissance du phyto : b	0.064	0.064	0.045	0.045
Sensibilité du broutage par le zoo à la température : b'_Z	0.076	0.076	0.084	0.084
Dépendance du nano à la durée du jour : f_{day}^N	1.5	1.5	1.5	1.5
Dépendance des diatomées à la durée du jour : f_{day}^D	1.5	1.5	1.5	1.5
Paramètres statistiques	Simulations			
	detIni	stoIni	detUpd	stoUpd
Décorrélacion temporelle (en mois) : τ	0	1	0	1
Moyenne de la distribution gaussienne μ	0	0	0	0
Écart-type des perturbations des échelles non résolues : σ_{us}	0	0.2	0	0.2
Écart-type des perturbations sur les paramètres : σ_{up}	0	0.3	0	0.3
Filtre laplacien de corrélation spatiale (en points de grille) :	0	5	0	5

TABLE 5.1 – Valeurs des paramètres perturbés, utilisées dans les différentes simulations. Tous les paramètres présentés dans ce tableau sont perturbés avec les paramétrisations stochastiques présentées en section 5.1.2. Les équations dans lesquelles les paramètres déterministes interviennent sont décrites au chapitre 4, section 3.2.2.2. Les abréviations nano, phyto, zoo font respectivement références au nanophytoplancton, au phytoplancton et au zooplancton

Pour analyser l'impact du réajustement des paramètres, on réalise une simulation déterministe d'un an sans perturbations stochastiques (simulation *detUpd*). Les figures 5.6 et 5.7 montrent que cette simulation affiche des concentrations en chlorophylle bien plus faibles sur tout le domaine (principalement pendant le bloom de printemps). Si la production de chlorophylle est améliorée aux abords du Gulf Stream (voir section 5.3.1), elle est largement sous-estimée partout ailleurs. Le réajustement des paramètres déterministes induit donc une forte diminution en chlorophylle. L'effet moyen des paramétrisations stochastiques stimulant la production en chlorophylle, cette simulation constitue un bon point de départ pour pouvoir réaliser une simulation d'ensemble en accord avec les données SeaWiFS.

Pour le vérifier, on utilise la même démarche que précédemment. Une simulation de 3 membres (simulation *stoUpd*), incluant les perturbations stochastiques et les paramètres réajustés (cf. table 5.1), est réalisée sur l'année 2005. Comme pour la simulation *stoIni*, l'initialisation est faite à partir de l'état biogéochimique de la simulation *detIni*. Les perturbations aléatoires introduites dans le modèle sont identiques à celles de la simulation *stoIni*.

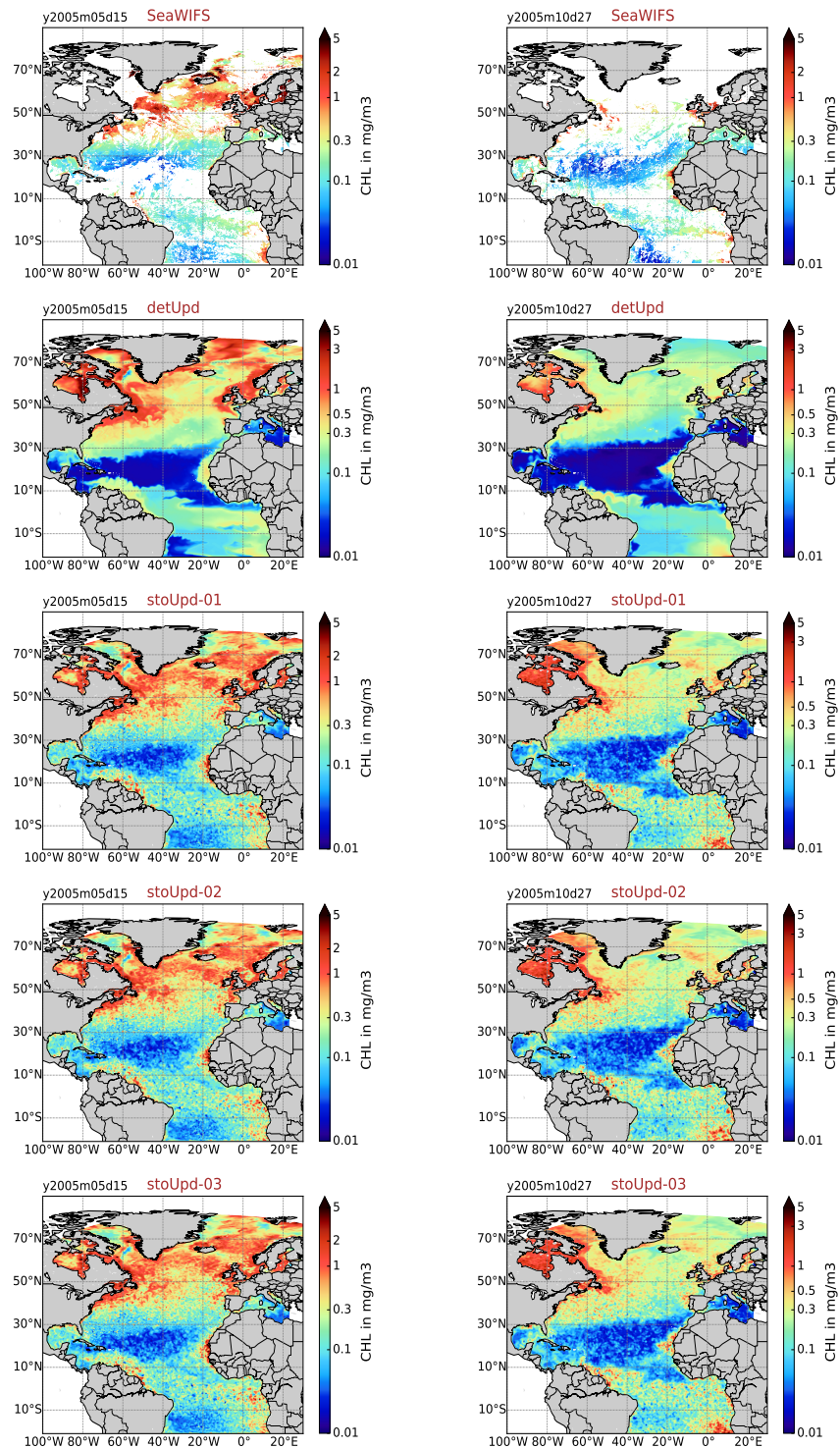


FIGURE 5.7 – Cartes de chlorophylle de surface des observations SeaWIFS (1^{ère} ligne), de la simulation déterministe *detUpd* (2^{ème} ligne) et des 3 membres de la simulation d'ensemble *stoUpd* (lignes 3 à 5). Les résultats sont présentés en moyenne de 5 jours pour le 15 mai (1^{ère} colonne) et le 27 octobre (2^{ème} colonne) de l'année 2005.

La figure 5.7 compare les concentrations en chlorophylle de la simulation *stoUpd* avec les données SeaWIFS pour le 15 mai et le 27 octobre 2005. Pour observer l'impact de l'introduction des perturbations aléatoires, les cartes de chlorophylle de surface de la simulation *detUpd* sont également présentées. Cette figure montre que les fortes concentrations en chlorophylle de surface, qui étaient simulées par la configuration de *stoIni* pendant le bloom de printemps, ont pu être significativement diminuées sur l'ensemble du domaine. Contrairement à la simulation *stoIni*, chaque réalisation de l'ensemble *stoUpd* est qualitativement comparable aux données SeaWIFS en ce qui concerne l'intensité et la répartition spatiale de la chlorophylle. Chaque membre affiche des distributions en chlorophylle plus « patchy » que les structures lissées de la simulation déterministe *detIni*. Ces effets de petite échelle seront directement associés avec la dispersion d'ensemble (cf. section suivante) et seront donc nécessaires pour comparer la simulation *stoUpd* avec les observations de la couleur de l'eau.

5.2.3 Conclusion partielle

Dans cette section, on a montré qu'en raison de la non-linéarité des équations, l'ajout de perturbations a tendance à augmenter les concentrations en chlorophylle de surface. À cause de cet effet moyen, les simulations (3 membres) surestiment la chlorophylle de surface par rapport aux données SeaWIFS. Pour compenser cet effet, nous avons proposé une version ajustée du jeu de paramètres déterministes. Ce réajustement a permis de réaliser 3 simulations dans lesquelles la représentation de la chlorophylle de surface était en accord avec les observations de la couleur de l'eau.

Pour conclure, il est nécessaire de souligner que le jeu de paramètres réajustés convient uniquement à notre approche probabiliste, c'est-à-dire à la simulation de ces sources d'incertitude. Les perturbations stochastiques étant introduites directement dans les équations biogéochimiques, elles ne peuvent pas être séparées du modèle déterministe. De la même manière, les paramètres biogéochimiques (déterministes) du modèle ne peuvent pas être découplés du niveau d'incertitude global du système. Nous avons, en effet, pu constater que les paramètres de la simulation (*detUpd*) ne sont pas adaptés à une simulation déterministe (mis à part au niveau du Gulf Stream). Dans la suite du document l'expression « paramétrisations stochastiques » fera ainsi référence à la combinaison perturbations stochastiques plus jeu de paramètres réajustés.

5.3 Impacts des paramétrisations stochastiques

Dans la section précédente, nous avons présenté une version probabiliste de PISCES capable de produire une simulation d'ensemble pouvant être comparée aux images SeaWIFS. Dans cette section, la simulation *stoUpd* est étendue à 60 membres. Les références à la simulation *stoUpd* désigneront donc maintenant une simulation d'ensemble de 60 membres réalisée sur l'année 2005. Dans cette section, aucune comparaison ne sera faite avec les observations satellites. Ce sera l'objet du chapitre suivant dans lequel les observations SeaWIFS et la simulation d'ensemble seront comparées de manière statistique. Dans cette section, l'objectif est de valider la pertinence des paramétrisations stochastiques à partir d'une comparaison avec la simulation déterministe.

Cela permettra de montrer que :

- la représentation moyenne de la chlorophylle faite par la simulation d'ensemble reste comparable à l'état déterministe.
- les paramétrisations stochastiques sont capables de générer de la dispersion en surface et dans l'ensemble de la couche euphotique.

La première partie analyse l'impact des paramétrisations stochastiques sur la chlorophylle de surface. La seconde partie se penche sur leurs impacts sur les distributions verticales de la chlorophylle.

5.3.1 Répartition de la chlorophylle de surface

Effets moyens

Dans un premier temps, l'impact des paramétrisations stochastiques est évalué en figure 5.8 à partir de cartes de la moyenne d'ensemble de la chlorophylle de surface pour le 31 mars, le 15 mai et le 27 octobre 2005. Pour faciliter la comparaison avec l'état déterministe, les différences entre la moyenne d'ensemble et la simulation déterministe *detIni* sont également présentées.

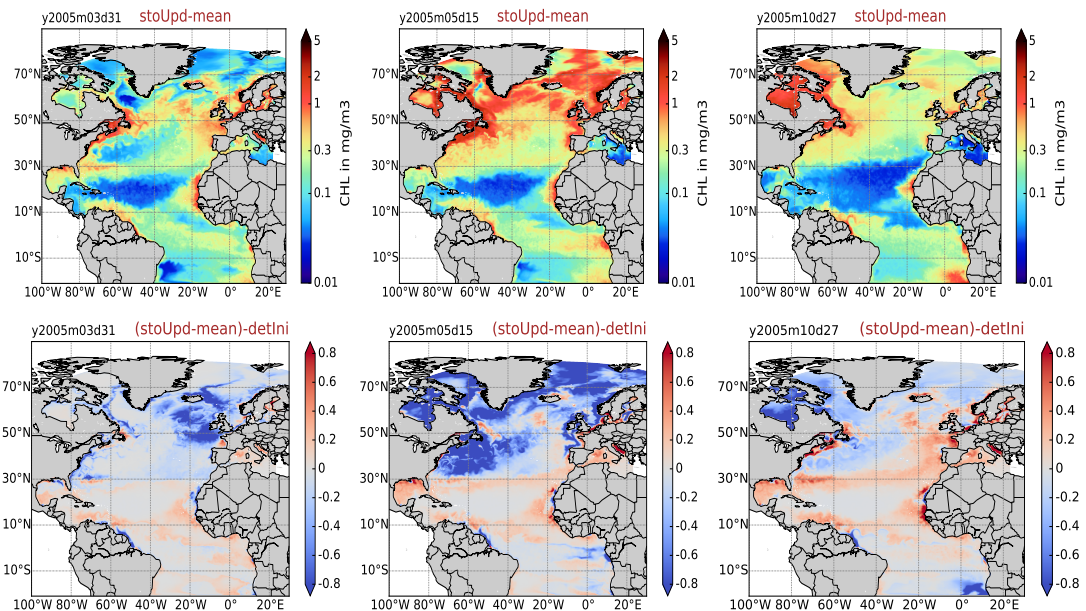


FIGURE 5.8 – Cartes de chlorophylle de surface de la moyenne d'ensemble de la simulation *stoUpd* (1^{ère} ligne) et de la différence entre la moyenne d'ensemble et la simulation *detIni* (2^{ème} ligne). Ces résultats sont présentés pour le 31 mars (1^{ère} colonne), le 15 mai (2^{ème} colonne) et le 27 octobre (3^{ème} colonne) de l'année 2005.

La figure 5.8 nous permet de constater que l'estimation de la moyenne d'ensemble de la chlorophylle de surface est proche de celle de la simulation déterministe *detIni*. Le réajustement des paramètres a donc permis de réduire l'augmentation de la chlorophylle causée par les perturbations stochastiques. L'état moyen de la simulation d'ensemble est ainsi recentré au niveau de la simulation déterministe. De surcroît, l'effet moyen des paramétrisations stochastiques induit des différences systématiques avec la simulation déterministe.

- Une diminution dans la partie nord du domaine (au-dessus de 50 °N) et dans la zone du Gulf Stream.
- Une augmentation dans la zone subéquatoriale, en particulier dans les zones frontales du gyre subtropical et dans les régions côtières.

Le plus souvent, la moyenne d'ensemble présente des concentrations en chlorophylle de surface plus faibles que la simulation déterministe. Cet effet est principalement le résultat du réajustement des paramètres du modèle. Cette tendance reste toutefois peu significative puisqu'elle se produit surtout dans la partie nord du domaine, zone dans laquelle les concentrations en chlorophylle sont généralement élevées.

Dans la section précédente, nous avons montré que le seul réajustement des paramètres (simulation *detUpd*) permettait de réduire la chlorophylle de surface dans la région du Gulf Stream. On constate ici que la simulation *stoUpd* conserve cette amélioration malgré l'introduction des perturbations stochastiques. À noter que cette surproduction avait été reconnue comme une déficience importante des simulations du modèle PISCES faites à partir du système de prévision de MERCATOR-Océan. Je précise que cet effet n'est pas uniquement un effet moyen puisqu'il a pu être observé dans chacun des membres de la simulation d'ensemble présentés en figure 5.7.

Une autre conséquence remarquable due aux paramétrisations stochastiques est l'augmentation de la chlorophylle de surface au niveau du gyre subtropical. Elle permet de réduire l'étendue spatiale de la zone oligotrophe identifiée comme trop développée dans la simulation déterministe *detini* (chapitre 4 section 4.2.1). Enfin, les effets combinés des diminutions septentrionales et des augmentations tropicales de la chlorophylle de surface permettent d'atténuer les forts gradients nord-sud initialement observés dans la simulation déterministe.

Écart-types

Afin d'évaluer le niveau de dispersion induit par les paramétrisations stochastiques, on représente en figure 5.9 les cartes des écart-types et des écart-types relatifs calculés en chaque point de grille comme la racine carrée de l'écart quadratique moyen entre les membres de l'ensemble et la moyenne d'ensemble. L'écart-type relatif est obtenu en normalisant l'écart-type par la moyenne d'ensemble.

L'analyse de ces cartes démontre que, malgré l'allure similaire de chacun des membres de la simulation d'ensemble (cf. figure 5.7), une importante dispersion est produite toute l'année dans la majeure partie du domaine. La dispersion résulte ainsi principalement de

l'effet cumulé de blooms localisés (cf. figure 5.7), très différents dans chacun des membres de l'ensemble. Je précise que si les effets moyens résultent de la non-linéarité des équations (perturbations des échelles biogéochimiques non résolues), ces événements localisés sont principalement dus à la simulation des incertitudes des paramètres du modèle. À noter que des expériences réalisées en découplant les 2 paramétrisations stochastiques ont été nécessaires pour s'assurer de ce résultat.

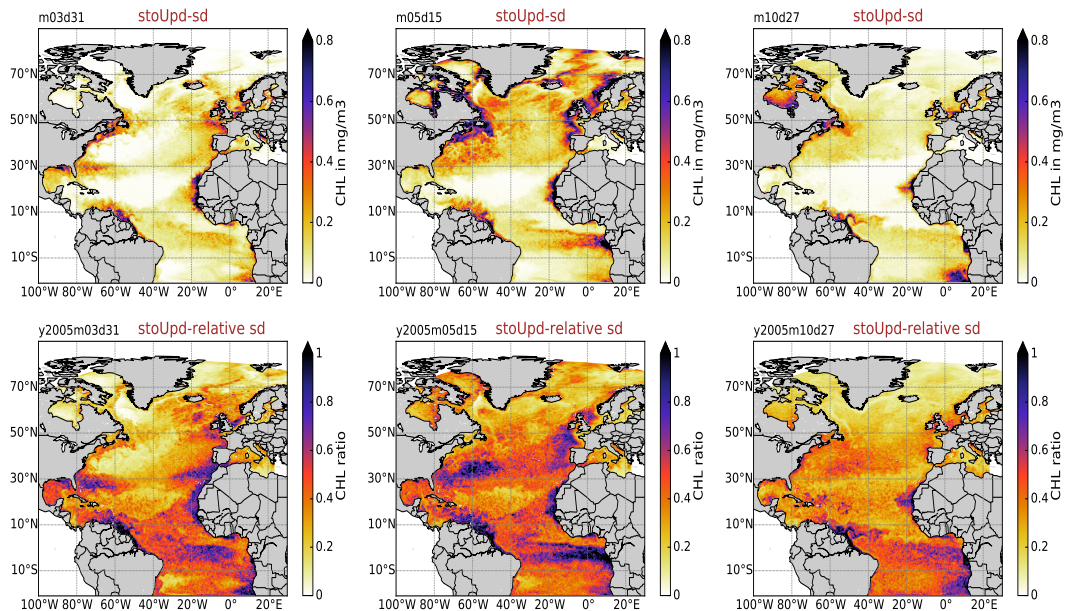


FIGURE 5.9 – Cartes de surface de l'écart-type (1^{ère} ligne) et de l'écart-type relatif de la simulation d'ensemble *stoUpd* (2^{ème} ligne). Ces résultats sont présentés sur la chlorophylle pour le 31 mars (1^{ère} colonne), le 15 mai (2^{ème} colonne) et le 27 Octobre (3^{ème} colonne) de l'année 2005.

La figure 5.9 indique que les zones où la dispersion relative est maximale sont localisées dans la moitié sud du domaine. Si on compare ces cartes d'écart-types à la variabilité interannuelle de la simulation *detIni* (chapitre 4 section 4.3.1), on observe des structures de variabilité très différentes. Ce constat permet de montrer que les incertitudes simulées par les paramétrisations stochastiques sont distinctes de celles dues aux paramètres extérieurs présentés dans le chapitre précédent. De plus, on montre ici que les paramétrisations stochastiques permettent de générer un niveau de dispersion plus élevé et mieux réparti spatialement.

5.3.2 Structures verticales de la chlorophylle

Étant donné le peu d'observations disponibles sur la verticale, évaluer la cohérence des distributions verticales de la simulation d'ensemble est plus délicat. Afin d'éviter de se restreindre à une comparaison, soit trop globale, soit trop localisée, nous considérerons que l'information de la simulation déterministe permet de vérifier la cohérence moyenne des structures verticales de la simulation d'ensemble. Les distributions verticales de grande échelle de la simulation *detIni* ayant en effet été validées au chapitre précédent, nous les utiliserons donc pour analyser l'impact moyen des paramétrisations stochastiques.

Effets moyens

Afin d'examiner l'impact moyen des paramétrisations stochastiques sur la verticale, la figure 5.10 présente des coupes longitudinales des structures moyennes de la simulation *stoUpd* et de la simulation déterministe *detIni* pour le 31 mars et le 15 mai 2005.

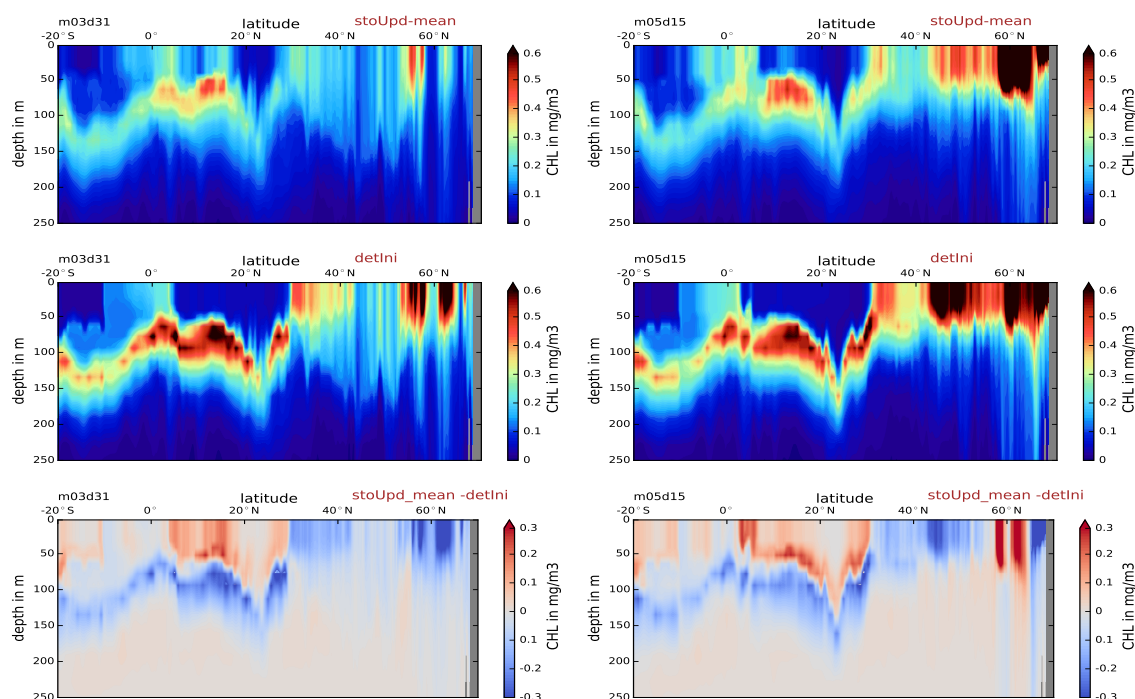


FIGURE 5.10 – Coupes verticales à 30°W de la chlorophylle pour le 31 mars (1^{ère} colonne), le 15 mai (2^{ème} colonne). La 1^{ère} ligne présente les structures moyennes de la simulation d'ensemble et la 2^{ème} ligne celles de la simulation déterministe *detIni*. La différence entre la moyenne d'ensemble et la simulation déterministe *detIni* est également présentée pour les 2 dates (3^{ème} ligne).

De manière générale, cette figure montre que les écarts entre la simulation *detIni* et la moyenne d'ensemble sont relativement faibles. Ils ne dépassent 0.3 mg de CHL/m³ qu'en de rares régions et uniquement en période de forte activité biologique. L'effet moyen des paramétrisations stochastiques a donc peu d'impact sur la structure verticale. En particulier, les profondeurs de propagation et la localisation des maxima de subsurface de la structure moyenne sont pratiquement inchangées par rapport à ceux de la simulation déterministe. Les maxima de subsurface restent prédominants dans les basses latitudes (entre 20°S et 40°N) et absents dans les hautes latitudes. Toutefois, leurs intensités et leurs étendues verticales semblent réduites par l'effet moyen des paramétrisations stochastiques.

Au mois de mai, l'augmentation moyenne de la chlorophylle de surface peut tout de même, dans la région équatoriale, faire disparaître certains maxima de subsurface observables dans la simulation déterministe. Compte tenu de la dispersion verticale (cf. figure 5.11), ce constat ne signifie pas que la simulation d'ensemble ne représente jamais l'existence de maxima de subsurface. Cependant, leur présence n'apparaît plus dans la structure verticale moyenne.

En définitive, les relations (moyennes) entre la chlorophylle de surface et l'allure des structures verticales, identifiables dans la simulation déterministe, semblent être conservées malgré les paramétrisations stochastiques.

Écarts-types

Pour évaluer le niveau de dispersion produit par les paramétrisations stochastiques, nous présentons en figure 5.11 des coupes verticales à 30°W de l'écart-type et de l'écart-type relatif de la simulation d'ensemble.

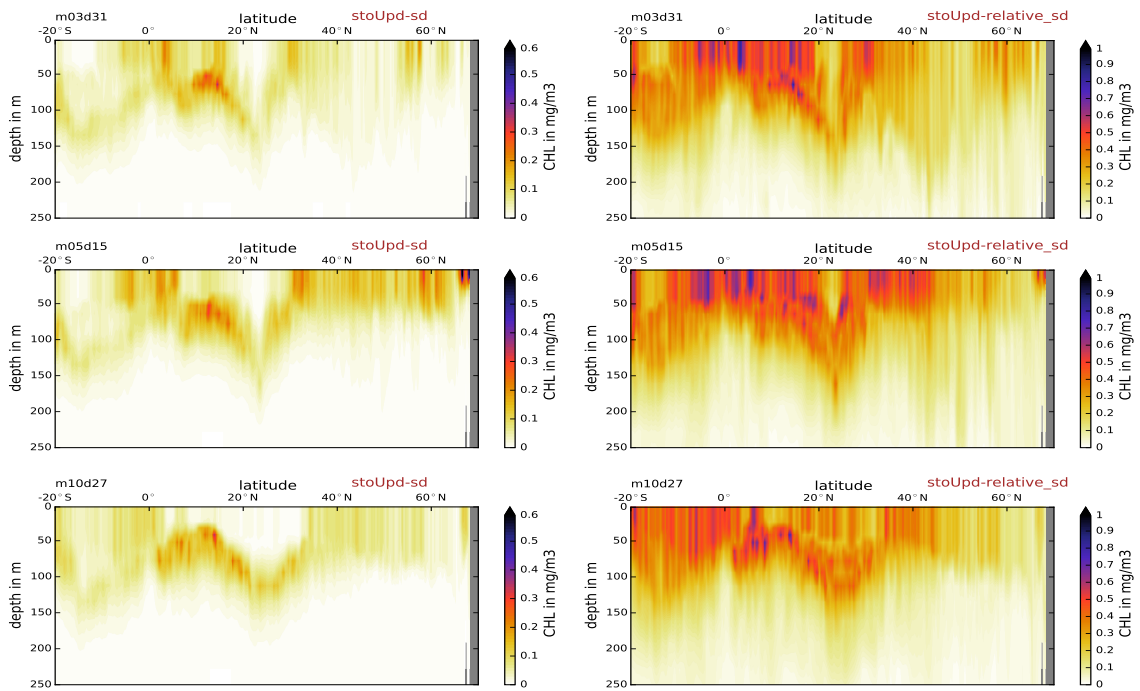


FIGURE 5.11 – Coupes verticales à 30°W de l'écart-type (1^{ère} colonne) et de l'écart-type relatif (2^{ème} colonne) de la simulation d'ensemble *stoUpd*. Les résultats sont présentés sur la chlorophylle pour le 31 mars (1^{ère} ligne), le 15 mai 2005 (2^{ème} ligne) et le 15 octobre 2005 (3^{ème} ligne).

Cette figure souligne qu'une dispersion significative est produite dans toute la zone euphotique. Les zones de plus faible dispersion semblent majoritairement localisées dans la moitié nord du domaine. Ce résultat est cohérent avec ce que nous avons observé pour la chlorophylle de surface (cf. figure 5.9). À noter qu'il est également satisfaisant de constater une dispersion nulle en dessous de la couche euphotique.

Comme pour la chlorophylle de surface, ces écarts-types peuvent être mis en parallèle avec la variabilité interannuelle de la simulation déterministe *detIni*, diagnostiquée au chapitre 4 section 4.3.1. Si l'ordre de grandeur de ces 2 variabilités est plus ou moins comparable, la simulation d'ensemble présente une dispersion mieux répartie dans les 100 premiers mètres de la colonne d'eau. Dans la simulation *detIni*, la variabilité interannuelle se réduisait en effet majoritairement aux profondeurs des maxima de subsurface.

Identifier une dispersion importante et relativement homogène sur la verticale peut également laisser présupposer que les membres de l'ensemble puissent présenter des profondeurs variables de maxima de subsurface. Ce point sera important pour pouvoir expliquer les écarts entre le modèle et des observations sur la verticale (cf. chapitre suivant).

Bilan

- L'effet moyen des paramétrisations stochastiques ne détériore pas l'information grande échelle de la simulation déterministe.
- Les paramétrisations stochastiques produisent une dispersion significative, bien répartie spatio-temporellement.

5.4 Conclusion

Dans ce chapitre, nous avons présenté une méthode permettant d'introduire des processus aléatoires dans le modèle PISCES. À partir de paramétrisations stochastiques spécifiques, nous avons ensuite pu démontrer que ces processus aléatoires pouvaient être utilisés pour simuler 2 sources d'incertitude : les incertitudes des paramètres biogéochimiques du modèle et les incertitudes dues aux effets des échelles biogéochimiques non résolues.

En première approche, ces paramétrisations stochastiques ont été intégrées dans la configuration biogéochimique de la simulation *detIni* dont les résultats avaient été présentés au chapitre précédent. Une première simulation d'ensemble de 3 membres a alors été réalisée. À cause de la non-linéarité des équations, nous avons constaté que l'introduction des perturbations stochastiques produisait en permanence un fort effet moyen. En conséquence, les membres de cette simulation surestimaient fortement les concentrations en chlorophylle de surface par rapport aux observations SeaWIFS. Pour contrecarrer cet effet moyen, nous avons proposé de réajuster la valeur de certains paramètres biogéochimiques du modèle PISCES. Avec ce nouveau jeu de paramètres, une nouvelle simulation d'ensemble de 3 membres, possédant des caractéristiques nettement plus proches des observations SeaWIFS, a ensuite pu être réalisée. Ces résultats ont justifié d'élargir cette simulation à 60 membres.

La structure moyenne de cette nouvelle simulation d'ensemble a été comparée avec la simulation déterministe *detIni*. Cela nous a permis de vérifier la pertinence des paramétrisations stochastiques (incluant l'ajustement des paramètres), à partir de la structure moyenne des distributions de la chlorophylle. En particulier, nous avons vérifié que la simulation d'ensemble conserve les caractéristiques de la simulation déterministe en termes de profondeur de propagation et de localisation des maxima de subsurface. Nous avons également montré que les sources d'incertitude simulées permettaient de produire une dispersion d'ensemble significative.

En définitive, nous avons montré dans ce chapitre que les paramétrisations stochastiques ont un impact fort sur l'estimation des champs de chlorophylle. Dans le contexte de l'assimilation de données, cela met donc en avant l'importance de considérer l'incertitude du modèle pour évaluer les erreurs de prévision.

On rappelle les principaux résultats de ce chapitre.

- Par l'intermédiaire de paramétrisations stochastiques spécifiques, l'utilisation de processus aléatoires permet de simuler certaines sources d'incertitudes d'un modèle biogéochimique.
- Le passage déterministe \rightarrow probabiliste modifie le niveau d'incertitude global du modèle.
- L'introduction de processus stochastiques dans un modèle implique de redéfinir la configuration initialement calibrée pour l'approche déterministe.
- L'effet des paramétrisations stochastiques est une superposition d'effets moyens et d'effets de petite échelle.
- Les structures dynamiques de la simulation déterministe sont conservées par la structure moyenne de la simulation d'ensemble à 60 membres *stoUpd*.
- Les membres de la simulation d'ensemble *stoUpd* définissent un niveau élevé d'incertitude (dispersion) sur la chlorophylle.
- L'incertitude du modèle impacte fortement l'estimation de la chlorophylle.

Dans le chapitre suivant, nous vérifierons que la dispersion d'ensemble possède des caractéristiques statistiques cohérentes avec les observations SeaWiFS.

Chapitre 6

Réponse probabiliste aux incertitudes biogéochimiques

Sommaire

6.1	Dispersion statistique de la chlorophylle de surface	112
6.1.1	Quantiles de surface	112
6.1.2	Évolution temporelle	116
6.1.3	Conclusion partielle	118
6.2	Fiabilité statistique de l'ensemble	119
6.2.1	Démarche	119
6.2.2	Histogrammes de rangs	120
6.2.2.1	Caractéristiques spatiales	120
6.2.2.2	Caractéristiques temporelles	123
6.3	Distributions verticales de la chlorophylle	126
6.3.1	Dispersion d'ensemble	126
6.3.2	Comparaison aux observations	129
6.3.2.1	Climatologie verticale	129
6.3.2.2	Diversité des comportements biologiques	131
6.4	Conclusion	135

Dans le chapitre précédent, la pertinence de la simulation d'ensemble de 60 membres établie à partir de paramétrisations stochastiques a pu être démontrée. On s'est pour cela basé uniquement sur les diagnostics usuels de moyenne et d'écart-type. Tant que les distributions de probabilité restent gaussiennes, cette description synthétique reste un moyen satisfaisant de caractériser l'ensemble. Lorsque l'allure des distributions devient asymétrique, définir l'ensemble à partir de son comportement moyen ne donne plus qu'une information incomplète. Il devient dès lors nécessaire de prendre en compte l'allure des distributions de probabilité. La démarche adoptée dans ce chapitre réside justement sur ce point particulier. On s'appliquera ainsi à caractériser la simulation d'ensemble *stoUpd* à partir de l'information statistique donnée par les différents membres de l'ensemble.

Les prévisions d'ensemble se généralisant, les outils statistiques pour les évaluer ne cessent de se développer (le livre de Wilks, 2006 en fait une revue très complète). Dans le cadre de cette thèse, l'idée n'est pas d'accumuler les scores statistiques mais de définir rigoureusement les caractéristiques de la simulation d'ensemble en vue de l'assimilation de données. L'un des objectifs de ce chapitre sera donc de comparer objectivement notre simulation d'ensemble avec les observations de la couleur de l'eau SeaWIFS à partir d'histogrammes des rangs.

La première partie de ce chapitre présente une analyse de la dispersion d'ensemble de la chlorophylle de surface telle que décrite par les quartiles de la distribution des membres de l'ensemble. La seconde section sera consacrée à une analyse des rangs de l'observation dans l'ensemble à l'aide des outils développés par Candille and Talagrand (2005). Elle permettra de vérifier le niveau de cohérence statistique entre les observations SeaWIFS et la distribution de probabilité définie par l'ensemble (diagnostic de fiabilité). Dans la dernière partie, on se focalisera sur les caractéristiques verticales de la dispersion avec une approche similaire à celle utilisée pour la chlorophylle de surface (première section de ce chapitre). Nous évaluerons également la capacité des paramétrisations stochastiques à simuler des profils verticaux représentatifs de l'existence conjuguée de plusieurs espèces de phytoplancton au sein des 2 compartiments modélisés par le modèle PISCES.

6.1 Dispersion statistique de la chlorophylle de surface

Dans cette première section, la dispersion d'ensemble est analysée en tenant compte des distributions de probabilité définies par les membres de l'ensemble. Cela permettra de comparer le niveau d'incertitude simulé par les paramétrisations stochastiques avec les observations SeaWIFS. En d'autres termes, l'objectif de cette section est de vérifier si l'enveloppe de l'ensemble permet d'inclure les observations de la couleur de l'eau.

Dans la première partie, nous étudierons la dispersion statistique à partir des quartiles de la distribution des membres de l'ensemble. La seconde partie portera sur l'évolution temporelle des différents membres de la simulation *stoUpd*.

6.1.1 Quantiles de surface

Afin d'analyser la dispersion d'ensemble en tenant compte des distributions de probabilité des membres de l'ensemble, on présente sur les figures 6.1 et 6.2 la réponse statistique de la chlorophylle de surface aux paramétrisations stochastiques pour le 15 mai et le 27

octobre 2005. La répartition spatiale de la chlorophylle est examinée à partir de cartes de quantiles, produits composites construits à partir des valeurs minimales, maximales, médianes et des quartiles des membres de l'ensemble en chaque point de grille du modèle. Je précise que ces cartes affichent un état statistique et n'attestent donc pas d'un état du modèle. Afin de faciliter la comparaison, les quantiles sont interpolés sur la grille au $1/12^\circ$ des observations SeaWiFS, également représentées sur ces figures.

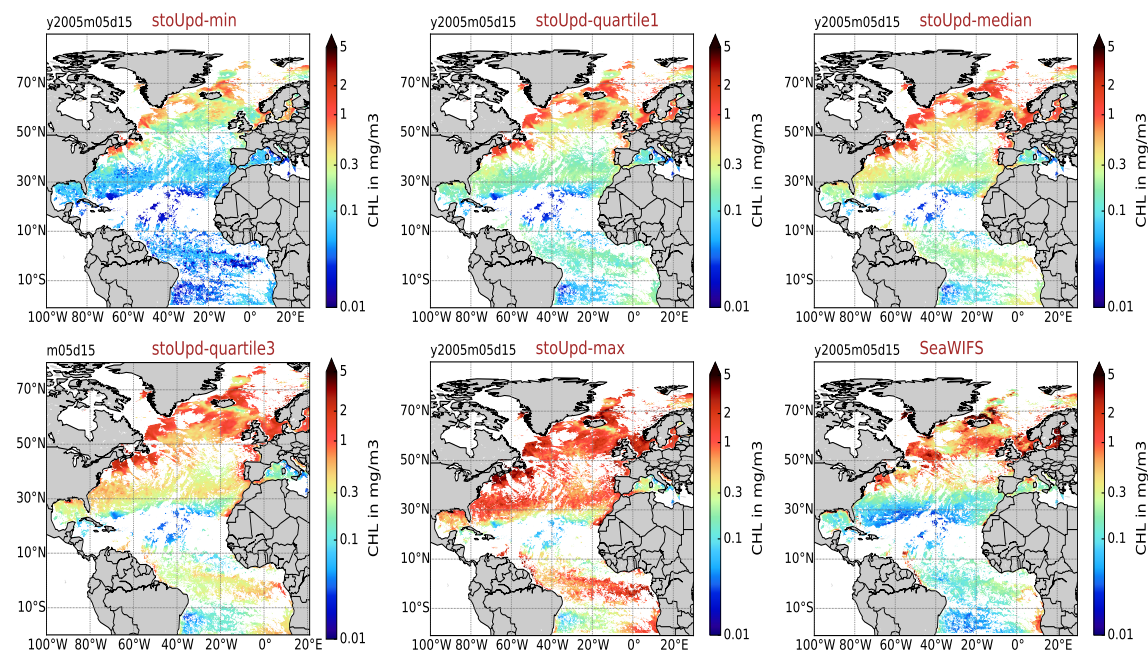


FIGURE 6.1 – Réponse statistique de la chlorophylle de surface aux paramétrisations stochastiques telle que décrite par la simulation d'ensemble à 60 membres pour le 15 mai 2005. De gauche à droite, la 1^{ère} ligne présente les cartes des minima, des 1^{ers} quartiles et des médianes de l'ensemble. Les images de la 2^{ème} ligne représentent les 3^{èmes} quartiles et les maxima de l'ensemble. Les observations satellites SeaWiFS sont également représentées.

Le point crucial mis en évidence par les figures 6.1 et 6.2 est que l'ensemble présente une forte étendue (écart entre les valeurs maximales et minimales) bien répartie sur l'ensemble du domaine. Cette large enveloppe statistique permet à l'ensemble de capturer la majorité des observations SeaWiFS. Plus précisément, sans tenir compte d'erreur d'observation, environ 70% des données d'observation sont incluses dans l'ensemble pour les 2 dates considérées. Il faut préciser que ces pourcentages sont significativement plus élevés si on ne comptabilise pas les données de régions comme la mer Méditerranée ou le pourtour côtier de l'Europe Septentrionale dans lesquelles l'ensemble est clairement sous-dispersif. Ces zones sont en effet le siège de comportements biogéochimiques distincts, bien différents de ceux que l'on rencontre dans l'océan ouvert. Les paramétrisations stochastiques ayant été calibrées empiriquement, on pouvait s'attendre à ce qu'elles n'expliquent pas complètement les écarts aux observations dans un domaine d'étude aussi vaste. En définitive, même en hiver, lorsque la dispersion est plus faible, l'ensemble est suffisamment dispersif pour représenter l'essentiel de la diversité des régimes biogéochimiques observables dans les données de la couleur de l'eau.

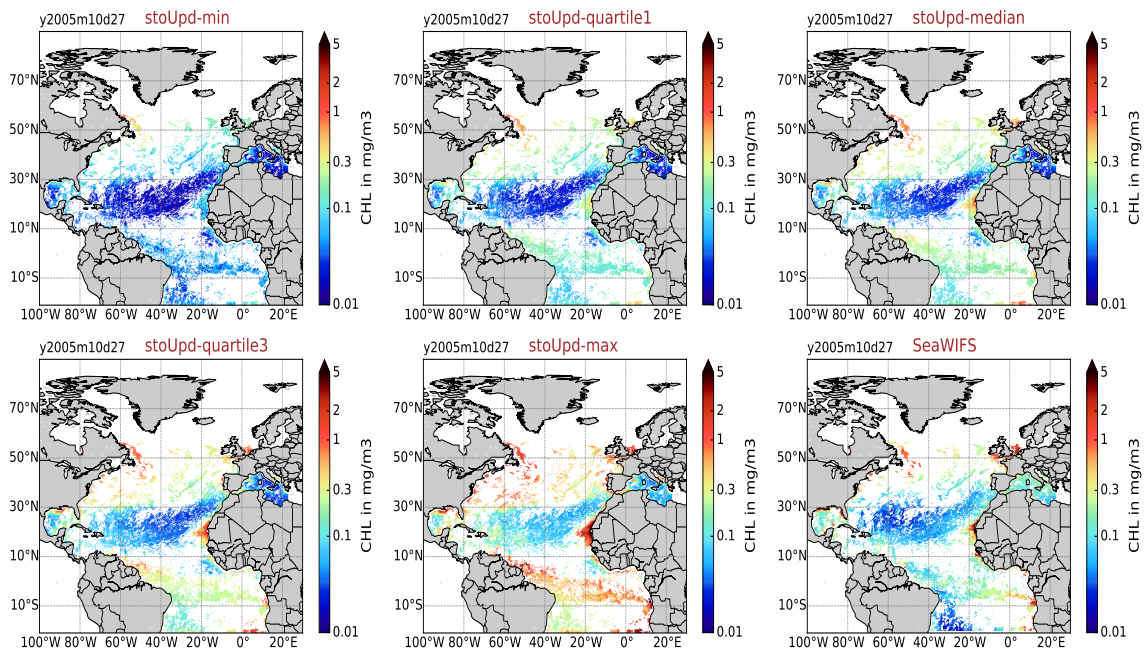


FIGURE 6.2 – Légende identique à la figure 6.1 pour le 27 octobre 2005.

Les figures 6.1 et 6.2 montrent que l'ensemble se caractérise par un écart entre le 3^{ème} quartile et le maximum systématiquement plus large qu'entre le minimum et le 1^{er} quartile. Les distributions de probabilité ont donc tendance à présenter une asymétrie positive, c'est-à-dire que la densité de probabilité des faibles concentrations en chlorophylle (inférieures à la moyenne) est plus élevée. L'introduction de perturbations stochastiques ne modifie donc pas la tendance log-normale de l'allure des distributions de chlorophylle, depuis longtemps mise en évidence (p.ex. Vollenweider et al., 1992; Campbell, 1995). Si ces cartes paraissent insuffisantes pour en être pleinement convaincu, le lecteur peut consulter la figure 7.2 du chapitre suivant qui présente des distributions des membres de l'ensemble pour la chlorophylle de surface sous la forme d'histogrammes.

Les cartes des maxima en chlorophylle permettent de montrer que le modèle stochastique est capable de simuler de très fortes concentrations en chlorophylle. Ces événements « extrêmes » reflètent la probabilité d'existence de blooms locaux de phytoplancton. Cette caractéristique est nécessaire pour que l'ensemble puisse représenter l'importante hétérogénéité spatiale présente dans les données de la couleur de l'eau.

Dans le but d'affiner l'analyse de la dispersion et la capacité de l'ensemble à inclure les observations, la figure 6.3 présente l'enveloppe de l'ensemble sur la chlorophylle de surface le long de coupes longitudinales à 20°W et 30°W. Les concentrations en chlorophylle de surface de la simulation déterministe *detIni* sont également affichées.

À cause d'une dynamique biogéochimique accrue, la dispersion d'ensemble apparaît logiquement plus forte pendant la période du bloom de printemps. De plus, la figure 6.3 montre bien que l'intensité de la dispersion est fortement dépendante de la zone géographique considérée.

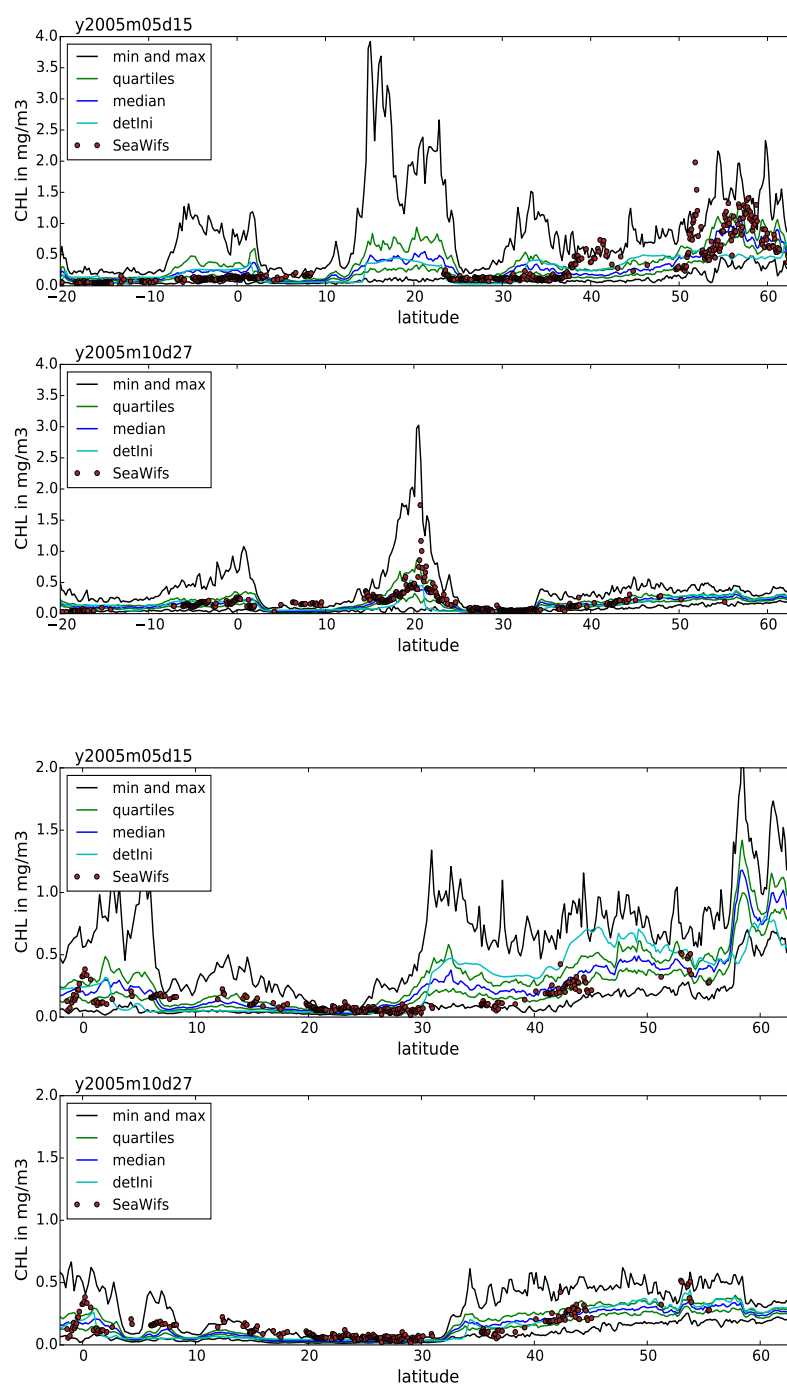


FIGURE 6.3 – Centiles d'ensemble de la chlorophylle de surface le long de coupes longitudinales à 20°W (2 figures du haut) et 30°W (2 figures du bas) pour le 15 mai et le 27 octobre 2005. Les maxima et minima sont représentés en noir, les 1^{ers} et 3^{èmes} quartiles en vert et la médiane en bleu foncé. Les points rouges indiquent les observations SeaWIFS. La simulation déterministe *detIni* est représentée en bleu cyan. À noter que l'incertitude des observations n'est pas considérée ici. Elle sera prise en compte dans la section 6.2.

La dispersion est ainsi maximale dans les régions du Gulf Stream, du courant équatorial et des upwellings côtiers, dans lesquelles l'importance des effets de petite échelle avait été précédemment identifiée (cf. chapitre 5 section 5.2). Une forte dispersion peut être observée aux alentours de l'upwelling de Mauritanie (10-20°N, coupe à 20°W), reconnu pour être à l'origine de fortes variabilités de production primaire (Estrada, 1980). Au niveau du gyre subtropical, la dispersion est relativement faible et les observations restent le plus souvent proches des minima de l'ensemble. Cela peut être perçu comme la conséquence de l'existence d'un seuil limite, les concentrations en chlorophylle ne pouvant jamais être négatives. Les perturbations stochastiques ont donc potentiellement plus "de facilité" à stimuler la production en chlorophylle. Les valeurs les plus extrêmes sont ainsi toujours de fortes concentrations. Dans la partie nord du domaine, l'ensemble est légèrement sous-dispersif pendant le bloom de printemps.

Récapitulatif

Ces figures confirment de manière satisfaisante la capacité de l'ensemble à inclure les observations de la couleur de l'eau SeaWIFS. Ainsi, les régions dans lesquelles l'ensemble est le plus dispersif sont généralement celles où les observations sont les plus hétérogènes. Inversement, lorsque la dispersion est faible, la plupart des observations affichent des concentrations en chlorophylle relativement basses et homogènes. Le modèle possède donc un niveau d'incertitude qui permet d'expliquer, dans la majorité des cas, les écarts entre le modèle et les observations. Par ailleurs, hormis quelques zones localisées très au nord, l'ensemble inclut toujours la simulation *detIni*. La simulation probabiliste conserve donc l'information du modèle déterministe.

Les figures présentées dans cette section ont également permis de souligner le caractère asymétrique des distributions de probabilité de la chlorophylle de surface. Ce dernier point sera particulièrement important dans le cadre de l'assimilation de données. Dans le chapitre 7, le caractère non gaussien des distributions de probabilité sera considéré en utilisant une méthode d'anamorphose.

6.1.2 Évolution temporelle

Dans cette partie, nous examinons l'évolution temporelle de la distribution de probabilité. À raison d'une sortie de modèle tous les 5 jours, la figure 6.4 représente, en plusieurs points de grille, les concentrations en chlorophylle de tous les membres de l'ensemble sur l'année 2005. Ces points de grille "stations" sont choisis pour être les plus représentatifs des différents comportements biogéochimiques. Leurs coordonnées géographiques sont affichées en figure 6.5. Lorsqu'elles sont disponibles, les observations SeaWIFS sont signalées sur la figure 6.4 par des cercles pleins rouges. Afin de rester cohérent avec la section précédente, les évolutions temporelles du minimum, du maximum, de la médiane et des quartiles de l'ensemble sont également représentées. On remarquera la présence des stations *BATS* et *INDIA*, dont le comportement biogéochimique est régulièrement étudié.

Le premier point que l'on peut identifier est la mise en place rapide de l'effet des perturbations stochastiques. Bien que la simulation d'ensemble soit initialisée durant une période de faible intensité biogéochimique, une dispersion significative s'établit dans la majeure

partie du domaine après seulement quelques pas de temps. Cette spécificité est d'intérêt majeur pour l'assimilation de données car il est nécessaire que les paramétrisations stochastiques puissent produire de la dispersion entre 2 étapes d'analyses. Aux stations *STAT_G* et *INDIA*, situées dans la partie nord du domaine, l'évolution temporelle de la chlorophylle montre qu'il faut attendre l'initiation du bloom de printemps avant qu'une dispersion significative s'établisse. La dispersion redevient faible à partir du mois de décembre. Dans ces cas, ce résultat souligne donc d'avantage la difficulté des paramétrisations stochastiques à représenter l'incertitude du modèle qu'une latence dans l'établissement de la dispersion.

Excepté en hiver et dans quelques zones particulières de la partie nord du domaine, les séries temporelles confirment le caractère fortement dispersif de l'ensemble dans la majeure partie du domaine pour toute la période de simulation. L'intensité de la dispersion respecte les cycles saisonniers avec une dispersion qui s'intensifie à partir de la période du bloom de printemps. Elle atteint son niveau maximal entre juin et juillet avant de stagner puis de décroître avec l'arrivée de l'hiver. Autour de l'équateur (stations *STAT_B* et *STAT_D*), la régularité du niveau de dispersion est en accord avec l'absence d'un rythme saisonnier prononcé.

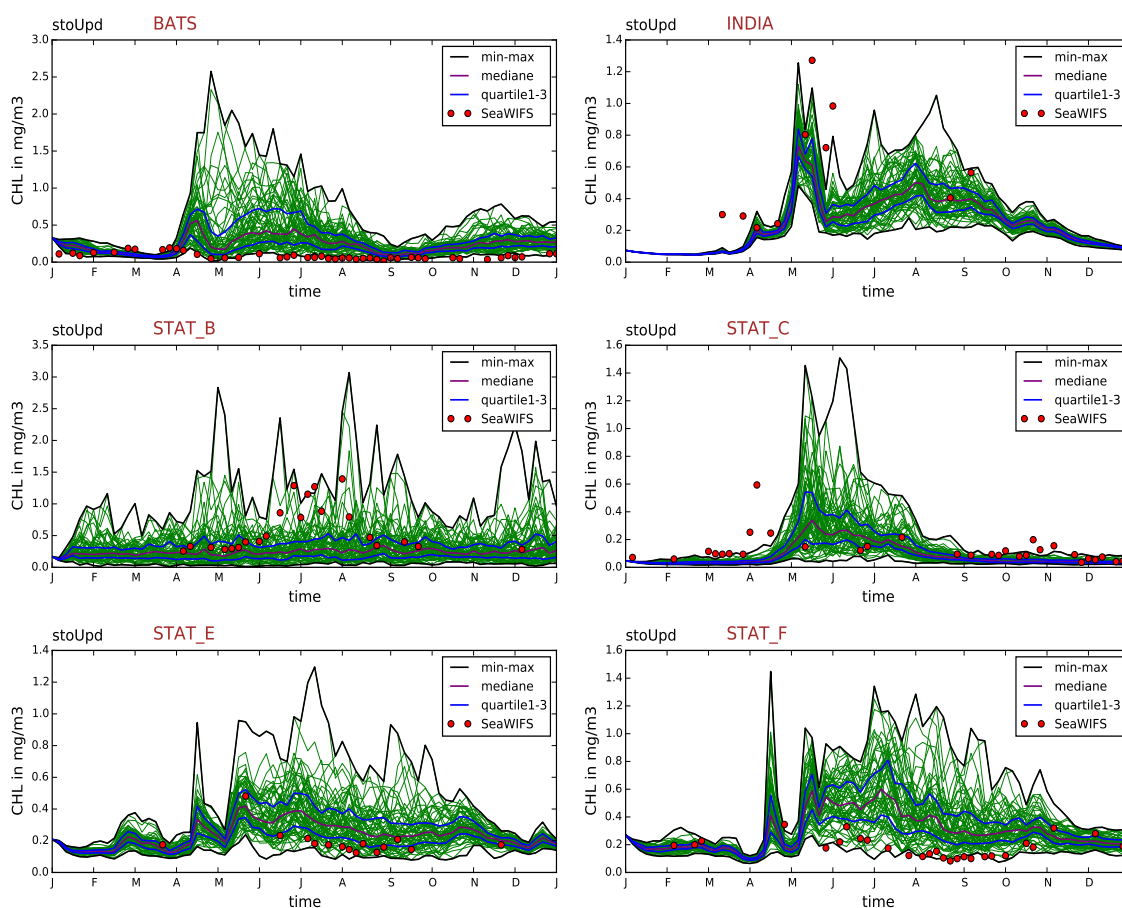


FIGURE 6.4 – Séries temporelles de la chlorophylle de surface simulée par les membres de l'ensemble en différents points de grille du domaine. Les observations SeaWIFS sont indiquées par des cercles pleins rouges. L'enveloppe statistique de l'ensemble est caractérisée par les quartiles de l'ensemble.

Les séries temporelles montrent également que les paramétrisations stochastiques ont un impact sur l'intensité du bloom du printemps (particulièrement à la station *BATS*) mais qu'elles ne modifient pas sa date de déclenchement. Cela pourrait plaider en faveur d'un déclenchement du bloom en grande partie dépendant des propriétés physiques de l'océan (eg. ventilation de la couche de mélange). L'évolution temporelle des quartiles d'ensemble confirme l'allure générale de la distribution de probabilité pour toute la période de simulation : une large enveloppe présentant une asymétrie positive avec un écart maximal entre le 3^{ème} quartile et le maximum de l'ensemble.

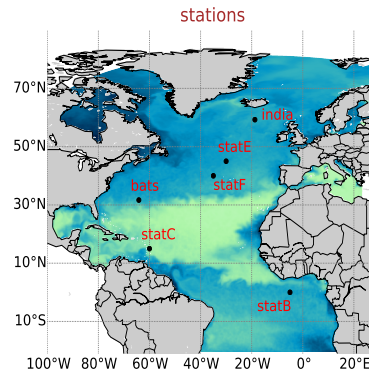


FIGURE 6.5 – Localisation des points de grille stations utilisés dans cette étude

6.1.3 Conclusion partielle

En résumé, l'étendue de la dispersion d'ensemble permet d'inclure la majorité des observations de la couleur de l'eau SeaWIFS. Nous avons toutefois montré que leur répartition dans l'enveloppe de l'ensemble pouvait significativement varier selon la région considérée.

Au niveau du Gulf Stream (stations *BATS* et *StatF*), les observations restent constamment confinées en proximité des minima de l'ensemble. Malgré le réajustement des paramètres du modèle (cf. chapitre précédent), la tendance à surestimer la chlorophylle subsiste donc dans la simulation d'ensemble. À la station *BATS*, cette caractéristique apparaît encore comme le résultat d'un fort bloom de printemps, quasiment inexistant dans les observations. En dépit d'une amélioration conséquente (par rapport à la simulation *de-Itini*) de la représentation de la chlorophylle dans la région du Gulf Stream, il convient de rappeler qu'une partie seulement des incertitudes ont été prises en compte. Il est de plus légitime d'envisager que les imperfections sur la représentation physique jouent un rôle non négligeable, en particulier au niveau du Gulf Stream.

Dans les régions Équatoriales (station *STAT_B*), les observations semblent relativement bien distribuées dans l'ensemble malgré une plus faible dispersion. Dans la partie nord du domaine, un nombre significatif d'observations n'est pas capturé par l'ensemble.

Pour terminer, il convient d'ajouter que les données de la couleur de l'eau possèdent, elles aussi, un niveau d'incertitude non négligeable. Il est donc nécessaire d'en tenir compte pour comparer objectivement la simulation d'ensemble avec les observations. Ce sera le cas dans la section suivante.

6.2 Fiabilité statistique de l’ensemble

L’objectif de cette section est de présenter une comparaison statistique de la simulation d’ensemble avec les observations de la couleur de l’eau SeaWIFS. On s’intéressera en particulier à la fiabilité statistique, mesure de la cohérence entre les distributions de probabilité des membres de l’ensemble et les distributions statistiques des observations.

Dans la première partie de cette section, nous décrivons la démarche d’évaluation statistique utilisée. Nous expliquerons en particulier comment les rangs des observations SeaWIFS dans l’ensemble sont calculés et nous définirons de manière simple le critère de fiabilité statistique qui sera examiné dans cette section. Dans la seconde partie de ce chapitre, ces rangs seront présentés sous forme d’histogrammes. L’analyse des histogrammes de rangs nous permettra d’évaluer la variabilité spatiale et temporelle de la fiabilité statistique de la simulation d’ensemble *stoUpd*.

6.2.1 Démarche

Fiabilité

De manière courante, on dit d’une mesure (ou d’un objet) qu’elle est fiable si elle reflète le résultat observé, attendu. Ma voiture est fiable si l’événement unique “démarrer” est atteint au moment où l’on tourne la clé. La fiabilité statistique utilisée ici n’est autre qu’un élargissement de cette définition au contexte probabiliste. Elle ne repose plus sur un événement unique mais sur un ensemble continu d’événements caractérisés par une distribution de probabilité. La fiabilité statistique représente alors la cohérence entre l’estimation des distributions de probabilité, définies ici par les différents membres de la simulation d’ensemble et la distribution statistique des vérifications, ici les observations de la couleur de l’eau. Par exemple, si 10% des observations estiment à plus de 4 mg/m³ les concentrations en chlorophylle de surface de l’Atlantique Nord, alors 10% des membres d’une simulation d’ensemble devront respecter ce chiffre pour que l’ensemble soit fiable.

À noter que la fiabilité est une condition nécessaire à la Bayesianité de l’ensemble. Cela signifie que l’ensemble doit être fiable pour que ses distributions de probabilité puissent être solutions du théorème de Bayes (cf. section 2.4). La fiabilité peut être évaluée à partir de nombreux outils statistiques. On peut citer le score de Brier, (Brier, 1950) dont l’extension à plusieurs événements continus, le CRPS (Continuous Rank Probability Score) sera examiné dans le chapitre suivant. Dans cette section, nous utiliserons uniquement les rangs des observations SeaWIFS pour évaluer la fiabilité de la simulation d’ensemble.

Rangs des observations

Dans cette section, la fiabilité de l’ensemble est uniquement évaluée à partir de la répartition des observations parmi les intervalles des membres de l’ensemble, c’est-à-dire des rangs des observations, que l’on présentera sous forme d’histogrammes. L’utilisation de ce diagnostic est principalement hérité de la prévision d’ensemble météorologique dans laquelle il a été initialement introduit par Anderson (1996). Il est encore très rarement utilisé en océanographie. L’utilisation de ce diagnostic présente pourtant l’avantage de

synthétiser une large quantité d'information et de parcourir l'espace des solutions sans faire d'hypothèses a priori sur les distributions de probabilité. Pour une description complète, le lecteur peut se référer à Hamill (2001) ou Candille and Talagrand (2005).

-Pour calculer les rangs des observations, on procède de la manière suivante :

En chaque point de grille, les membres de l'ensemble sont classés suivant l'ordre croissant des concentrations en chlorophylle. Le rang de l'ensemble est égal à l'indice de l'intervalle entre les membres classés dans lequel est située l'observation. Le rang 0 correspond à une observation inférieure au minimum de l'ensemble et le rang 1 correspond à une observation supérieure au maximum de l'ensemble. Avec 60 membres, les observations se répartissent donc parmi 61 intervalles (59 intervalles entre les membres plus les indices 0 et 1). Le calcul des rangs est ensuite réitéré sur plusieurs points de grille et/ou pendant plusieurs pas de temps. Le cumul des rangs permet d'établir un histogramme (de rangs) dont l'allure va caractériser la fiabilité de l'ensemble.

Pour qu'un ensemble soit fiable, chacun des membres de l'ensemble doit être capable de représenter un événement (observation) avec la même fréquence d'occurrence. En d'autres termes, la probabilité d'une observation d'être indexée du rang i avec $0 \leq i \leq 1$ doit être de $1/N+1$ avec N le nombre de membres de l'ensemble. Un ensemble parfaitement fiable possédera donc un histogramme de rangs plat. On distingue généralement 3 formes d'histogrammes de rangs :

- l'histogramme en \cup , qui signifie qu'un nombre trop important d'observations est en dehors des bornes de l'ensemble. L'ensemble est en général sous-dispersif (ou possède un double biais, Hamill, 2001).
- l'histogramme en \cap , qui signifie que la majorité des observations est incluse entre seulement quelques intervalles des membres de l'ensemble. L'ensemble est surdispersif.
- l'histogramme asymétrique, qui signifie que les observations sont réparties dans la partie haute ou basse de l'ensemble. La dispersion présente un biais.

6.2.2 Histogrammes de rangs

6.2.2.1 Caractéristiques spatiales

Pour être statistiquement significatifs, les histogrammes de rangs doivent être basés sur un grand nombre d'observations. Sur la figure 6.6, les histogrammes sont ainsi construits avec les rangs de toutes les observations disponibles sur le domaine à une date donnée (en moyenne environ 300 000 données). La fiabilité estimée est donc ici une fiabilité spatiale, que l'on évalue en figure 6.6 pour le 31 mars, le 15 mai et 27 octobre 2005. Les cartes de la distribution spatiale de ces rangs sont également présentées. À noter que l'interpolation des données du modèle au $1/4^\circ$ sur la grille des observations SeaWIFS au $1/12^\circ$ est faite par interpolation bilinéaire.

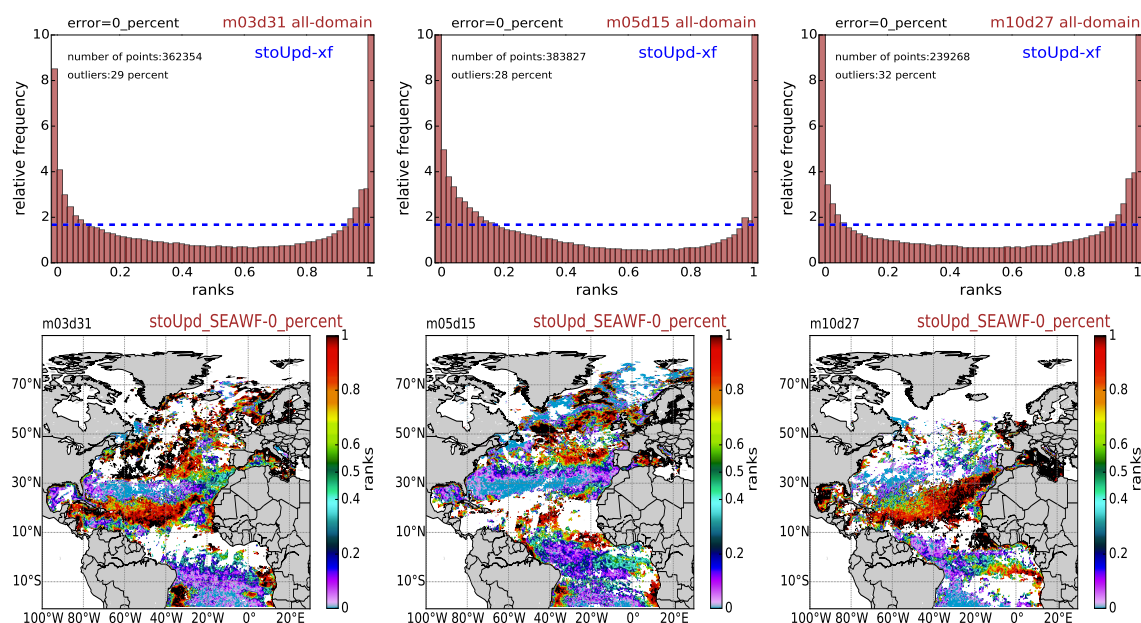


FIGURE 6.6 – Rangings des observations SeaWiFS dans l'ensemble calculés sans erreur d'observation sur la totalité du domaine pour le 31 mars (1^{ère} colonne), le 15 mai (2^{ème} colonne) et le 27 octobre (3^{ème} colonne) 2005. La ligne du haut présente ces rangs sous forme d'histogrammes. La fréquence d'occurrence est exprimée en pourcentage du nombre total d'observations. Les lignes en pointillés bleus précisent la fréquence d'occurrence d'un histogramme plat (parfaite fiabilité de l'ensemble.) Les cartes de la ligne du bas affichent la répartition spatiale de ces rangs.

Le point principal, mis en avant par ces premiers histogrammes, est que l'ensemble possède un degré de fiabilité globalement satisfaisant. Ces histogrammes sont même qualitativement comparables à ceux obtenus par Candille et al. (2015) dans l'évaluation probabiliste de la température et de la salinité du Gulf Stream et à ceux qui sont obtenus dans d'autres champs d'étude comme la prévision hydrologique (Rousset-Regimbeau, 2007). Malgré une tendance générale à la sous-dispersion, les observations ne s'accumulent pas uniquement dans la partie haute ou la partie basse de l'ensemble. Tous les membres de l'ensemble sont donc nécessaires pour capturer l'information des observations.

En plus de caractériser la fiabilité statistique, l'analyse des histogrammes permet de confirmer et de quantifier certains des résultats mis en avant par les cartes de quantiles de surface (figures 6.1 et 6.2). La prédominance de l'allure en \cup témoigne ainsi d'une part trop importante d'observations situées en dehors des bornes de l'ensemble (environ 30%). En mai, l'allure asymétrique de l'histogramme de rangs souligne la tendance du modèle à surestimer la production en chlorophylle.

La distribution spatiale des rangs présente une structure plutôt « patchy » qui dénote d'une répartition des observations dans l'ensemble possédant une forte dépendance spatiale. On peut distinguer simultanément des zones où la chlorophylle est toujours surestimée (comme autour de 30°N) et des régions où elle est systématiquement sous-estimée (comme au niveau du gyre subtropical). L'allure en \cup dénote également de la présence d'un double biais spatial qui s'explique par le mélange statistique de régimes biogéochimiques différents.

En définitive, la variabilité spatiale indique que les histogrammes de rangs doivent être considérés en fonction des différents régimes biologiques. Ce point sera donc l'objet de la section suivante. Avant cela, nous analysons l'impact des erreurs d'observation sur la répartition des rangs.

Impact de l'erreur d'observation

Jusqu'ici, toutes les comparaisons avec les observations SeaWIFS ont été faites sans prendre en compte les erreurs d'observation. Les données satellites étant elles aussi incertaines (cf. chapitre 3), il est important de considérer ces incertitudes pour évaluer objectivement la qualité de l'ensemble. Le besoin d'intégrer ces erreurs est de plus motivé par leur prise en compte en assimilation de données.

Pour le calcul des rangs, une solution consiste à augmenter la dispersion en ajoutant aux membres de l'ensemble un bruit blanc normalement distribué dont l'écart-type est défini par l'incertitude des observations. L'utilisation de cette méthode a été préalablement testée et validée sur des ensembles de prévision météorologique de l'ECMWF dans les travaux de Saetra et al. (2004). Sur les données SeaWIFS, l'erreur communément admise est d'environ 30% (Gregg and Casey, 2004). Pour chaque observation \mathbf{y}^0 , un bruit blanc gaussien de moyenne nulle et d'écart-type égal à 30% ($\sigma = 0.3\mathbf{y}^0$) est ajouté à chacun des membres x_i selon l'équation 6.1.

$$x'_i = x_i + \alpha \quad \text{avec} \quad \alpha \in \mathcal{N}(\mu = 0, \sigma^2 I) \quad (6.1)$$

Calculée de cette manière, l'erreur d'observation est donc indirectement simulée en générant un nouvel ensemble dont la dispersion est étendue proportionnellement à l'incertitude des observations. Les rangs sont ensuite calculés dans les intervalles entre les membres x'_i de ce nouvel ensemble. Même si l'introduction de nombres aléatoires altère l'unicité de la solution, l'allure des histogrammes dépend très peu de la suite aléatoire utilisée.

La figure 6.7 présente les histogrammes et les cartes de surface des rangs des observations SeaWIFS dans l'ensemble, calculés cette fois en tenant compte d'une erreur d'observation de 30%. Lorsque l'on compare les histogrammes de cette figure avec ceux de la figure 6.6, on constate que l'allure en \cup est conservée (tout de même réduite), mais que le nombre d'observations en dehors de l'ensemble (« outliers ») a été considérablement réduit (en moyenne divisé par 2). Les observations incluses dans l'enveloppe de l'ensemble sont, de manière générale, bien réparties entre tous les intervalles définis par les différents membres.

Si les différenciations spatiales de grande échelle, soulignées précédemment, restent identiques, les cartes des rangs de la figure 6.7 présentent des distributions spatiales significativement plus hétérogènes. Bien que l'ensemble reste légèrement sous-dispersif, considérer une erreur d'observation permet donc de réduire les biais et d'améliorer l'estimation de la fiabilité (déjà satisfaisante sans erreur d'observation). Les améliorations les plus significatives sont principalement à observer pendant l'hiver (31 mars) et dans la partie nord du domaine. Dans la suite de ce manuscrit, tous les rangs présentés seront calculés avec une erreur d'observation de 30%.

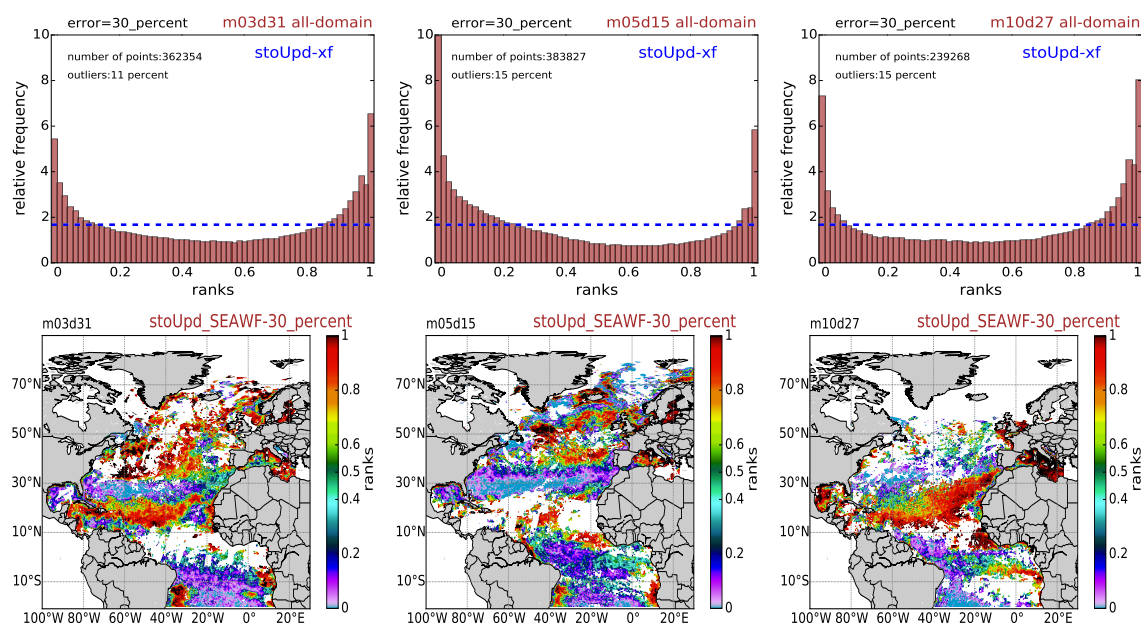


FIGURE 6.7 – Rangs des observations SeaWIFS dans l'ensemble calculés avec une erreur d'observation de 30% sur la totalité du domaine pour le 31 mars (1^{ère} colonne), le 15 mai (2^{ème} colonne) et le 27 octobre (3^{ème} colonne) 2005. La ligne du haut présente ces rangs sous forme d'histogrammes. La fréquence d'occurrence est exprimée en pourcentage du nombre total d'observations. Les lignes en pointillés bleus précisent la fréquence d'occurrence d'un histogramme plat (parfaite fiabilité de l'ensemble) Les cartes de la ligne du bas affichent la répartition spatiale de ces rangs.

6.2.2.2 Caractéristiques temporelles

L'allure des histogrammes présentés dans la partie précédente a mis en évidence l'existence d'un double biais, résultant du mélange statistique de caractéristiques biogéochimiques variables d'une région à une autre. Pour distinguer ces différences spatiales, les rangs sont désormais calculés séparément pour les provinces de Longhurst 4, 6, 7 et 9 (cf. figure 4.3 du chapitre 4 pour un rappel sur la délimitation des provinces). Sur la figure 6.8, les histogrammes de rangs sont tout d'abord construits sans distinction temporelle. L'utilisation de l'ensemble des observations SeaWIFS disponibles pendant l'année 2005 (tous les 5 jours) permet de se munir d'un large échantillon statistique compris entre 1 million (province 4) et 3 millions (province 7) de données.

Ces histogrammes montrent qu'excepté aux abords du Gulf Stream, où le fort biais positif du modèle est confirmé, la fiabilité de l'ensemble est globalement satisfaisante. Mis à part dans la province 6, moins de 10% des observations sont à l'extérieur de l'enveloppe de l'ensemble. L'existence d'« outliers » (cf. figures précédentes) est donc majoritairement le résultat de paramétrisations stochastiques parfois peu adaptées aux comportements biogéochimiques de certaines régions côtières, principalement localisées dans les hautes latitudes. Une légère sous-estimation (biais négatif) des quantités de chlorophylle peut également être observée aux alentours du gyre subtropical (province 7).

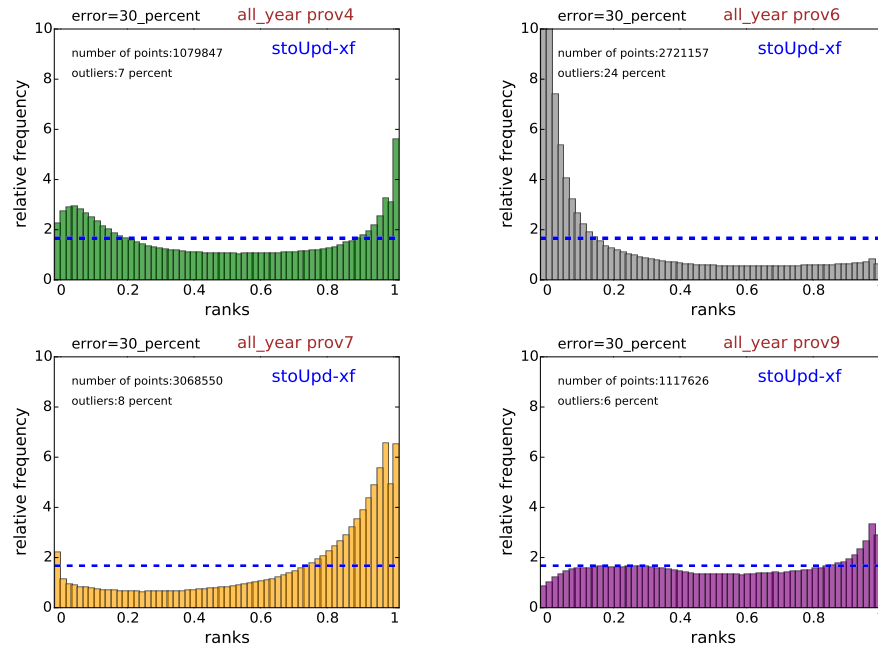


FIGURE 6.8 – Rangs des observations SeaWIFS dans l’ensemble calculés avec une erreur d’observation de 30% sur toute l’année 2005 dans les provinces de Longhurst 4 (1^{ère} ligne, 1^{ère} colonne), 6 (1^{ère} ligne, 2^{ème} colonne), 7 (2^{ème} ligne, 1^{ère} colonne) et 9 (2^{ème} ligne, 2^{ème} colonne). Les rangs sont présentés sous forme d’histogrammes. La fréquence d’occurrence est exprimée en pourcentage du nombre total d’observations. Les lignes en pointillés bleus précisent la fréquence d’occurrence d’un histogramme plat (parfaite fiabilité de l’ensemble).

Variabilités saisonnières

Dans la figure 6.9, les histogrammes de rangs sont représentés en considérant la variabilité saisonnière des régimes biogéochimiques. Dans la partie plus au nord du domaine (province 4), cette figure met en évidence une différenciation été-hiver particulièrement frappante. L’hétérogénéité des observations n’est pas toujours bien représentée et l’ensemble affiche un biais négatif relativement important pendant l’hiver. En dehors de cette période, l’ensemble capture la majorité des observations (il est même surdispersif en automne) et la fiabilité reste convaincante.

Dans les autres régions, l’allure des histogrammes est peu sensible aux variations saisonnières. Le fort biais positif existant au niveau du Gulf Stream est confirmé pour toute l’année, avec des valeurs particulièrement élevées pendant le bloom de printemps. Dans la province 9, nous pouvons remarquer un nombre important d’observations localisées dans la partie haute de l’histogramme en période estivale. Cela peut s’expliquer par l’influence de l’upwelling du Benguela, mal représentée dans le modèle.

Conclusion

Les histogrammes de rangs ont permis de montrer la capacité de l'ensemble à représenter de manière fiable l'information statistique des observations SeaWIFS. La récurrence d'un double biais apparaît comme la conséquence d'une tendance à surestimer la production en chlorophylle pendant le bloom de printemps et à la sous-estimer, plus régulièrement, dans le gyre subtropical et les zones côtières des hautes latitudes. Il est important de préciser que tenir compte d'une erreur d'observation de seulement 30% est relativement optimiste, en particulier dans les zones côtières (cf. chapitre 3). Montrer que la dispersion générée par les paramétrisations stochastiques permet d'inclure la quasi-totalité des observations en respectant leur distribution statistique est donc un point important de cette thèse.

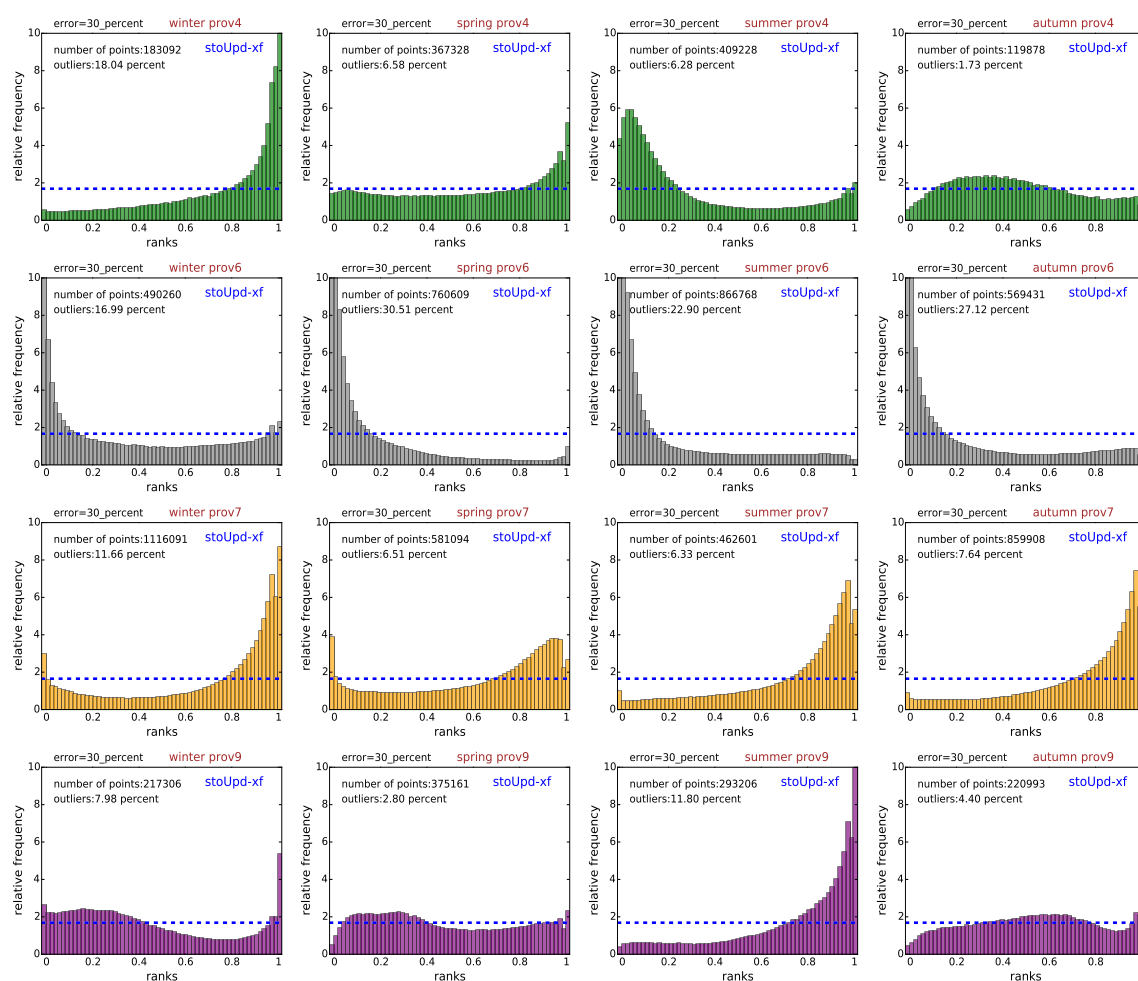


FIGURE 6.9 – Rangs des observations SeaWIFS présentés par saison, sous forme d'histogrammes. La 1^{ère} colonne correspond aux mois d'hiver (janvier à mars), la 2^{ème} colonne au printemps (avril à juin), la 3^{ème} colonne à l'été (juillet à septembre) et la dernière colonne à l'hiver (octobre à décembre). Les rangs sont calculés avec 30% d'erreur d'observation dans les provinces de Longhurst 4 (1^{ère} ligne), 6 (2^{ème} ligne), 7 (3^{ème} ligne) et 9 (4^{ème} ligne). La fréquence d'occurrence est exprimée en pourcentage du nombre total d'observations. Les lignes en pointillés bleus précisent les occurrences d'un histogramme plat (fiabilité parfaite).

6.3 Distributions verticales de la chlorophylle

Au chapitre précédent, on s'est assuré de la capacité des paramétrisations stochastiques à générer de la dispersion dans la couche euphotique à partir de coupes verticales de l'écart-type des membres de l'ensemble. Dans cette section, l'objectif est de préciser les caractéristiques spatiales et temporelles de cette dispersion.

Dans la première partie de cette section, nous adopterons une démarche analogue à celle utilisée pour la chlorophylle de surface (cf. section 6.1). La dispersion verticale de la simulation d'ensemble sera ainsi évaluée statistiquement, en s'appuyant sur la représentation des quartiles de l'ensemble. Dans la seconde partie, les structures verticales de l'ensemble seront comparées avec des observations. Nous insisterons, en particulier, sur la capacité des paramétrisations stochastiques à simuler l'existence de distributions verticales pouvant être assimilées à différentes espèces au sein d'une même classe de phytoplancton.

6.3.1 Dispersion d'ensemble

Coupes longitudinales

Pour analyser la dispersion d'ensemble de la chlorophylle dans la colonne d'eau, la figure 6.10 présente les quartiles des membres de l'ensemble le long de coupes longitudinales à 30°W. Ces résultats sont présentés pour le 15 mai et le 27 octobre 2005. Le calcul des maxima, minima, de la médiane et des quartiles est identique à celui que nous avons fait pour la chlorophylle de surface (section 6.1.1). De la même manière, ces structures verticales ne représentent donc pas un état biogéochimique du modèle, mais un état composite des 60 membres de la simulation d'ensemble.

Comme nous pouvions le présager à partir des coupes verticales d'écart-type présentées au chapitre précédent, les paramétrisations stochastiques produisent une importante dispersion dans la couche euphotique. Contrairement au cas de la chlorophylle de surface, l'enveloppe statistique n'est pas particulièrement plus développée pendant le bloom de printemps. Dans les zones où les maxima de subsurface sont prédominants, une dispersion maximale est observée à leurs niveaux de profondeur. En l'absence de maxima de subsurface, la dispersion reste relativement constante dans les 100 premiers mètres. Elle diminue ensuite rapidement jusqu'à devenir nulle en dessous de la zone euphotique (environ 150 mètres).

La figure 6.10 montre également que les différents quantiles présentent des structures verticales comparables. Leurs caractéristiques sont, de plus, identiques à celles de la simulation déterministe (cf. chapitre précédent). Je précise que ce point peut s'expliquer par l'invariance verticale des perturbations stochastiques (cf. section 5.1.1.2). Dans les coupes verticales des minima de l'ensemble, on constate toutefois une prédominance accentuée des maxima de subsurface (DCM). A contrario, les zones de prédominance des DCM ont tendance à être réduites dans les coupes des maxima de l'ensemble (jusqu'à 40°N). Ce constat permet de présupposer l'existence de différents types de profils verticaux. Ce point sera approfondi dans la section 6.3.2.2.

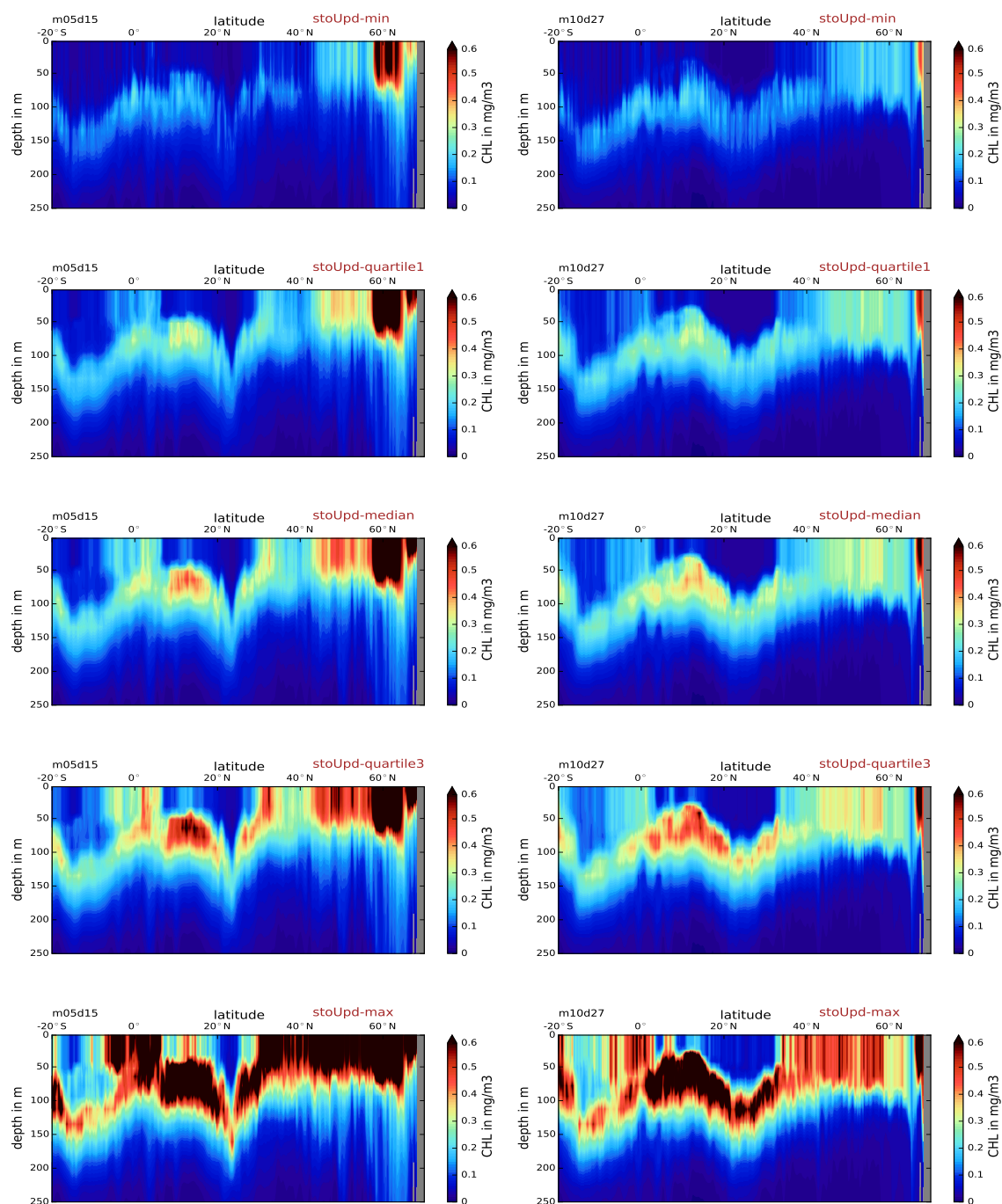


FIGURE 6.10 – Réponse statistique de la chlorophylle aux paramétrisations stochastiques telle que décrite par la simulation d'ensemble pour le 15 mai (1^{ère} colonne) et le 27 octobre (2^{ème} colonne) 2005. Les quantiles sont identifiés le long de coupes longitudinales à 30°W. De haut en bas, on représente le minimum de l'ensemble, le premier quartile, la médiane, le troisième quartile et le maximum de l'ensemble.

Sur la figure 6.10, nous pouvons également constater que l'allure des distributions de probabilité est comparable à celle obtenue pour la chlorophylle de surface. L'ensemble est ainsi caractérisé par une large étendue et une distribution asymétrique, définie par un écart maximal entre le 3^{ème} quartile et le maximum de l'ensemble. Le caractère non gaussien des distributions de probabilité est donc confirmé dans toute la colonne d'eau.

Évolutions temporelles

Afin d'analyser l'évolution temporelle de la dispersion d'ensemble sur la verticale, la figure 6.11 présente les séries temporelles des minima, des maxima et de la médiane de l'ensemble aux points de grille stations *STAT_B* et *INDIA*. La localisation géographique de ces points de grille a été précisée sur la figure 6.5. Pour des raisons de lisibilité les 1^{er} et 3^{ème} quartiles de l'ensemble ne sont pas représentés. Sur cette figure, ces quantiles sont mis en parallèle avec la structure verticale de la simulation déterministe *detIni*.

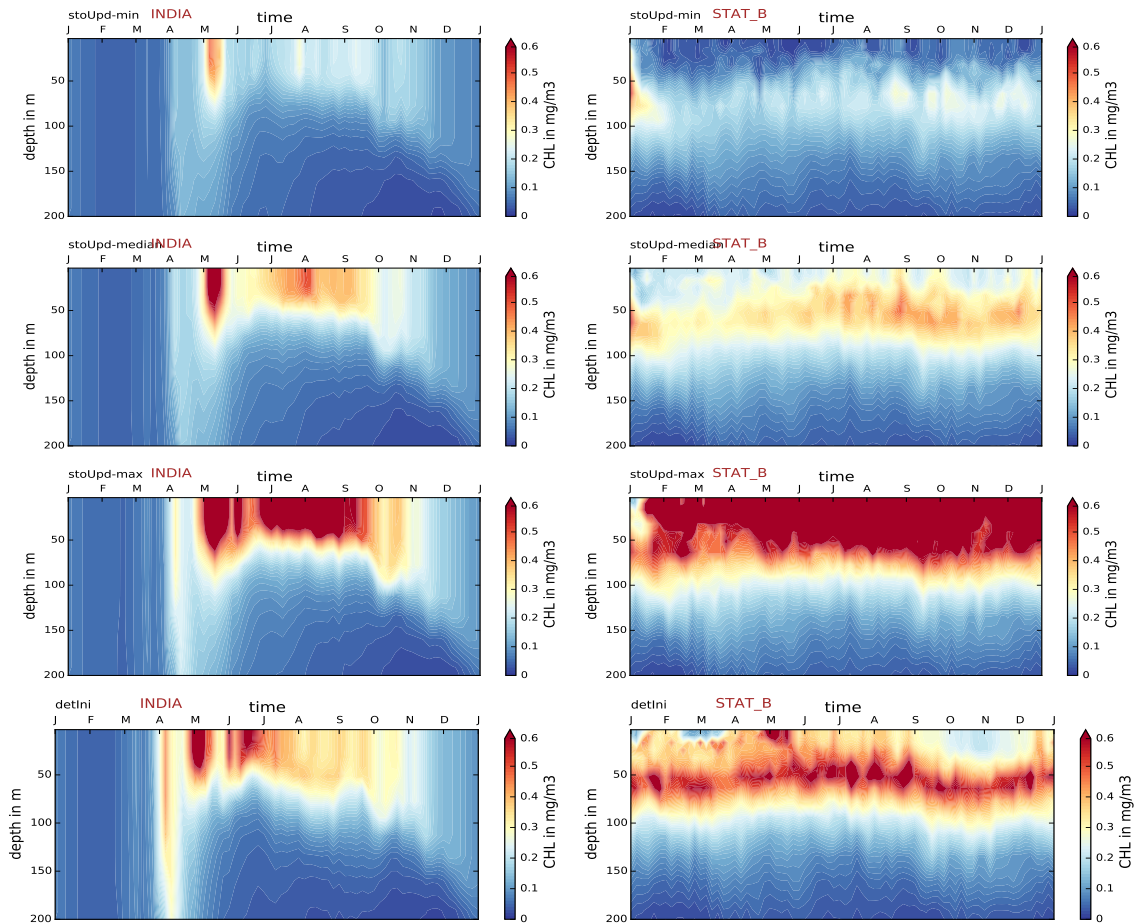


FIGURE 6.11 – Séries temporelles du minimum (1^{ère} ligne), de la médiane (2^{ème} ligne) et du maximum de l'ensemble (3^{ème} ligne) aux stations *STAT_B* (1^{ème} colonne) et *INDIA* (2^{ème} colonne). Les séries temporelles de la simulation *detIni* sont présentées (dernière ligne).

À la station *STAT_B*, nous pouvons constater que les paramétrisations stochastiques génèrent, dès le début de la simulation, de la dispersion dans toute la zone euphotique. Comme préalablement constaté (cf. figure 6.4), il faut attendre l’initiation du bloom de printemps pour qu’une dispersion s’établisse à la station *INDIA*. Les difficultés pour générer de la dispersion en hiver dans la partie nord du domaine (observées en section 6.1.2 sur les séries temporelles de la chlorophylle de surface), sont ainsi confirmées pour l’ensemble de la couche euphotique. Excepté durant cette période (non représentative de l’ensemble de la partie nord du domaine), une dispersion importante est toujours générée. Ces séries temporelles montrent également que la représentation des états médians de la simulation d’ensemble est proche celle de la simulation *detIni*. L’information verticale de la simulation déterministe est donc incluse dans l’ensemble avec une forte probabilité d’existence.

6.3.2 Comparaison aux observations

6.3.2.1 Climatologie verticale

Jusqu’à présent, l’analyse de la dispersion verticale de l’ensemble s’est faite sans effectuer de comparaisons avec des observations. Elle a permis de vérifier que les paramétrisations stochastiques produisent de la variabilité, en conservant toujours des structures verticales cohérentes avec celles de la simulation déterministe. Ce constat fait, il est maintenant envisageable de moyenniser spatialement la simulation d’ensemble afin de pouvoir la comparer aux données climatologiques de la chlorophylle, utilisées précédemment pour comparer les modèles PISCES et LOBSTER. Je précise que le niveau d’erreur de ces profils climatologiques est relativement élevé (Sauzède et al., 2015). L’objectif n’étant pas de quantifier les écarts avec la simulation d’ensemble, nous n’en tiendrons pas compte ici. Étant donné le faible échantillon statistique de ces données, nous n’évaluerons pas non plus la cohérence statistique entre l’ensemble et ces observations. L’objectif de cette section est uniquement de s’assurer que le niveau d’incertitude, produit par les paramétrisations stochastiques sur la verticale, est suffisant pour pouvoir comparer la simulation d’ensemble à des observations.

La figure 6.12 (page suivante) permet de comparer l’enveloppe de la distribution verticale de l’ensemble avec des profils verticaux climatologiques dans les provinces de Longhurst 4, 6, 7 et 9. L’enveloppe de l’ensemble est calculée à partir de la moyenne mensuelle des minima, maxima et de la médiane de l’ensemble en chaque point de grille, ensuite moyennée par province de Longhurst (plutôt qu’à partir de l’enveloppe de l’ensemble moyen). Cette figure 6.12 étant équivalente à la figure 4.2 du chapitre 4, les profils verticaux de la simulation déterministe sont également présentés.

L’analyse de la figure 6.12 montre que l’ensemble est capable de capturer la quasi-totalité de l’information verticale contenue dans les observations climatologiques. Même lorsque la simulation déterministe présente des profils verticaux éloignés de ceux des observations (p.ex. mois de mai, provinces 4 et 6), l’étendue verticale de l’ensemble permet de se rapprocher significativement de la climatologie. À noter que les profils verticaux de la simulation déterministe sont eux-aussi inclus dans l’ensemble. Comme déjà évoqué, ils suivent en général la médiane de l’ensemble.

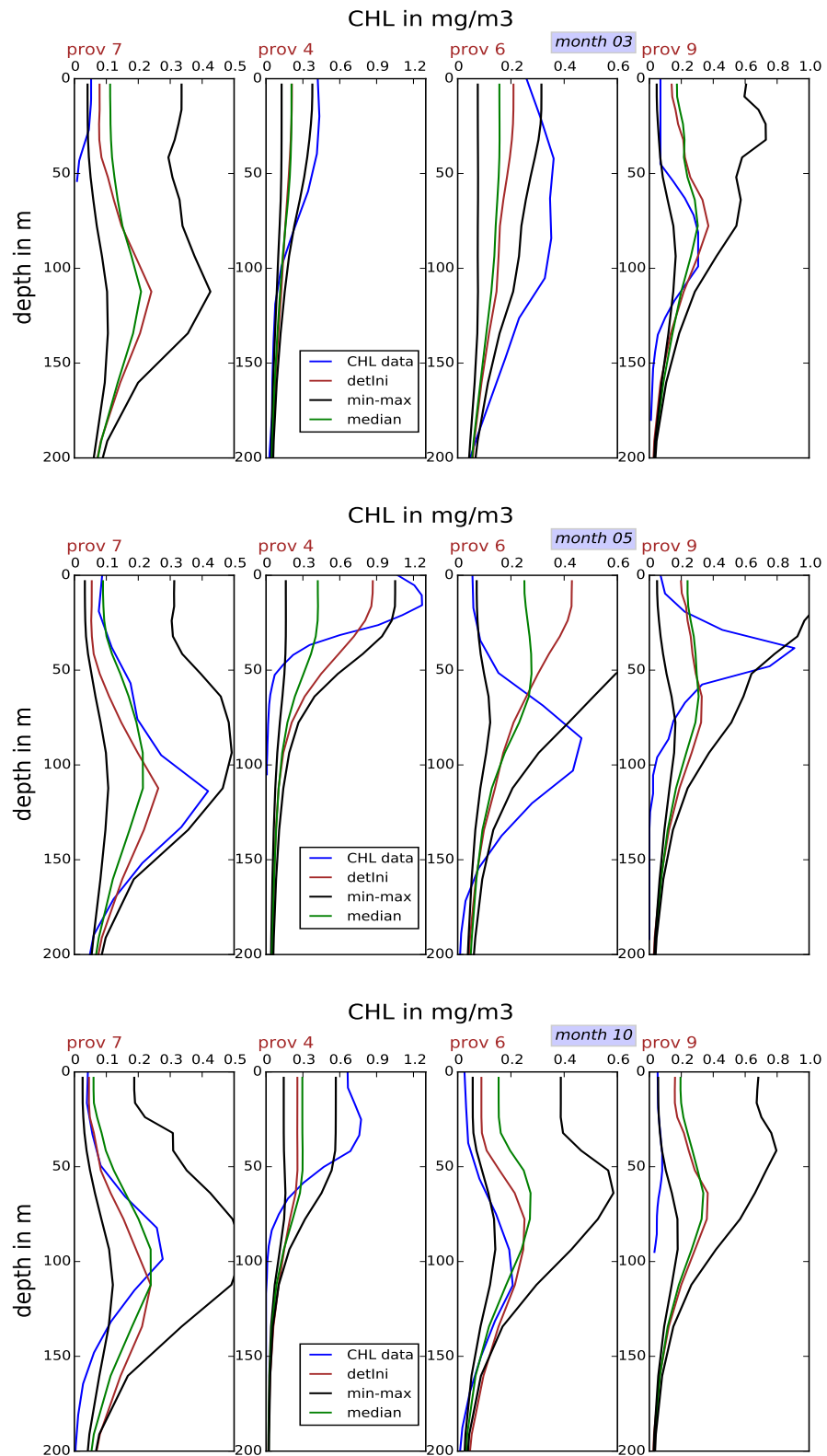


FIGURE 6.12 – Profils verticaux en chlorophylle en moyenne spatiale pour les provinces de Longhurst 4, 6, 7 et 9. Les courbes noires représentent les moyennes mensuelles par province de Longhurst des minimums de l'ensemble et des maximums de l'ensemble. La courbe bleue représente les données de chlorophylle climatologiques et la courbe rouge représente les profils verticaux de la simulation déterministe *detIni*.

Des écarts importants entre la médiane et la simulation déterministe peuvent néanmoins être observés, notamment au mois de mai dans les provinces 4 et 6. Dans la région du Gulf Stream (province 6), on constate que la diminution (en moyenne) de la chlorophylle permet à la médiane de l'ensemble d'être plus proche des observations que la simulation déterministe. Au contraire, à cause d'une tendance à la surestimation, les observations restent proches du minimum de l'ensemble dans les hautes couches de la zone euphotique (50 premiers mètres). La dispersion d'ensemble diminuant rapidement avec la profondeur, l'existence d'un fort maximum de subsurface est ensuite difficilement capturé par l'ensemble. Une analyse approfondie des caractéristiques de l'ensemble dans cette région (station *BATS*) fera l'objet de la section suivante. Dans la partie nord du domaine (province 4), les observations sont toujours proches du maximum de l'ensemble dans les premiers mètres de la couche euphotique. La rapide diminution de la dispersion avec la profondeur est cette fois cohérente avec la diminution rapide des quantités de chlorophylle avec la profondeur (due à une couche euphotique moins profonde). Dans ces régions, la diminution de la chlorophylle produite par le réajustement des paramètres du modèle PISCES n'est pas entièrement compensée par l'effet moyen des paramétrisations stochastiques. Cela renforce l'idée qu'un ajustement plus fin des paramètres stochastiques du modèle permettrait d'améliorer la différenciation nord-sud des régimes biogéochimiques.

Pour conclure, malgré une mise en place empirique des paramétrisations stochastiques, la dispersion produite sur la verticale permet à l'ensemble de contenir l'information verticale présente dans les observations climatologiques proposées par [Sauzede et al. \(2015\)](#).

6.3.2.2 Diversité des comportements biologiques

Dans les sections précédentes, les profils verticaux ont été examinés de manière statistique. Si cela a permis d'analyser la dispersion de la simulation d'ensemble, cette vision ne suffit pas pour évaluer sa capacité à distinguer une variété de profils verticaux. En plus de permettre de générer de la dispersion, les paramétrisations stochastiques présentent justement l'intérêt de pouvoir représenter simultanément plusieurs comportements biogéochimiques. C'est ce que nous allons voir dans cette section.

Pour analyser la capacité du système stochastique à simuler une diversité biologique, la figure 6.13 présente les profils verticaux de chacun des membres de la simulation d'ensemble. Ces distributions verticales sont représentés à la station *BATS* et sont comparées avec des données d'observation in-situ disponibles sur le site du BIOS (Bermuda Institute of Ocean Sciences) à l'adresse <http://bats.bios.edu>. Je rappelle que les analyses précédentes ont mis en évidence les difficultés de l'ensemble pour inclure les observations de la couleur de l'eau dans la région où se situe cette station. Elle présente toutefois l'avantage de disposer d'un jeu d'observations au large spectre spatio-temporel.

Le premier point mis en évidence par la figure 6.13 est la capacité de l'ensemble à simuler l'existence d'une large variété de profils verticaux. À chaque pas de temps, il est possible d'isoler des profils affichant de forts maxima de subsurface et des profils affichant, au contraire, une diminution rapide de la chlorophylle avec la profondeur. Nous pouvons donc simuler différents comportements verticaux, pouvant être associés à l'existence de plusieurs régimes biologiques.

En contradiction avec certains résultats précédents, nous pouvons observer que l'ensemble n'inclut pas toujours la simulation déterministe. Cette caractéristique est particulière à certains points de la zone du Gulf Stream (et des hautes latitudes) dans laquelle le réajustement des paramètres du modèle n'est pas entièrement compensé par l'effet moyen dû à l'introduction des perturbations stochastiques. Cet effet est ici plutôt positif puisque la simulation déterministe surévaluait la production en chlorophylle.

Comme nous avons déjà pu le constater, une grande partie des membres de l'ensemble ont néanmoins encore tendance à surestimer les concentrations en chlorophylle dans cette région. L'enveloppe de l'ensemble peine ainsi à capturer toute l'information verticale des observations qui restent toujours proches du minimum de l'ensemble.

Dans la figure 6.13, les profils verticaux dont l'allure est proche de celle des observations sont surlignés en noir. Cela permet de montrer que même lorsque l'ensemble n'inclut pas les observations (p.ex. 11 novembre 2005), nous pouvant tout de même trouver une réalisation de l'ensemble dont l'allure est comparable aux observations. Il serait donc possible de retrouver par assimilation de données, la valeur des paramètres biogéochimiques permettant de simuler, de façon déterministe, un profil vertical cohérent avec les observations.

La cohérence de la représentation de la chlorophylle de surface apparaît ici comme une condition nécessaire (mais pas suffisante). De manière presque systématique, les réalisations de l'ensemble dont les profils verticaux sont en accord avec les données in-situ sont ceux dont la représentation de la chlorophylle de surface est la plus proche des observations. En s'appuyant sur l'hypothèse, pour la première fois mise en évidence par Morel and Berthon (1989), selon laquelle l'allure des profils verticaux de la chlorophylle dépend fortement des concentrations de surface, on propose, en figure 6.14, de classer l'ensemble des distributions verticales obtenues en fonction de leur concentration en chlorophylle de surface.

Globalement, la figure 6.14 permet de distinguer 2 grands types de profils verticaux : les profils verticaux présentant des maxima de subsurface et les profils verticaux sans maxima de subsurface. Lorsque la concentration en chlorophylle de surface est faible (1^{ère} ligne, 2 premières colonnes), l'ensemble simule toujours l'existence de maxima de subsurface. Lorsque la chlorophylle de surface augmente (1^{ère} ligne, 3^{ème} colonne et 2^{ème} ligne, 2 premières colonnes), l'existence de maxima de subsurface n'apparaît plus comme la seule structure verticale possible de l'ensemble. Pour de fortes concentrations en chlorophylle (2^{ème} ligne, 3^{ème} colonne), les profils verticaux ne simulent jamais de maxima de subsurface. D'après Uitz et al. (2006), la gamme des concentrations en chlorophylle de surface peut être reliée au type d'espèce prédominante du milieu, auquel sera associé un comportement vertical particulier. Représenter différents types de profils verticaux, comme c'est le cas ici, pourrait ainsi signifier que l'on simule l'existence conjuguée de plusieurs espèces au sein des 2 classes de phytoplancton du modèle PISCES.

En perturbant de manière stochastique les paramètres, c'est-à-dire en faisant varier spatialement et temporellement leur valeur, la part de prédominance relative des 2 classes de phytoplancton est vraisemblablement modifiée. Il est en effet possible que dans un membre, les diatomées présentent régulièrement un taux de croissance très faible comparé aux taux de croissance moyens de l'ensemble. C'est alors le comportement biogéochimique du nanophytoplancton qui régira la distribution de la chlorophylle.

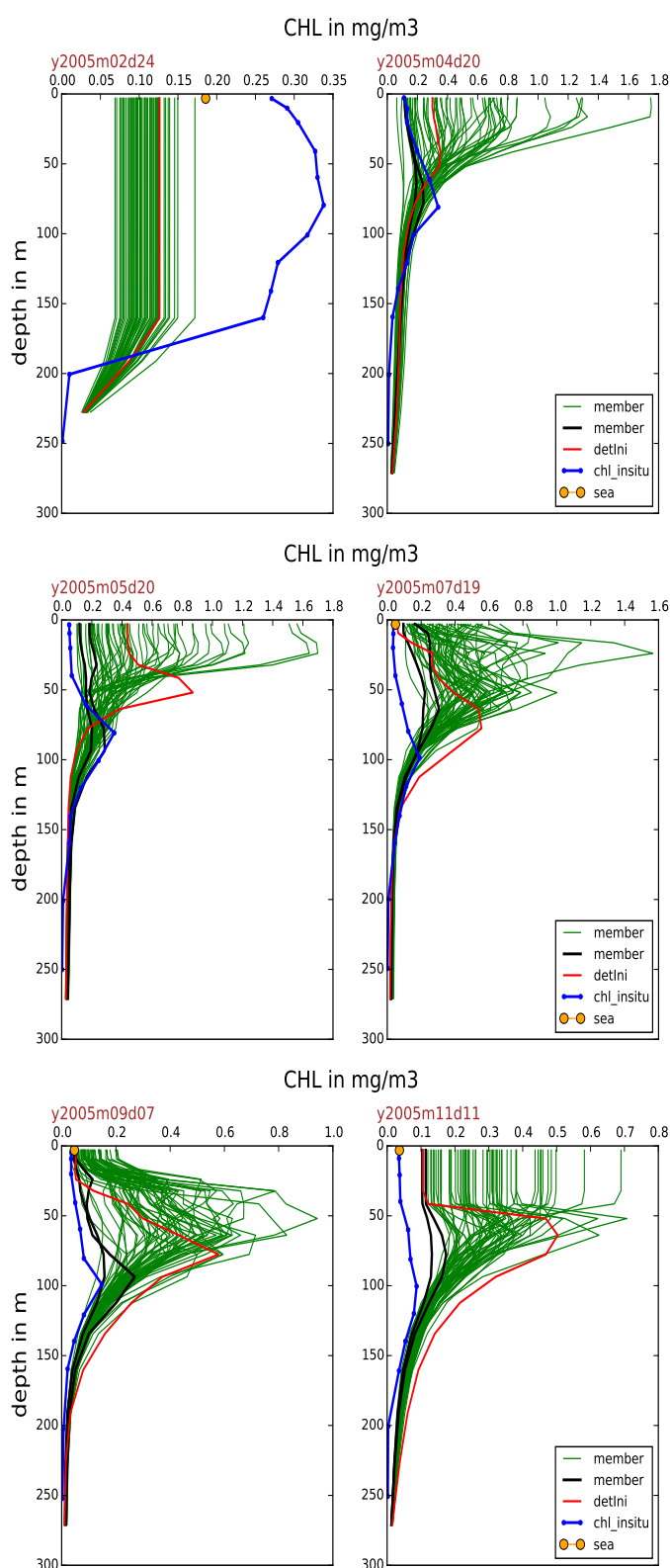


FIGURE 6.13 – Profils verticaux des 60 membres de la simulation d'ensemble au point de grille le plus proche de la station BATS. Les observations disponibles à cette station sont représentées en bleu et les membres de l'ensemble en vert. Pour chaque date les réalisations de l'ensemble dont l'allure du profil vertical se rapproche le plus de l'observation sont représentées en noir. Le profil vertical de la simulation déterministe est représenté en rouge.

Au chapitre 4 section 4.3.3, nous avons constaté que la surestimation de la production en chlorophylle était due à un bloom printannier de diatomées. Lorsque l'ensemble simule uniquement des faibles concentrations en chlorophylle, le modèle simule une prédominance du nanophytoplancton. Les concentrations en chlorophylle seront alors mieux reproduites par le modèle. À noter que d'autres types de relations peuvent être également constatés tel que le lien entre la chlorophylle de surface et la profondeur des maxima de subsurface.

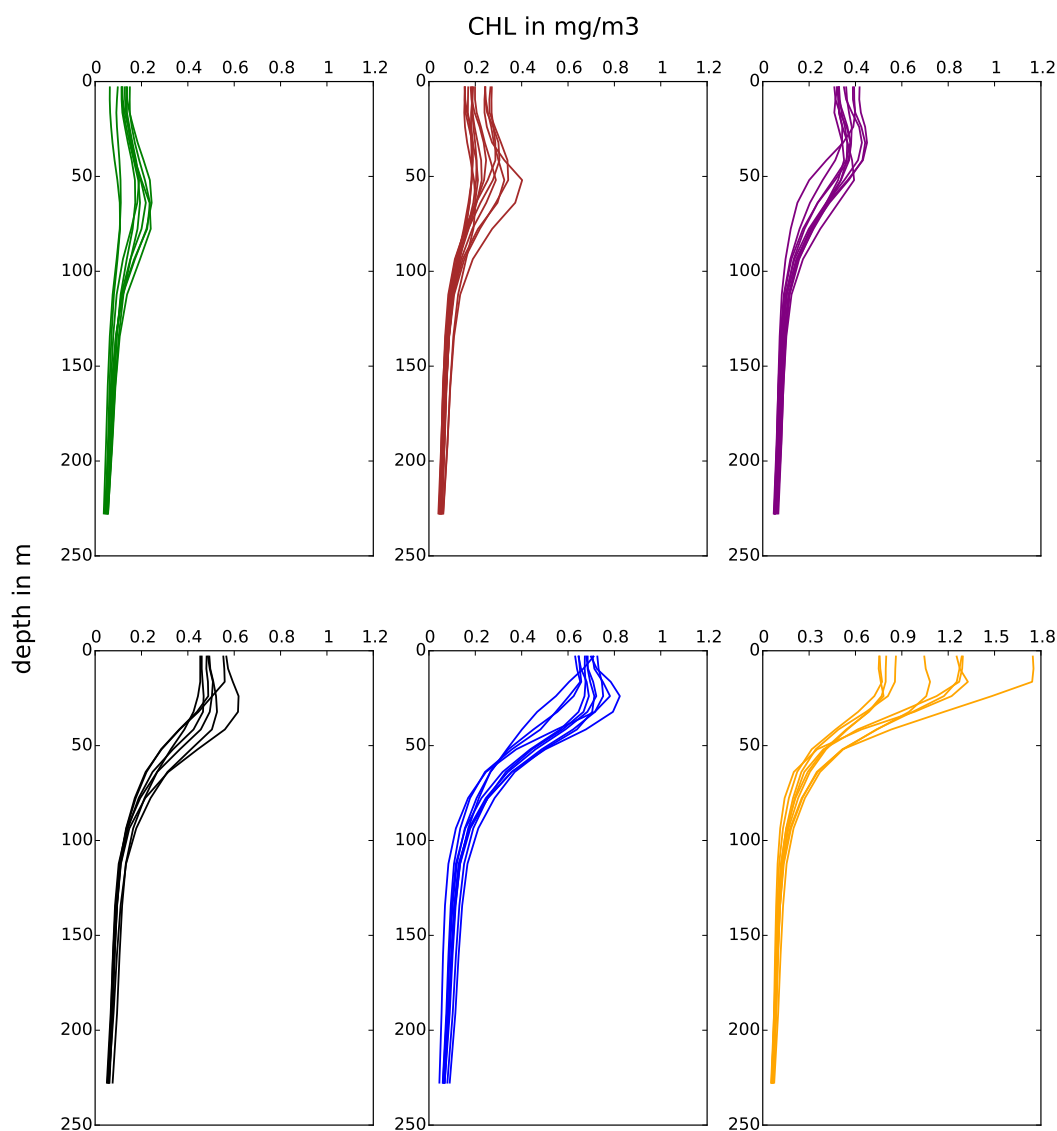


FIGURE 6.14 – Profils verticaux des 60 membres de la simulation d'ensemble classés en fonction des concentrations en chlorophylle de surface. Les profils sont rangés comme suit : Pour la ligne du haut : 1^{ère} colonne $\text{CHL} \leq 0.15 \text{ mg/m}^3$; 2^{ème} colonne $0.15 \text{ mg/m}^3 < \text{CHL} \leq 0.3 \text{ mg/m}^3$, 3^{ème} colonne $0.3 \text{ mg/m}^3 < \text{CHL} \leq 0.45 \text{ mg/m}^3$. Pour la ligne du bas : 1^{ère} colonne $0.45 \text{ mg/m}^3 < \text{CHL} \leq 0.6 \text{ mg/m}^3$; 2^{ème} colonne $0.6 \text{ mg/m}^3 < \text{CHL} \leq 0.75 \text{ mg/m}^3$; 3^{ème} colonne $\text{CHL} > 0.75 \text{ mg/m}^3$.

6.4 Conclusion

Dans ce chapitre, afin d'exploiter l'information de la simulation d'ensemble, nous avons utilisé une approche statistique présentant l'avantage de ne pas nécessiter d'hypothèses a priori sur l'allure des distributions de probabilité. Une analyse approfondie des caractéristiques dispersives de l'ensemble a ainsi pu être réalisée.

Un des avantages d'une simulation d'ensemble, mis en avant dans ce chapitre, est de pouvoir effectuer une comparaison objective avec des observations. Une approche probabiliste permet en effet de caractériser le niveau de fiabilité d'une simulation d'ensemble. Plutôt que de quantifier l'écart d'un modèle avec les observations (cas déterministe), l'approche probabiliste fournit une mesure de probabilité à la simulation, par le modèle, d'un événement observé. L'association future d'un coût monétaire aux prévisions opérationnelles, notamment pour la pêche, va très certainement favoriser ce type d'évaluation qui ne peut être réalisé qu'à partir de simulations d'ensemble.

Dans un cadre plus axé « perspectives », nous avons pu mettre en avant le potentiel de notre approche probabiliste pour simuler des comportements biogéochimiques non explicitement représentés dans le modèle. En plus de montrer que notre démarche permettait d'apporter des informations statistiques qui seront pertinentes pour l'assimilation de données, l'objectif sous-jacent de ce chapitre a ainsi été d'encourager son utilisation en modélisation biogéochimique. À la vue de l'impact des paramétrisations stochastiques, on peut même se demander, en extrapolant quelque peu, si il peut être justifiable d'envisager la prévision en biogéochimie, sans tenir explicitement compte de l'incertitude due à la formulation des modèles.

On rappelle les principaux résultats de ce chapitre.

- La dispersion de la simulation d'ensemble permet de capturer la majorité des observations de la couleur de l'eau SeaWIFS.
- La simulation d'ensemble présente un niveau de fiabilité satisfaisant.
- Les structures verticales de la simulation d'ensemble sont cohérentes avec des données climatologiques.
- Les paramétrisations stochastiques permettent de simuler une variété de comportements biologiques pouvant être associée à une diversité d'espèces.
- L'information probabiliste d'une simulation d'ensemble doit être considérée de manière statistique.
- L'approche probabiliste peut être perçue comme une alternative crédible à la modélisation biogéochimique.

Dans le chapitre suivant, nous présenterons une séquence d'analyses d'ensemble réalisée à partir d'observations de la couleur de l'eau.

Chapitre 7

Une séquence d'analyses 3D de données de la couleur de l'eau

Sommaire

7.1	Méthode d'analyse d'ensemble	139
7.1.1	Le filtre de Kalman	139
7.1.2	Le filtre de Kalman transformé (ETKF)	140
7.1.3	Prise en compte des non-gaussianités	143
7.2	Un nouvel ensemble d'états actualisés	145
7.2.1	Impact des analyses sur l'état moyen	146
7.2.2	Incertitudes des analyses	148
7.2.3	Extrapolation verticale	151
7.3	Comparaison statistique aux données de la couleur de l'eau	155
7.3.1	Rangs des observations MERIS	155
7.3.2	Score probabiliste CRPS	159
7.3.2.1	Variabilité temporelle	160
7.3.2.2	Décomposition du score	161
7.4	Correction des variables non observées	164
7.4.1	Dispersion des séries temporelles	164
7.4.2	Dispersion des profils verticaux	166
7.5	Conclusion	168

Dans les précédents chapitres, nous avons présenté les caractéristiques d'une simulation d'ensemble de 60 membres réalisée à partir de paramétrisations stochastiques. Plutôt que de s'intéresser directement à l'assimilation de données, nous avons choisi d'approfondir l'étude de cette simulation. Une démarche statistique a pour cela été suivie, attestant qu'une simulation d'ensemble devait être définie en tenant compte des distributions de probabilité caractérisant les différentes sources d'incertitude. L'objectif sous-jacent était de montrer qu'une approche probabiliste, telle que celle affichée ici, propose une alternative crédible à la modélisation biogéochimique déterministe. Dans cet exercice, nous sommes toutefois toujours restés reliés avec le contexte initial d'assimilation de données de la couleur de l'eau. Dans ce dernier chapitre, nous présenterons des expériences d'assimilation de données réalisées à partir des distributions de probabilité définies par la simulation d'ensemble.

Les résultats exposés ici s'appuient uniquement sur l'étude des impacts de la correction de l'état biogéochimique du modèle par les observations de la couleur de l'eau (étape d'analyse). Ils ne proviennent donc pas à proprement parlé d'expériences d'assimilation de données. Après correction, les états actualisés ne seront pas propagés par le modèle. Les raisons pour lesquelles nous limitons ici l'assimilation de données à l'étape d'analyse sont principalement d'ordre technique. Tout d'abord, l'implémentation du schéma d'assimilation d'ensemble dans un système de la taille du modèle NATL025/PISCES est loin d'être triviale. Ensuite, le coût informatique de l'assimilation de données est très élevé. Même si l'on restreint la période d'étude à 6 mois, le temps de calcul requis serait d'environ 30 000 heures. Disposer de ces heures de calcul nécessite donc de s'être assuré au préalable des performances du système. Le fait de présenter ici uniquement des analyses n'est pas un choix par défaut. C'est une étape obligatoire permettant de justifier la réalisation d'une expérience complète d'assimilation de données sur une période relativement longue. Tous les calculs nécessaires aux analyses ont été effectués en utilisant le logiciel *SESAM*, développé au sein de l'équipe MEOM. Une description détaillée de cet outil peut être trouvée à l'adresse <http://lgge.osug.fr/meom/Utils/SESAM/sesam.html>.

Toutes les analyses présentées dans ce chapitre ont été réalisées à partir du schéma d'analyse du Filtre d'Ensemble de Kalman Transformé (ETKF). Les observations utilisées sont les données satellites au $1/24^\circ$ SeaWIFS-HR (cf. chapitre 3). Exploiter ces observations permettra également d'illustrer la capacité du système à intégrer de l'information de haute résolution spatiale dans un modèle de plus basse résolution (ici $1/4^\circ$). Même si cette question restera examinée ici de manière relativement sommaire, elle devra cependant être prochainement approfondie pour répondre aux besoins des futures missions spatiales.

Les principaux points qui seront abordés dans ce chapitre sont :

1. Montrer qu'une méthode d'analyse d'ensemble a été implémentée pour la biogéochimie dans la configuration NEMO/PISCES.
2. S'assurer que les distributions de probabilités issues des paramétrisations stochastiques permettent d'assurer une bonne performance des analyses.
3. Évaluer les caractéristiques statistiques des distributions en chlorophylle de l'ensemble analysé.
4. Vérifier les impacts sur les variables non observées.

Dans la première section de ce chapitre, nous présentons la démarche utilisée pour corriger les distributions de probabilité de la simulation d'ensemble *stoUpd* à partir des observations de la couleur de l'eau. On détaillera la méthode d'analyse du filtre ETKF ainsi que les techniques employées pour prendre en compte des erreurs d'observation et la non-gaussianité des distributions a priori. La deuxième section décrit les résultats d'une succession d'analyses réalisées tous les 5 jours de l'année 2005. Nous examinerons en particulier l'impact des corrections sur les distributions horizontales et verticales de la chlorophylle. Dans la troisième section, nous présenterons une évaluation statistique de l'ensemble analysé comparable à celle du chapitre précédent. Outre des histogrammes de rangs, le « Continuous Rank Probability Score » (CRPS) sera utilisé pour quantifier l'influence des analyses sur la distribution statistique des membres de l'ensemble. Dans la dernière partie de ce chapitre, on s'intéressera à l'impact des analyses sur les distributions des variables non observées.

7.1 Méthode d'analyse d'ensemble

Les expériences d'assimilation réalisées dans cette thèse ont été effectuées à partir du schéma d'analyse du filtre de Kalman d'ensemble transformé (ETKF). Cette méthode est une adaptation du filtre de Kalman pour les simulations d'ensemble. Cette première section décrit précisément le schéma d'analyse et les paramètres de la séquence d'analyses présentée dans les sections suivantes.

7.1.1 Le filtre de Kalman

La méthode du filtre de Kalman a été développée en 1960 par R.E Kalman (Kalman, 1960). On peut le voir comme une extension de l'interpolation statistique, décrite dans le chapitre 2, appliquée à un système dynamique. Des notations similaires sont donc utilisées ici. En tant que méthode séquentielle, le filtre de Kalman se construit sous forme d'une succession d'analyses.

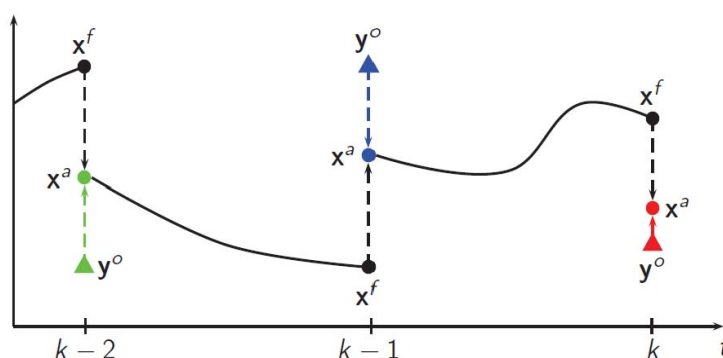


FIGURE 7.1 – Principe simplifié de l'assimilation séquentielle de données.

Un état préalablement corrigé (analysé) $\mathbf{x}_{t_{k-1}}^a$ est propagé par le modèle pour établir une nouvelle prévision $\mathbf{x}_{t_k}^f$, qui est à son tour corrigée pour donner $\mathbf{x}_{t_k}^a$ (cf. figure 7.1). Comme précédemment, on se place dans la situation où le modèle \mathcal{M} donne une prévision $\mathbf{x}_{t_k}^f$ de l'état d'un système observé par $\mathbf{y}_{t_k}^o$. Les erreurs de prévision sont caractérisées par la matrice de covariance $\mathbf{P}_{t_k}^f$ et les erreurs d'observation par la matrice de covariance \mathbf{R}_{t_k} . L'étape d'analyse est une régression linéaire identique à celle de l'interpolation statistique. En ajoutant l'indice temporel, cette étape est rappelée par les équations 7.2 et 7.1.

$$\mathbf{K}_{t_k} = \mathbf{P}_{t_k}^f \mathbf{H}_{t_k}^T (\mathbf{R}_{t_k} + \mathbf{H}_{t_k} \mathbf{P}_{t_k}^f \mathbf{H}_{t_k}^T)^{-1} \quad (7.1)$$

$$\mathbf{x}_{t_k}^a = \mathbf{x}_{t_k}^f + \mathbf{K}_{t_k} (\mathbf{y}_{t_k}^o - \mathbf{H}_{t_k} \mathbf{x}_{t_k}^f) \quad (7.2)$$

Pour l'étape de prévision, l'état analysé $\mathbf{x}_{t_k}^a$ devient l'ébauche d'une prévision $\mathbf{x}_{t_{k+1}}^f$ à laquelle sera associée la matrice de covariance d'erreur $\mathbf{P}_{t_{k+1}}^f$. Le vecteur $\mathbf{x}_{t_{k+1}}^f$ et la matrice $\mathbf{P}_{t_{k+1}}^f$ sont obtenus à partir de $\mathbf{x}_{t_k}^a$ et $\mathbf{P}_{t_k}^a$ (cf. équation 2.8) en les propageant dans le temps par la dynamique du modèle \mathcal{M} . L'information apportée par les observations s'accumulent ainsi au cours du temps dans les équations du modèle. Comme préalablement mentionné, l'étape de propagation n'a pas été mise en place dans cette thèse. L'objectif de cette première section était de rappeler toutefois les expressions énoncées en section 2.4, utiles à la compréhension de l'étape d'analyse du filtre ETKF.

7.1.2 Le filtre de Kalman transformé (ETKF)

Le Filtre de Kalman d'Ensemble Transformé a été introduit par Bishop et al. (2001). Pour ce qui est de la méthode d'analyse (décrite ci-après), l'algorithme est exactement équivalent à celui développé pour le filtre SEEK (évoqué en section 2.4.3), implémenté dans le logiciel SESAM utilisé ici. Outre le fait que c'est un filtre déterministe (les observations ne sont pas perturbées pendant l'étape d'analyse), la principale différence avec un filtre de Kalman d'ensemble traditionnel comme l'EnKF, réside dans l'espace de résolution. L'étape d'analyse de l'ETKF fait, en effet, intervenir une matrice de transformation (notée \mathbf{T}) afin de réaliser les corrections dans l'espace réduit des anomalies de l'ensemble $\delta\mathbf{x}_i^f$.

Au pas de temps t_k , on dispose d'un ensemble de vecteurs d'état $\{\mathbf{x}_i^f : i = 1, 2, \dots, N\}$ représentant les $N = 60$ prévisions du modèle. L'analyse se base sur l'estimé moyen $\overline{\mathbf{x}}^f$ (équation 7.3) à partir duquel on calcule pour chaque membre i une anomalie $\delta\mathbf{x}_i^f$ correspondant à l'écart à la moyenne (équation 7.4).

$$\overline{\mathbf{x}}^f = \frac{1}{N} \sum_{i=1}^N \mathbf{x}_i^f \quad (7.3)$$

$$\delta\mathbf{x}_i^f = \mathbf{x}_i^f - \overline{\mathbf{x}}^f \quad (7.4)$$

Les anomalies caractérisent une matrice \mathbf{S}^f dont la $(i)^{\text{ème}}$ colonne est $\mathbf{S}_{(i)}^f = (1/\sqrt{N-1})\delta\mathbf{x}_i^f$. Cette matrice permet de construire la matrice de covariance \mathbf{P}^f (équation 7.5). La distribution des erreurs est donc soumise à l'hypothèse gaussienne : les incertitudes sont définies par leur moyenne et leur covariance.

$$\mathbf{P}^f = \mathbf{S}^f \mathbf{S}^{fT} \quad (7.5)$$

Le principe de l'analyse de l'ETKF est de calculer le gain de Kalman dans l'espace réduit des anomalies, en écrivant la matrice de covariance d'erreur d'analyse sous la forme d'un produit de matrices d'anomalies. Le changement d'espace se fait par le calcul d'une matrice de transformation \mathbf{T} satisfaisant l'équation 7.6.

$$\mathbf{P}^a = \mathbf{S}^a \mathbf{S}^{aT} = \mathbf{S}^f \mathbf{T} \mathbf{T}^T \mathbf{S}^{fT} \quad (7.6)$$

Comme pour l'analyse par filtre de Kalman, le calcul de la matrice de covariance d'analyse \mathbf{P}^a est fait à partir du calcul du gain de Kalman \mathbf{K} (équations 7.7 et 7.8).

$$\mathbf{K} = \mathbf{P}^f \mathbf{H} (\mathbf{R} + \mathbf{H} \mathbf{P}^f \mathbf{H}^T)^{-1} \quad (7.7)$$

$$\mathbf{P}^a = (\mathbf{I} - \mathbf{K} \mathbf{H}) \mathbf{P}^f \quad (7.8)$$

L'opérateur d'observation \mathbf{H} permet de passer de l'espace du modèle à l'espace des observations. L'estimé moyen de l'ensemble dans l'espace des observations est donc $\overline{\mathbf{H}\mathbf{x}^f}$ auquel on associe la matrice des anomalies $\mathbf{Y}^f = \mathbf{H}\mathbf{S}^f$. Dans cet algorithme, pour des raisons d'efficacité numérique, la matrice d'erreur d'observation \mathbf{R} est supposée diagonale. Nous montrerons dans le paragraphe suivant qu'une méthode d'inflation des variances est utilisée pour limiter l'impact de cette approximation. Si on remplace 7.5 dans 7.7, nous pouvons obtenir une nouvelle expression pour la matrice \mathbf{P}^a :

$$\mathbf{P}^a = \mathbf{S}^f \mathbf{S}^{fT} \left(\mathbf{I} - \mathbf{S}^f \mathbf{Y}^{fT} (\mathbf{Y}^f \mathbf{Y}^{fT} + \mathbf{R})^{-1} \mathbf{H} \right) \quad (7.9)$$

en utilisant les égalités 7.10 et 7.11.

$$\mathbf{Y}^{fT} (\mathbf{Y}^f \mathbf{Y}^{fT} + \mathbf{R})^{-1} = \mathbf{I} + (\mathbf{Y}^{fT} \mathbf{R}^{-1} \mathbf{Y}^f)^{-1} \mathbf{Y}^T \mathbf{R}^{-1} \quad (7.10)$$

$$\mathbf{I} - (\mathbf{I} + \mathbf{A})^{-1} \mathbf{A} = (\mathbf{I} - \mathbf{A})^{-1} \quad \text{avec} \quad \mathbf{A} = \mathbf{Y}^{fT} \mathbf{R}^{-1} \mathbf{Y}^f \quad (7.11)$$

Afin de répondre à l'équation 7.6, la matrice \mathbf{P}^a peut s'exprimer par l'équation 7.12.

$$\mathbf{P}^a = \mathbf{S}^a \mathbf{S}^{aT} = \mathbf{S}^f \left(\mathbf{I} + \mathbf{Y}^f \mathbf{R}^{-1} \mathbf{Y}^{fT} \right)^{-1} \mathbf{S}^{fT} \quad (7.12)$$

On définit ensuite la matrice $\mathbf{\Gamma} = \mathbf{Y}^f \mathbf{R}^{-1} \mathbf{Y}^{fT}$. En faisant une décomposition en éléments propres, cette matrice peut être réécrite via l'équation 7.13 à partir de la matrice unitaire \mathbf{U} et de la matrice diagonale $\mathbf{\Lambda}$, constituées par respectivement les vecteurs propres et les valeurs propres de $\mathbf{\Gamma}$.

$$\mathbf{\Gamma} = \mathbf{U} \mathbf{\Lambda} \mathbf{U}^T \quad (7.13)$$

Cette expression (équation 7.13) définit la matrice de transformation \mathbf{T} qui permet de calculer les anomalies (équation 7.14) et la moyenne d'ensemble $\bar{\mathbf{x}}^a$ (équation 7.15).

$$\delta \mathbf{x}_i^a = \mathbf{S}^f \mathbf{T} \quad \text{avec} \quad \mathbf{T} = \sqrt{N-1} \mathbf{U} (\mathbf{I} + \mathbf{\Lambda})^{-1/2} \mathbf{\Lambda}^{1/2} \mathbf{U}^T \quad (7.14)$$

$$\bar{\mathbf{x}}^a = \bar{\mathbf{x}}^f + \mathbf{S}^f \mathbf{U} (\mathbf{I} + \mathbf{\Lambda})^{-1} \mathbf{Y}^{fT} \mathbf{R}^{-1} (\mathbf{y}_0 - \mathbf{H} \bar{\mathbf{x}}^f) \quad (7.15)$$

L'étape d'analyse consiste ensuite à actualiser les i membres de l'ensemble :

$$\mathbf{x}_i^a = \bar{\mathbf{x}}^a + \delta \mathbf{x}_i^a \quad (7.16)$$

Lorsque la dimension s de l'espace des observations est de grande taille comparé à la taille N de l'ensemble ($s \gg N$), le coût numérique est largement réduit grâce à cette méthode. Le filtre de Kalman transformé ne calcule en effet pas explicitement les covariances d'erreurs dans l'espace des observations (de taille $s \times s$) mais des matrices d'anomalies de taille ($s \times N$). À noter que cette méthode nécessite d'inverser la matrice d'erreur d'observation \mathbf{R} pour le calcul de la matrice $\mathbf{\Lambda}$. Elle ne reste donc efficace que sous l'hypothèse d'une matrice diagonale.

Traitement des erreurs d'observation

Les observations utilisées pour les analyses sont les données SeaWIFS-HR au $1/24^\circ$ (cf chapitre 3). En ce qui concerne les erreurs associées à ces observations, on considérera qu'elles suivent une loi gaussienne, qu'elles sont non biaisées (espérance égale à la vérité) et que la matrice de covariance est diagonale : $\mathbf{R} = \sigma^2 \mathbf{I}$ avec $\sigma = 0.3 \times CHL$ (écart-type de 30%, Gregg and Casey, 2004). À noter que supposer une matrice diagonale signifie que les erreurs sont spatialement non corrélées. Si cette hypothèse est nécessaire à la méthode d'analyse (cf paragraphe précédent), les erreurs de mesure entre 2 points d'observation proches ne sont en réalité pas entièrement déconnectées.

Afin de tenir compte de corrélations d'erreur en conservant une matrice diagonale, on procède à une inflation des termes de la matrice de covariance des erreurs d'observation. En pratique, cela revient à considérer que l'incertitude des observations est augmentée en fonction de la distance de corrélation spatiale. Parfaire ces corrélations n'étant pas l'objectif de cette étude, on se contentera d'utiliser une longueur de corrélation constante de 2.3° . Cette valeur a été déterminée expérimentalement en vue de cette séquence d'analyses.

En résumé, on a donc $\mathbf{R} = \alpha(\sigma^2 \mathbf{I})$ avec $\sigma = 0.3 \times CHL$. La constante $\alpha = 2.3 \times 24$ correspond à la racine carrée du nombre de données d'observations contenu dans un rayon de 2.3° autour du point de grille considéré. À noter que lorsque la densité spatiale des observations est très variable, l'utilisation de cette technique, relativement approximative, peut poser problème. Cette méthode empirique reste en effet plutôt destinée à des observations possédant une couverture spatiale relativement régulière, comme c'est globalement le cas des données SeaWIFS-HR.

Traitement des incertitudes du modèle

Si l'approche stochastique permet de caractériser les incertitudes du modèle, définir les covariances d'erreur à partir de 60 membres reste une approximation. À partir d'un nombre réduit de membres, des effets d'échantillonnage de la matrice de covariance peuvent être à l'origine de corrélations spatiales de grande distance non réalistes. Ce phénomène ne doit donc pas être négligé dans un domaine de la taille de l'Atlantique Nord, pour lequel la dimension horizontale est largement supérieure aux longueurs de corrélations. Pour réduire les coûts de calcul et gommer l'impact d'une définition imparfaite des covariances spatiales de grande distance, l'étendue spatiale des corrections est réduite en introduisant une matrice de localisation. La version utilisée ici est toutefois spécifique à la méthode d'analyse du filtre SEEK ou de l'ETKF (renommé LETKF). Une description détaillée peut être trouvée dans Brankart et al. (2009). De manière simplifiée, cette technique est basée sur l'utilisation d'une fonction de corrélation $\gamma(r) = \exp(-r^2/l^2)$, qui atténue de manière exponentielle les corrélations spatiales en fonction de la distance r . De plus, au delà d'une longueur l_c (rayon de coupure), les covariances sont imposées nulles. Pour cette étude, les valeurs $l = 4$ et $l_c = 16$ points de grille, reprises des travaux de Candille et al. (2015), ont été utilisées.

7.1.3 Prise en compte des non-gaussianités

Dans le schéma d'analyse de l'ETKF, l'hypothèse de gaussianité de la prévision constitue une approximation forte comparée aux distributions de probabilité effectivement simulées (cf. figure 7.2). Dans la majorité des cas, l'allure des distributions de la chlorophylle est plus proche d'une distribution log-normale que d'une distribution gaussienne. Pour se rapprocher d'une distribution gaussienne, certaines études ont recours à une transformation log-normale. Cette transformation étant spécifique, elle reste généralement insuffisante (p.ex. Ciavatta et al., 2011). Afin de rendre compte de la non-gaussianité des distributions de manière plus générique, on appliquera plutôt une transformation anamorphique qui dépend de la prévision d'ensemble elle-même. Cette méthode a préalablement été utilisée dans le contexte d'assimilation de données biogéochimiques dans le cas d'expériences jumelles (p.ex. Simon and Bertino, 2009; Doron et al., 2011) comme avec des données réelles de la couleur de l'eau (p.ex. Doron et al., 2013; Fontana et al., 2013).

La méthode d'anamorphose utilisée ici est issue des travaux de Béal et al. (2010). Une description mathématique complète peut être trouvée dans Brankart et al. (2012). Le principe est d'effectuer une transformation¹ bijective f qui redessine les centiles de la distribution de probabilité a priori afin d'obtenir une distribution gaussienne. On suppose par exemple en un point de grille la variable d'état x . Les centiles x_k de la distribution de probabilité de cette variable x sont définis par l'ensemble de sorte que $p(x < x_k) = r_k$. Dans notre cas, on utilise $r_k = 0, 0.02, 0.04, 0.06, 0.08, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 0.92, 0.94, 0.96, 0.98, 1$. La transformation anamorphique f est alors définie par le système d'équation 7.17 ci-après :

1. Anamorphose est un mot d'origine grecque qui signifie transformer. Les premiers exemples d'anamorphoses sont issus des travaux de perspective *De prospecta pingendi*, menés par le peintre Piero della Francesca entre 1460-1480.

$$f(x) = \begin{cases} z_1 & \text{pour } x < x_1 \\ z_k + \frac{z_{k+1}-z_k}{x_{k+1}-x_k}(x-x_k) & \text{pour } x_k \leq x \leq x_{k+1} \\ z_p & \text{pour } x > x_p \end{cases} \quad (7.17)$$

où les z_k sont les centiles correspondant aux r_k d'une distribution gaussienne $\mathcal{N}(0, 1)$. La transformation f est univoque sur l'intervalle $[x_1, x_p]$. La fonction réciproque g permettant de repasser dans l'espace original (espace des états) est donc uniquement définie dans l'intervalle $[z_1, z_p]$ (équation 7.18). Nous verrons dans la suite que cette caractéristique, propre à cette méthode précise d'anamorphose, aura des conséquences sur les analyses.

$$g(z) = \begin{cases} x_1 & \text{pour } z < z_1 \\ x_k + \frac{x_{k+1}-x_k}{z_{k+1}-z_k}(z-z_k) & \text{pour } z_k \leq z \leq z_{k+1} \\ x_p & \text{pour } z > z_p \end{cases} \quad (7.18)$$

La figure 7.2 illustre l'impact de l'anamorphose sur la distribution de l'ensemble a priori. On constate que l'anamorphose permet de se rapprocher significativement d'une distribution gaussienne des erreurs. L'avantage de cette méthode est qu'elle s'adapte aux variations des distributions de probabilité. C'est une transformation locale dont l'expression va ainsi varier spatialement en fonction de l'allure des distributions définies par l'ensemble. Les matrices d'anomalies utilisées pour les analyses seront calculées dans cet espace anamorphosé.

L'hypothèse de gaussianité s'appliquant aussi aux observations, la même transformation leur sera appliquée. Les observations sont ainsi calculées dans l'espace anamorphosé ($\mathbf{y}^0 \rightarrow \mathbf{y}_a^0$). Pour définir la valeur de l'écart-type des erreurs d'observation dans ce nouvel espace, on projette également les 2 valeurs $\beta^- = \mathbf{y}^0 - \alpha\sigma$ et $\beta^+ = \mathbf{y}^0 + \alpha\sigma$, qui représentent les pentes de la distribution gaussienne des observations dans l'espace anamorphosé ($\beta^{+;-} \rightarrow \beta_a^{+;-}$). Connaissant la valeur de α , choisie par l'utilisateur (en général petite), on peut évaluer l'écart-type dans l'espace anamorphosé à partir du calcul de $(\beta^+ - \beta^-)/2\alpha$. Je précise que l'anamorphose transforme les distributions marginales en distributions gaussiennes mais elle ne garantit pas la gaussianité des distributions conjointes (entre variables).

Bilan des paramètres de l'analyse

- Les observations de la couleur de l'eau utilisées sont les données satellites SeaWiFS-HR au 1/24° décrites dans le chapitre 3.
- Les corrections sont réalisées selon le schéma d'analyse du Filtre d'Ensemble de Kalman Transformé (ETKF), équivalent à celui du filtre SEEK.
- Une localisation de 2.3° est intégrée au schéma d'analyse ($ETKF \rightarrow LETKF$).
- Une anamorphose est appliquée aux distributions pour réduire les effets des non-gaussianités.
- L'analyse est multivariée. En plus de la chlorophylle, 8 variables sont impactées par les analyses. Elles seront présentées en section 7.4.
- L'analyse est réalisée tous les 5 jours pendant l'année 2005.

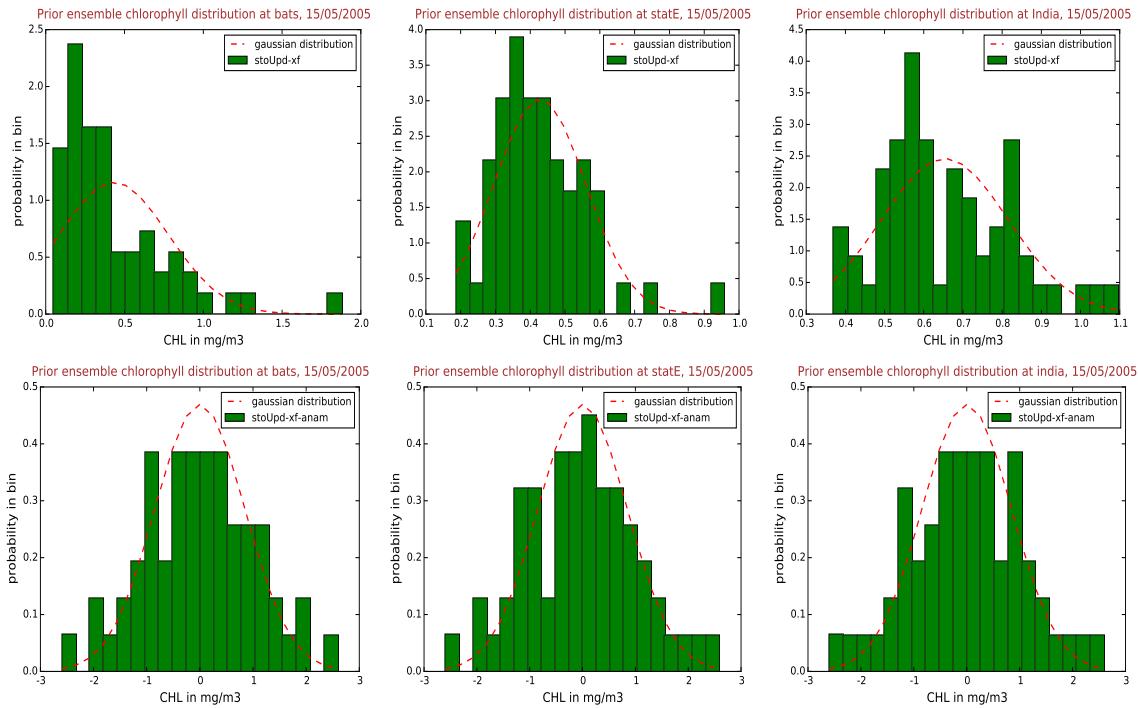


FIGURE 7.2 – Distributions de probabilité de l'ensemble a priori (1^{ère} ligne) aux points de grille BATS, STAT_E et INDIA (cf. figure 6.5) pour le 15 mai 2005. Les histogrammes de la 2^{ème} ligne indiquent ces mêmes distributions de probabilité après anamorphose. La comparaison avec une distribution gaussienne est explicitée par l'intermédiaire des courbes en pointillés rouges.

7.2 Un nouvel ensemble d'états actualisés

À partir des paramètres détaillés dans la section précédente, la simulation d'ensemble de 60 membres décrite dans les 2 précédents chapitres (renommée *stoUpd-xf*) est corrigée par les observations de la couleur de l'eau SeaWIFS-HR. Tous les 5 jours, il en résulte un nouvel ensemble de 60 membres d'états dits actualisés (noté *stoUpd-xa*). J'insiste sur le caractère tridimensionnel de cette succession d'analyses. Entre 2 analyses, l'ensemble corrigé n'est pas propagé par le modèle. Les résultats présentés dans ce chapitre correspondent donc à une séquence d'analyses 3D des données de la couleur de l'eau sur l'année 2005. Cette séquence peut être perçue comme une expérience d'assimilation de données qui ne garderait pas la mémoire des données intégrées dans les modèles durant les cycles précédents.

Dans cette section, les résultats de cette séquence d'analyses sont présentés en termes de concentration en chlorophylle. On montre en particulier que l'estimation moyenne est améliorée et que l'incertitude des prévisions est réduite. La seconde partie de cette section développe l'impact des analyses sur la structure verticale. En résumé, l'objectif de cette section est de montrer que la méthode d'analyse d'ensemble a été implémentée avec succès selon des paramètres permettant d'intégrer les observations satellites de haute résolution en respectant la dynamique du modèle.

7.2.1 Impact des analyses sur l'état moyen

En incorporant les données des capteurs MERIS et MODIS pour combler une partie des zones non observées (cf. chapitre 3), les images SeaWIFS-HR proposent une information de surface particulièrement bien distribuée spatio-temporellement (voir figure 7.3). L'utilisation de ces données de haute résolution a pu mettre en évidence la nécessité de tenir compte de corrélations spatiales des erreurs d'observation (voir section 7.1.2). Sans la prise en compte de ces corrélations, l'information du modèle est en quelque sorte "noyée" par la redondance des observations. Dans un premier temps, les résultats des analyses sont présentés en figure 7.3 par des cartes de surface de la moyenne d'ensemble de la chlorophylle. Pour mieux visualiser l'impact des analyses, la moyenne de l'ensemble a priori et les observations SeaWIFS-HR sont également représentées.

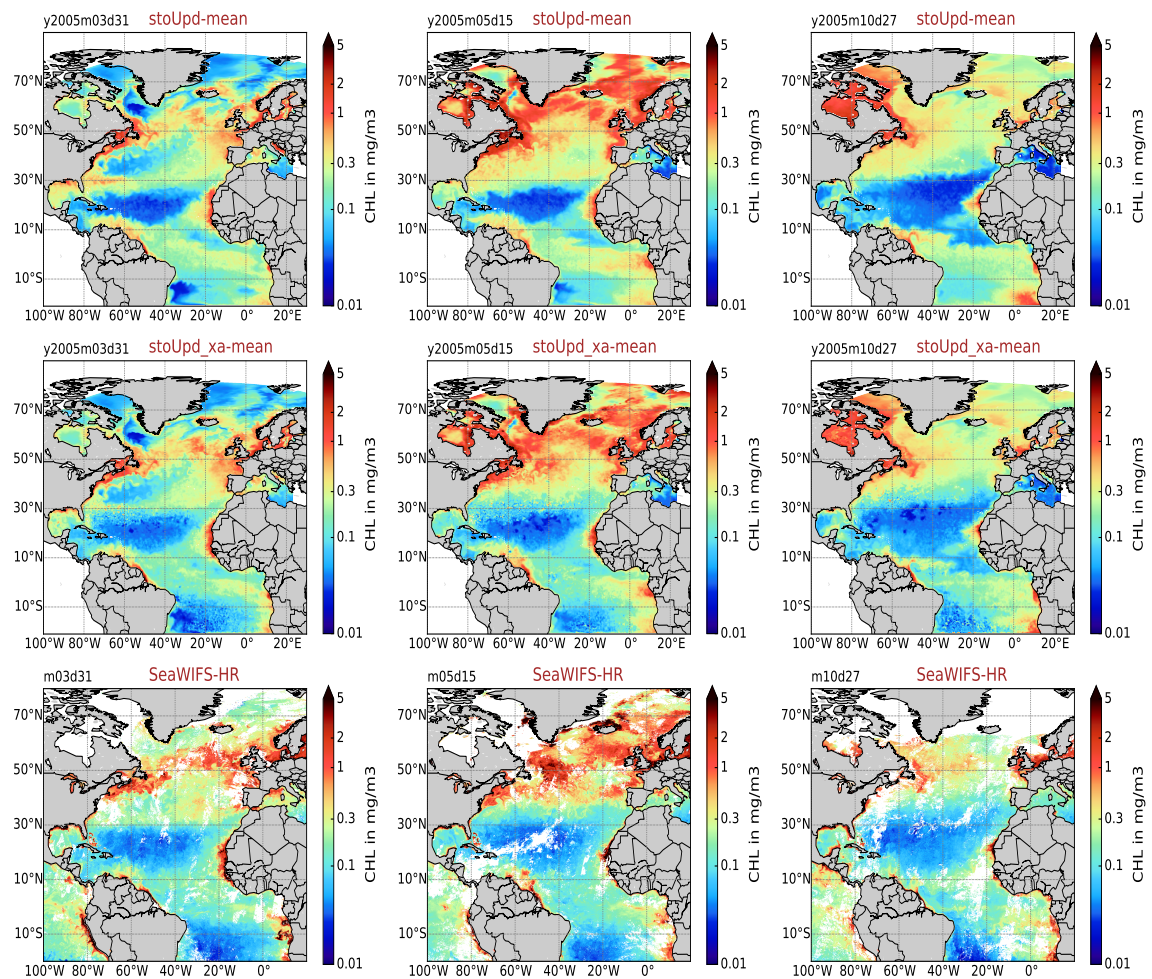


FIGURE 7.3 – Cartes de chlorophylle de surface de la moyenne de l'ensemble a priori (1^{ère} ligne), de la moyenne de l'ensemble après analyse (2^{ème} ligne) et des observations SeaWIFS-HR utilisées pour l'analyse (3^{ème} ligne). Ces résultats sont présentés pour le 31 mars (1^{ère} colonne), le 15 mai (2^{ème} colonne) et le 27 octobre (3^{ème} colonne) de l'année 2005.

Cette figure montre que l'état moyen de l'ensemble actualisé se rapproche significativement des observations de la couleur de l'eau utilisées pour l'analyse. Par rapport à la moyenne de l'ensemble a priori, les concentrations en chlorophylle sont augmentées au niveau du gyre subtropical et diminuées dans la région du Gulf Stream ainsi que dans la partie nord du domaine. Les forts gradients nord-sud (évoqués dans les chapitres précédents) sont par conséquent sensiblement réduits par l'analyse. Une structure spatiale légèrement plus « patchy » peut également être détectée dans les régions des moyennes et hautes latitudes. Les corrections effectuées dans les régions non couvertes par les observations paraissent convaincantes, puisqu'elles n'induisent pas de variations brutales des quantités de chlorophylle. De manière globale, l'analyse ne provoque pas de modification majeure sur la répartition spatiale moyenne de l'état a priori. La structure dynamique initialement définie par le modèle est donc respectée, tout en corrigeant les concentrations de chlorophylle. Ce point est particulièrement important afin de pouvoir conserver les améliorations de l'analyse après une étape de propagation par le modèle.

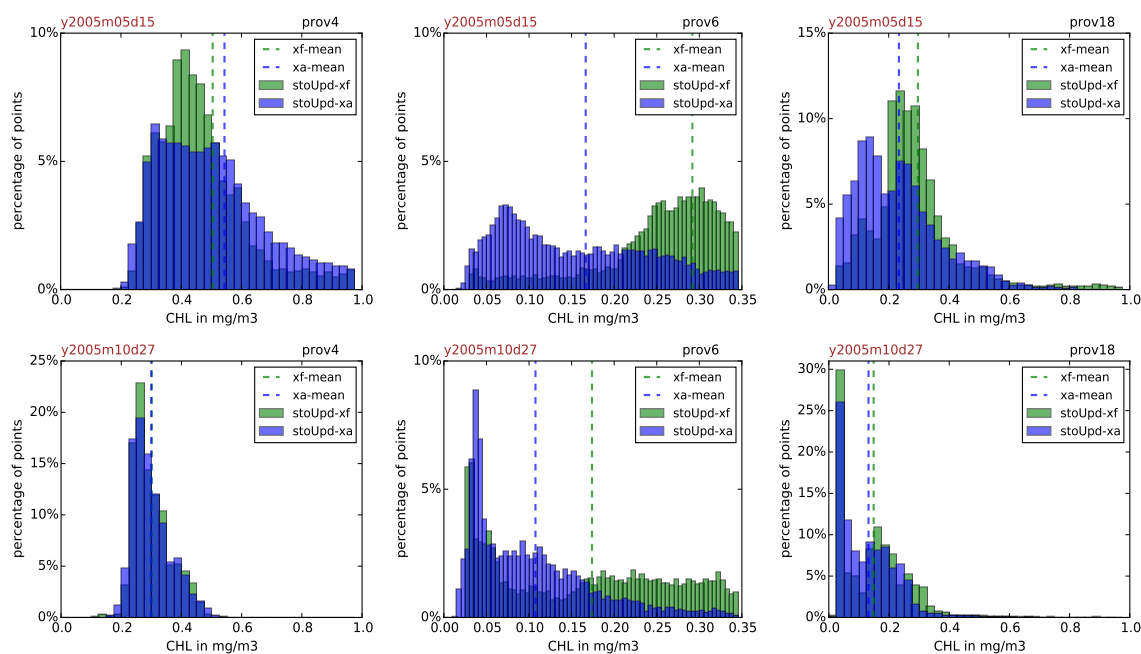


FIGURE 7.4 – Histogrammes des moyennes de l'ensemble a priori (vert) et de l'ensemble analysé (violet) de la chlorophylle de surface à tous les points de grille situées à l'intérieur des provinces de Longhurst 4,6 et 18. La superposition des 2 histogrammes est représentée en bleu. Les lignes en pointillés indiquent la concentration moyenne en chlorophylle de surface des différentes provinces. Ces résultats sont présentés pour le 15 mai (1^{ère} ligne) et le 27 octobre (2^{ème} ligne) de l'année 2005

Si l'ensemble actualisé conserve la structure moyenne du modèle, l'analyse doit aussi pouvoir rendre compte de la variabilité spatiale de petite échelle des observations qui n'est pas correctement simulée par le modèle. Pour cela, la figure 7.4 présente les histogrammes des moyennes d'ensemble de la chlorophylle de surface des simulations *stoUpd-xf* et *stoUpd-xa*, obtenus à partir de tous les points de grille situés dans les provinces de Longhurst 4, 6 et 18 (cf. figure 6.5). Nous pouvons remarquer sur ces histogrammes que même lorsque

l'impact sur la production moyenne globale en chlorophylle des provinces est faible (lorsque les lignes en pointillés bleus et verts sont proches, p.ex. provinces 4 et 18), l'analyse a un impact sur la distribution spatiale des quantités (ici moyennes) de chlorophylle. La variabilité spatiale (hétérogénéité) des observations peut ainsi modifier la distribution spatiale du modèle, sans pour autant affecter la répartition moyenne de la chlorophylle sur la plus grande échelle. Dans la province 6 (région du Gulf Stream), les histogrammes soulignent toutefois une forte diminution de la moyenne de l'ensemble lors de l'analyse. Si cet effet permet un meilleur accord avec les observations satellites, nous verrons cependant que l'ensemble analysé ne possède pas les caractéristiques statistiques attendues dans cette région (cf. section suivante).

Récapitulatif

Excepté dans les régions non couvertes par les observations, les améliorations sur l'état moyen étaient attendues par la construction de la méthode d'analyse (basée sur la correction des anomalies par rapport à l'état moyen). Ces premiers résultats ont donc permis de vérifier la pertinence de la méthode d'analyse et de confirmer la bonne définition a priori des corrélations spatiales sur la chlorophylle. Nous avons également constaté que les corrections ne déstructurent pas la représentation moyenne de l'ensemble a priori.

Néanmoins, puisque l'analyse conserve un degré non nul d'incertitude, considérer uniquement la moyenne de l'ensemble reste insuffisant.

7.2.2 Incertitudes des analyses

Suite à l'analyse, l'effet attendu est une diminution des incertitudes. Les observations étant imparfaites, le niveau d'incertitude résiduel ne doit cependant jamais devenir nul. L'ensemble doit en effet conserver un niveau de dispersion significatif, caractérisant l'incertitude de la nouvelle estimation.

Sur la figure 7.5 (1^{ère} ligne), des cartes de surface affichent les valeurs de l'écart-type relatif en chlorophylle des membres analysés. Comme lors des chapitres précédents, il est calculé à chaque point de grille comme le rapport $\sqrt{(x_i^a - \overline{x_i^a})^2} / \overline{x_i^a}$ entre l'écart-type des membres de l'ensemble et la moyenne d'ensemble. L'indice supérieur a fait référence au terme "analyse".

Cette figure montre que l'ensemble analysé conserve une dispersion importante, d'environ 30%, sur la quasi-totalité du domaine. Dans certaines zones des hautes latitudes et autour de 30°N, une dispersion particulièrement faible peut néanmoins être observée. Une explication sera donnée au paragraphe suivant. De manière générale, la répartition spatiale de la dispersion suit celle de la simulation *stoUpd-xf* (cf. chapitre 5, figure 5.9). Elle reste donc cohérente avec l'incertitude a priori du modèle. Le rapport entre l'écart-type relatif de l'ensemble analysé et celui de l'ensemble a priori est également présenté en figure 7.5 (2^{ème} ligne). Par rapport à la simulation *stoUpd-xf*, une diminution de la dispersion d'environ 10-20% peut ainsi être observée sur la totalité du domaine. L'analyse réduit donc l'incertitude de la prévision, tout en conservant presque partout un niveau d'incertitude cohérent avec son estimation a priori.

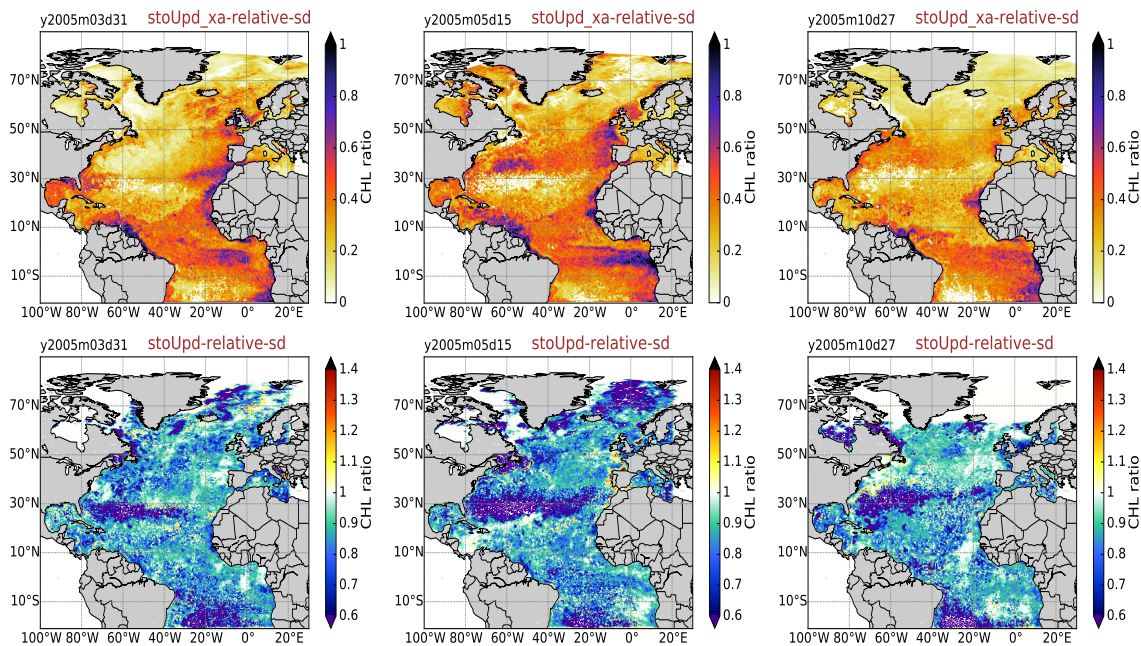


FIGURE 7.5 – Cartes de surface de l'écart-type relatif de la chlorophylle de l'ensemble analysé (1^{ère} ligne) et du rapport avec l'écart-type relatif de l'ensemble a priori (analysé/a priori, 2^{ème} ligne). Ces résultats sont présentés pour le 31 mars (1^{ère} colonne), le 15 mai (2^{ème} colonne) et le 27 octobre (3^{ème} colonne) de l'année 2005.

Afin de caractériser plus précisément la dispersion (sans hypothèse gaussienne), la figure 7.6 présente en quelques points de grille les séries temporelles des membres de l'ensemble a priori (courbes vertes) et de l'ensemble analysé (courbes bleues). À la station BATS, on identifie clairement la très faible dispersion préalablement observée sur les cartes d'écart-type de la figure 7.5. Dans cette situation, même si l'état moyen est amélioré, la dispersion résiduelle est beaucoup trop faible et l'ensemble ne peut absolument pas être fiable. Lors d'une étape future d'assimilation de données, l'estimation de l'erreur modèle serait donc très sous-estimée. L'intégration des observations n'apportera presque plus d'information. L'intérêt de l'approche probabiliste est perdu : l'ensemble s'effondre (ou "collapse").

Cette absence de dispersion provient du traitement des erreurs d'observation par la méthode d'anamorphose décrite en section 7.1.3. Pour les définir, les valeurs de l'observation et de l'écart-type de la distribution, supposée gaussienne, des erreurs (approximé par $y_o \pm \sigma$) sont calculées dans l'espace anamorphosé décrit par les centiles de l'ensemble a priori. Lorsque les observations sont en dehors de l'enveloppe de l'ensemble, l'anamorphose contraint ces valeurs à prendre la valeur de l'extremum de l'ensemble le plus proche (dans l'espace anamorphosé). L'erreur d'observation devient donc nulle dans l'espace anamorphosé. Puisque qu'un grand nombre d'observations est considéré dans le domaine de localisation, une dispersion strictement nulle ne se produit que dans des cas réduits : lorsque les erreurs de toutes les observations situées dans le rayon de localisation sont nulles. Pour une expérience d'assimilation, il serait primordial de remédier à cela en affectant, par exemple, un niveau d'erreur minimum aux observations. Dans cette séquence d'analyses, faire ce constat permet d'insister une nouvelle fois sur l'importance d'inclure les observations dans

l'enveloppe de l'ensemble. D'autres méthodes d'anamorphose (p.ex. Simon and Bertino, 2012) définissent des distributions non limitées par les bornes de l'ensemble a priori. Nous ne souhaitons pas utiliser ce type de méthode afin d'éviter d'utiliser des corrélations d'erreur dans un espace des phases qui ne serait pas exploré par les membres de l'ensemble a priori. Ce point nous paraît important pour obtenir des corrections cohérentes sur la verticale et sur les variables non observées (cf. sections 7.2.3 et 7.4). Puisque la majorité des observations est incluse dans l'enveloppe de l'ensemble, cette particularité reste relativement limitée spatio-temporellement.

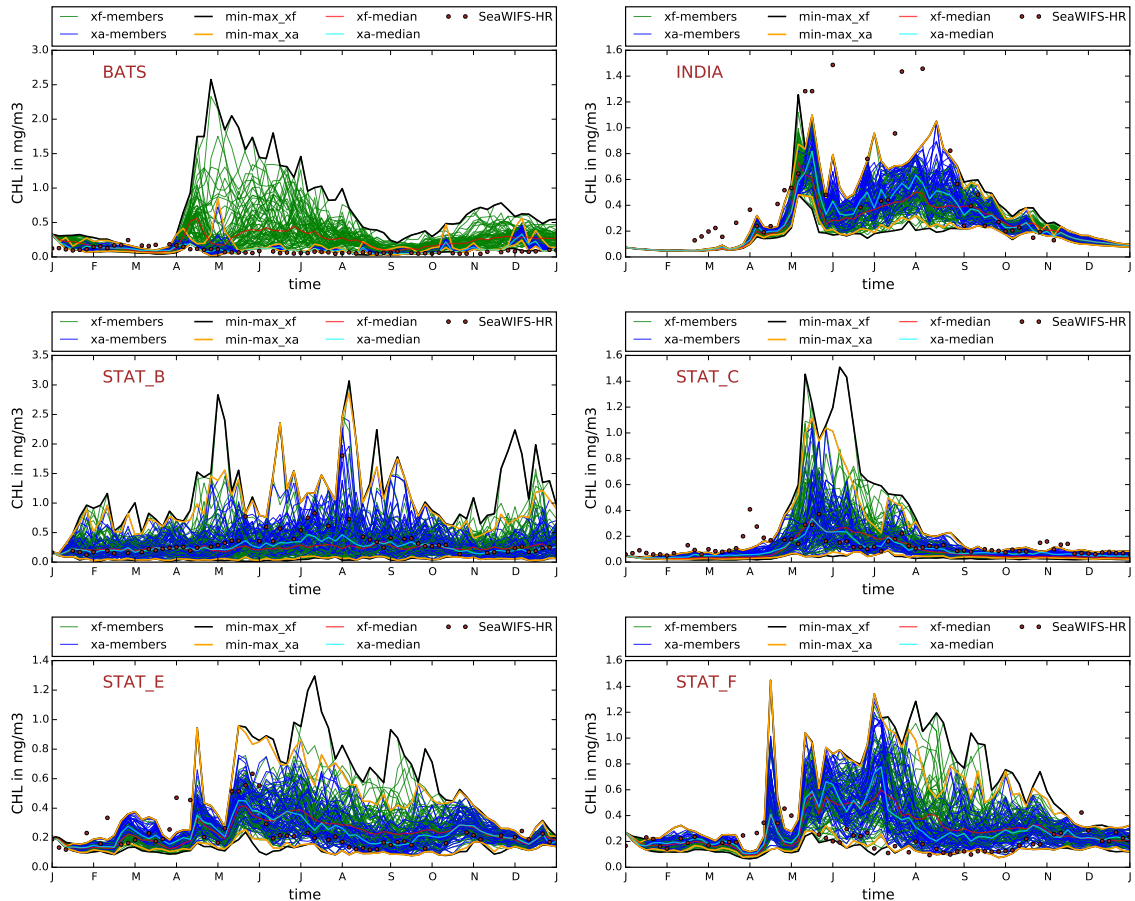


FIGURE 7.6 – Séries temporelles de la chlorophylle de surface simulée par les membres de l'ensemble a priori (vert) et de l'ensemble analysé (bleu) en plusieurs points de grille du domaine (cf. figure 6.5). Les observations SeaWIFS-HR sont indiquées par des cercles rouges. Les médianes (courbes brunes et cyan) et les extrema des ensembles (courbes noires et orange) sont également représentés.

En dehors de cette difficulté qui n'apparaît que très localement, l'analyse permet de réduire significativement le niveau d'incertitude. La médiane de l'ensemble se rapproche toujours des observations et les maxima sont bien diminués dans les régions pour lesquelles le niveau d'incertitude est élevé. L'ensemble a priori ayant régulièrement tendance à surestimer les concentrations, l'analyse impacte surtout les membres qui simulent de fortes quantités de chlorophylle. Grâce à la méthode d'anamorphose utilisée ici, les membres de l'ensemble analysé restent toujours inclus dans l'intervalle de valeur de l'ensemble a priori.

La pertinence des caractéristiques dynamiques de l'ensemble a priori est donc fondamentale. A noter que nous ne commentons pas ici les observations SeaWIFS-HR, qui sont indiquées sur la figure 7.6 de manière informative. Une comparaison statistique avec les observations satellites fera l'objet de la section 7.3.

Bilan

De manière globale, la méthode d'analyse permet d'améliorer la représentation de la chlorophylle de surface tout en diminuant l'incertitude résiduelle de prévisions. L'utilisation d'une approche d'ensemble permet de définir l'espace des solutions en fonction de l'erreur de prévision. L'ensemble analysé reste ainsi toujours cohérent avec le niveau d'incertitude de la simulation a priori. Dans certaines régions localisées, telles qu'aux abords du Gulf Stream, la prise en compte des erreurs d'observation devra être améliorée pour éviter l'effondrement de l'ensemble. Il sera, par exemple, envisageable de définir un niveau minimal d'incertitude à appliquer pour les cas où les observations sont situées en dehors de l'enveloppe de l'ensemble (dans l'espace anamorphosé). Dans la section suivante, nous examinerons l'impact des analyses dans la colonne d'eau.

7.2.3 Extrapolation verticale

L'extrapolation verticale d'observations restreintes à la surface est une difficulté majeure en assimilation de données. L'avantage d'utiliser une approche d'ensemble est de pouvoir définir les corrélations verticales directement à partir de la dynamique du modèle. Cela explique que la cohérence de la structure verticale de la simulation d'ensemble (du moins avec la simulation déterministe) ait fait l'objet d'une attention particulière lors de la mise en place des paramétrisations stochastiques (cf chapitre 5). Dans cette section, on s'intéresse à l'impact des analyses sur les distributions verticales de la chlorophylle.

État moyen

La figure 7.7 présente des coupes verticales à 30°W de la moyenne de l'ensemble a priori (1^{ère} ligne) et de l'ensemble après analyse (2^{ème} ligne) pour le 15 mars et le 27 octobre 2005. Pour faciliter la comparaison, le rapport entre ces 2 structures moyennes est également représenté (3^{ème} ligne).

Cette figure montre que l'analyse a un impact important (entre 10 et 40%) dans l'ensemble de la couche euphotique. En dessous de cette zone, l'analyse n'a pas d'effets. On constate également que l'information apportée par les observations ne modifie pas en moyenne les caractéristiques verticales de profondeur de propagation et de localisation des maxima de subsurface. La structure verticale de l'ensemble analysé conserve donc les principaux éléments de la dynamique verticale du modèle vérifiés dans les chapitres précédents. Une bonne représentation de la structure verticale moyenne a priori apparaît donc essentielle, démontrant de l'importance de la définition initiale des paramétrisations stochastiques. Comme précédemment, l'analyse conserve toutefois un degré d'incertitude qu'il est nécessaire de considérer.

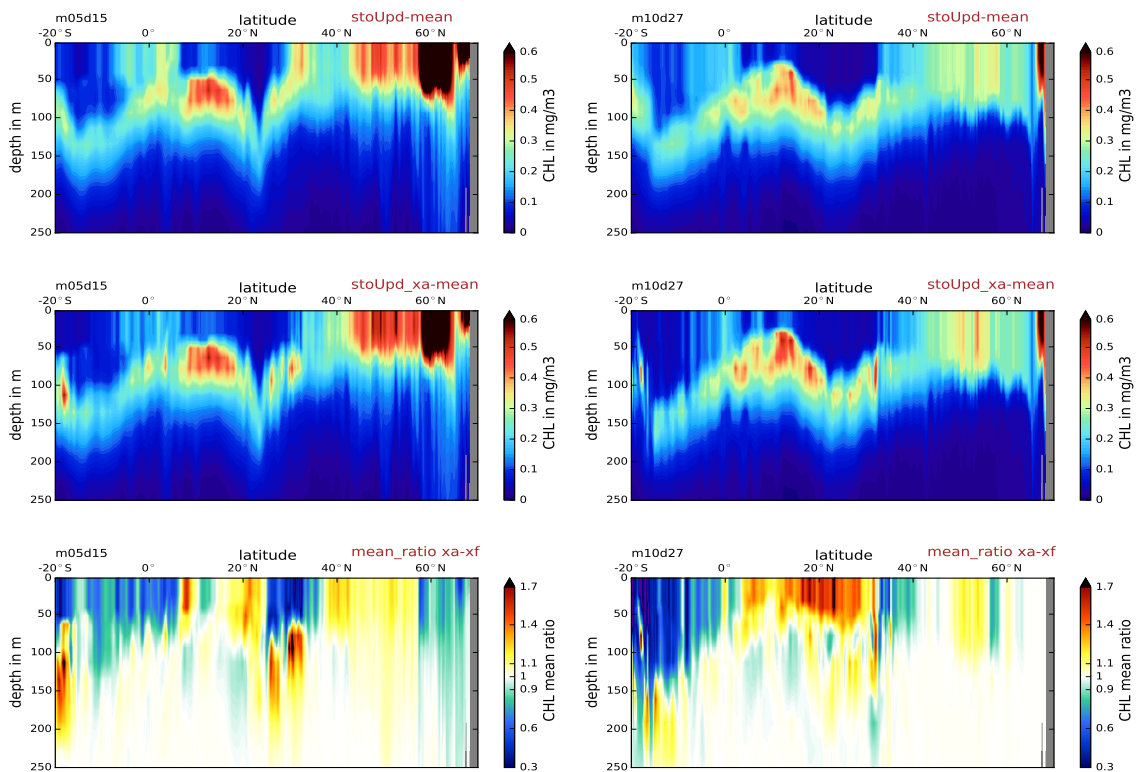


FIGURE 7.7 – Coupes verticales en chlorophylle à 30°W de la moyenne de l'ensemble a priori (1^{ère} ligne), de la moyenne de l'ensemble actualisé par filtre d'ensemble ETKF (2^{ème} ligne) et du rapport entre les 2 (analysé/a priori, 3^{ème} ligne). Ces résultats sont présentés pour le 15 mai (1^{ère} colonne) et le 27 octobre (2^{ème} colonne) de l'année 2005.

Incertitude des analyses

La figure 7.8 présente des coupes verticales à 30°W de l'écart-type relatif de l'ensemble analysé (1^{ère} ligne) ainsi que de son rapport avec l'écart-type relatif de l'ensemble a priori (2^{ème} ligne). Elle montre que l'analyse conserve un degré d'incertitude significatif, compris entre 30% et 50%, dans l'ensemble de la couche euphotique. La dispersion est légèrement plus forte aux profondeurs des maxima de subsurface, zones dans lesquelles l'incertitude du modèle est la plus élevée (cf. section 5.3.2). Comme attendu, l'analyse réduit donc sensiblement la dispersion (entre 10-20%) par rapport à la simulation a priori.

Diversité des profils verticaux

Dans le chapitre précédent, nous avons mis en évidence la capacité des paramétrisations stochastiques à simuler l'existence d'une variété de profils verticaux, pouvant être attribuée à une diversité d'espèces. Pour conserver ce bénéfice, qui résulte de l'approche probabiliste, il est important de vérifier que la diminution de la dispersion, due à l'analyse, ne dégrade pas cette caractéristique. En figure 7.9, nous présentons pour cela les profils verticaux des membres de l'ensemble a priori (courbes vertes) et de l'ensemble analysé (courbes bleues).

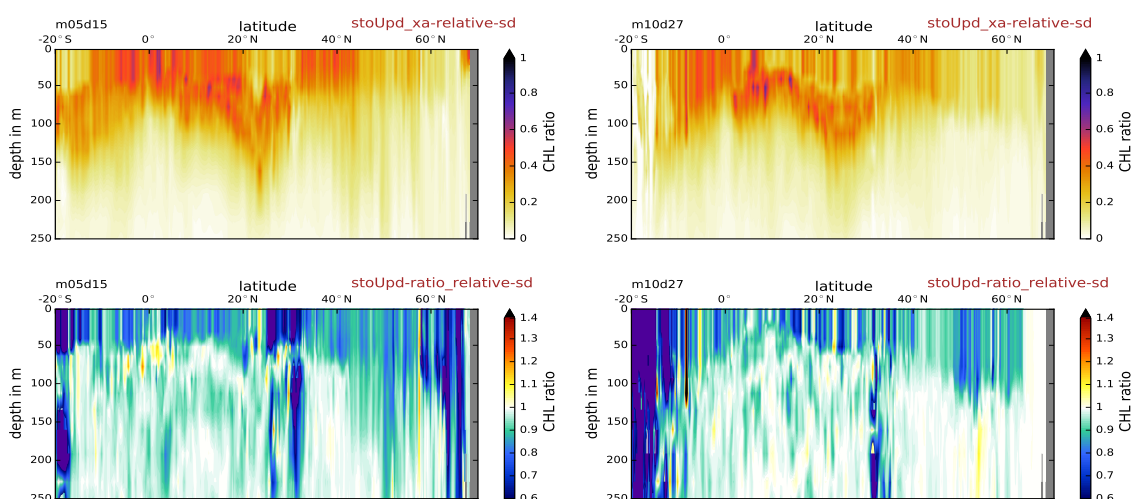


FIGURE 7.8 – Coupes verticales en chlorophylle à 30°W de l'écart-type relatif de l'ensemble actualisé par filtre d'ensemble ETKF (1^{ère} ligne) et du rapport avec l'écart-type relatif de l'ensemble a priori (analysé/a priori, 2^{ème} ligne). Ces résultats sont présentés pour le 15 mai (1^{ère} colonne) et le 27 octobre (2^{ème} colonne) de l'année 2005.

La figure 7.9 (au verso) montre que les profils verticaux de l'ensemble analysé respectent les corrélations verticales définies par l'ensemble a priori. Lorsque la dispersion de surface est (trop) faible comme souvent à la station BATS, les membres tendent ainsi vers un profil vertical unique. Lorsqu'elle est plus importante, les membres analysés reproduisent différents comportements verticaux. L'analyse permet donc de conserver l'existence d'une diversité d'espèce, à l'intérieur des 2 classes de phytoplancton du modèle PISCES, résultant de l'approche probabiliste (cf. section 6.3.2.2). D'autres expériences seraient toutefois nécessaires pour étudier précisément la sensibilité des représentations verticales aux paramètres de l'analyse. Dans toutes les situations, les membres analysés restent toujours inclus dans l'enveloppe de l'ensemble a priori. Ce résultat souligne encore que les performances de l'analyse dépendent principalement de la définition des caractéristiques verticales, issues des paramétrisations stochastiques. Si l'ensemble a priori contient des corrélations verticales satisfaisantes, les améliorations faites par l'analyse sur la chlorophylle de surface doivent permettre d'améliorer les profils verticaux.

Récapitulatif

De manière globale, ces premiers résultats présentant l'extrapolation verticale de corrections réalisées à partir d'observations de la couleur de l'eau, sont plutôt convaincants. L'analyse permet de réduire la dispersion, tout en conservant une diversité de profils verticaux, représentative d'une variété de comportements biologiques, cohérente avec les observations de surface. Une comparaison avec des profils verticaux en chlorophylle serait toutefois nécessaire pour examiner de manière plus approfondie l'impact des analyses.

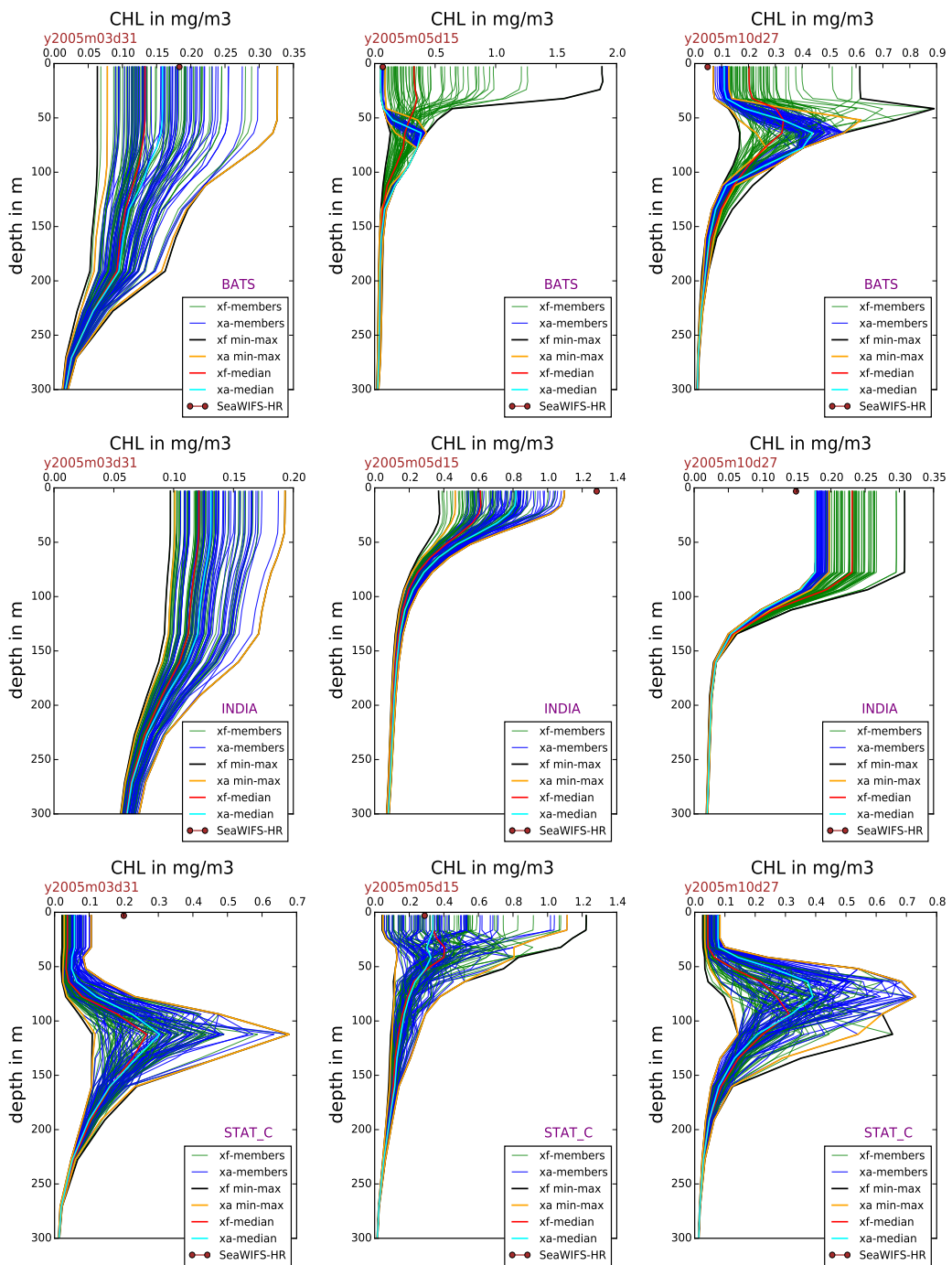


FIGURE 7.9 – Profils verticaux en chlorophylle des membres de l'ensemble a priori (vert) et analysé (bleu) aux points de grille (cf. figure 6.5) correspondant aux stations BATS (1^{ère} ligne), INDIA (2^{ème} ligne) et STAT_C (3^{ème} ligne). Les observations SeaWIFS-HR sont indiquées par des cercles rouges. Les médianes (courbes rouges et bleu cyan) et les extrema (courbes noires et orange) sont également représentés.

7.3 Comparaison statistique aux données de la couleur de l'eau

Dans cette section, les impacts de l'analyse sur l'estimation de la chlorophylle de surface sont évalués dans une comparaison statistique avec les observations. Afin d'utiliser un jeu d'observations (presque) indépendant², on s'appuie sur les données satellites MERIS dont les caractéristiques ont été présentées au chapitre 3. La démarche adoptée dans la première partie de cette section est calquée sur celle du chapitre précédent (section 6.2.2.1). La cohérence statistique entre l'ensemble analysé et les observations sera démontrée par l'intermédiaire d'histogrammes de rangs. Afin d'enrichir l'évaluation statistique de l'ensemble, le CRPS (Continuous Rank Probability Score) sera introduit dans la seconde partie de cette section. Outre le fait d'apporter une vision quantitative, ce diagnostic permettra de mesurer la résolution statistique de l'ensemble.

7.3.1 Rangs des observations MERIS

Dans cette première section, on s'intéresse uniquement aux rangs des observations MERIS, que l'on présente sous forme d'histogrammes. La propriété statistique évaluée est donc la fiabilité, c'est-à-dire la cohérence statistique entre la distribution des membres de l'ensemble et les observations. En guise de rappel, pour qu'un ensemble soit parfaitement fiable, tous les membres de l'ensemble doivent représenter l'état observé avec la fréquence d'occurrence, ce qui se traduit par un histogramme de rangs plat.

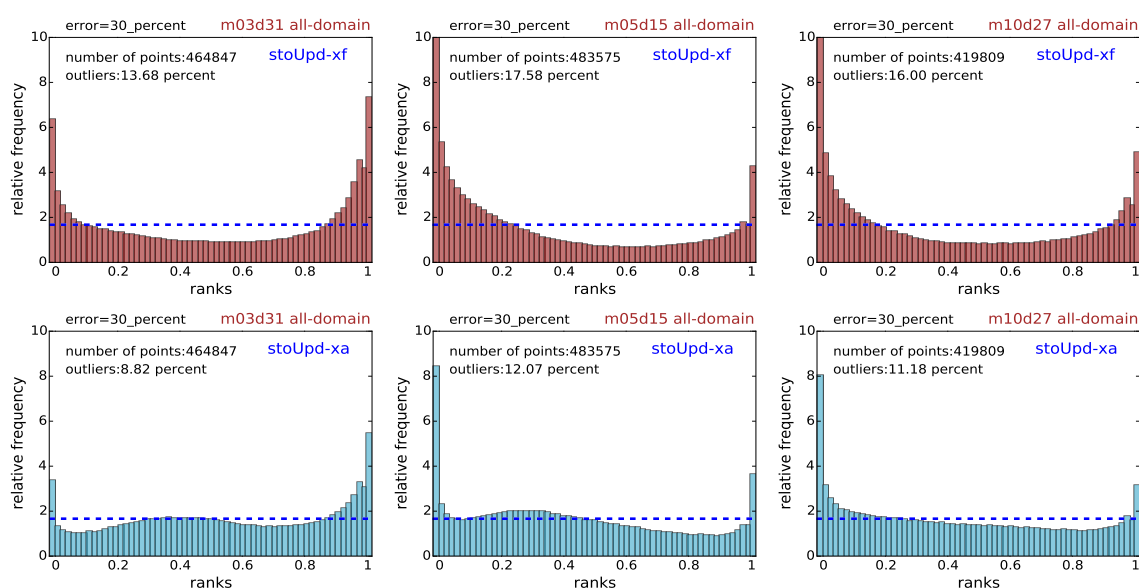


FIGURE 7.10 – Histogrammes de rangs des observations MERIS, calculés pour la totalité du domaine avec une erreur d'observation de 30% au 31 mars (1^{ère} colonne), 15 mai (2^{ème} colonne) et 27 octobre (3^{ème} colonne) 2005. La ligne du haut présente les rangs dans l'ensemble a priori et celle du bas dans l'ensemble analysé. La fréquence d'occurrence est en pourcentage (sur le nombre total d'observations). Les lignes en pointillés précisent la fréquence d'occurrence d'un histogramme plat.

2. Les données MERIS interviennent dans la construction des observations SeaWiFS-HR, cf. chapitre 3

La figure 7.10 présente les histogrammes de rangs de l'ensemble a priori (1^{ère} ligne) et de l'ensemble analysé (2^{ème} ligne), établis à une date donnée avec toutes les observations disponibles dans le domaine. Une erreur d'observation de 30% est prise en compte pour le calcul des rangs. Les histogrammes ont été construits d'une manière similaire à ceux présentés en section 6.2.2.1.

Avant toute chose, nous pouvons constater que les histogrammes de rangs de l'ensemble a priori présentent une forme très semblable à ceux obtenus, précédemment, avec les observations SeaWiFS. La fiabilité de cet ensemble est donc peu dépendante du jeu d'observation utilisé. Ce résultat permet de mettre en évidence la robustesse des paramétrisations stochastiques mises en place dans cette thèse.

En ce qui concerne l'ensemble analysé, l'impact des corrections de la chlorophylle par les observations SeaWiFS-HR apparaît assez clairement : pour les 3 dates considérées, les histogrammes sont fortement aplatis. L'analyse améliore donc significativement la fiabilité de l'ensemble. Le nombre d'observations en dehors de l'ensemble est réduit d'environ 5 à 6% et la sous-dispersion (forme en U des histogrammes) est largement diminuée par l'analyse. L'apport des informations de la couleur de l'eau permet donc de réduire la dispersion de l'ensemble (cf. section précédente), tout en améliorant la distribution statistique des membres de l'ensemble.

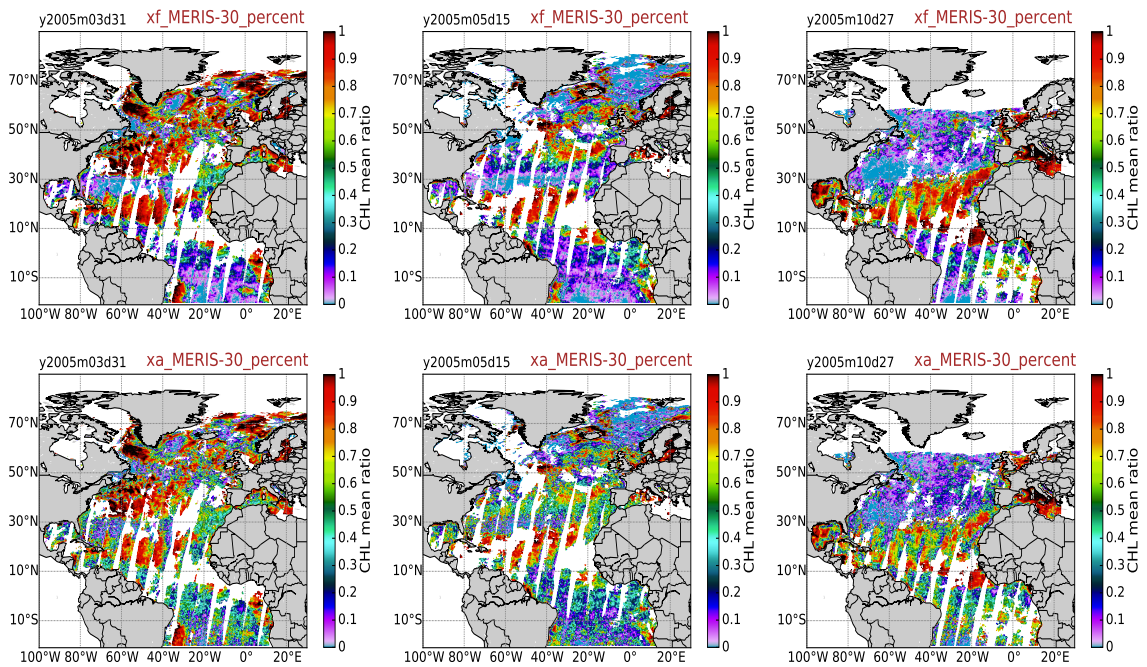


FIGURE 7.11 – Cartes des rangs des observations MERIS, calculés pour la totalité du domaine avec une erreur d'observation de 30% au 31 mars (1^{ère} colonne), 15 mai (2^{ème} colonne) et 27 octobre (3^{ème} colonne) 2005. La ligne du haut présente la répartition spatiale des rangs dans l'ensemble a priori et celle du bas dans l'ensemble analysé.

Si on s'appuie sur la figure 7.11 qui affiche la répartition spatiale des rangs des observations MERIS, nous observons une distribution des rangs bien plus hétérogène pour l'ensemble analysé. Ce résultat souligne une amélioration locale de la fiabilité et une meilleure représentation des valeurs du centre de l'intervalle des rangs. Les caractéristiques de la dispersion d'ensemble présentent ainsi une variabilité spatiale plus marquée et paraissent donc mieux adaptées aux différents comportements biogéochimiques. Certaines différenciations spatiales de grande échelle subsistent néanmoins. L'ensemble analysé peine ainsi toujours à inclure les observations au niveau du Gulf Stream (à relier avec la faible dispersion) et au niveau du gyre subtropical (tendance du modèle à sous-estimer la chlorophylle).

Répartition par provinces biogéochimiques

Pour distinguer plus précisément les variations spatiales de la fiabilité, les rangs des observations MERIS sont présentés séparément pour les provinces de Longhurst 4, 6, 7 et 9. La figure 7.12 présente les histogrammes de ces rangs calculés en considérant toutes les dates d'analyse (donc toutes les observations correspondantes).

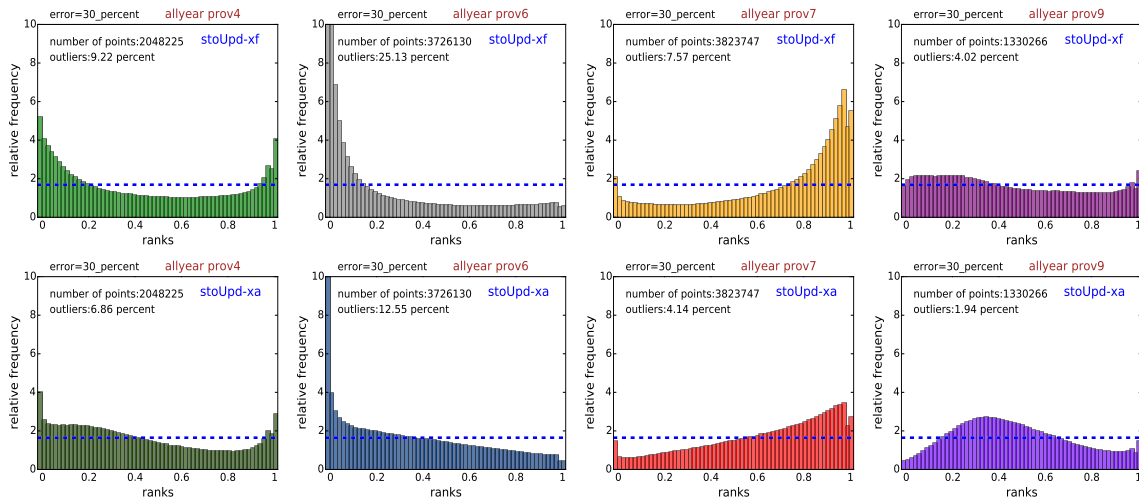


FIGURE 7.12 – Histogrammes de rangs des observations MERIS, calculés sur toute l'année 2005 avec une erreur d'observation de 30% dans les provinces de Longhurst 4 (1^{ère} ligne, 1^{ère} colonne), 6 (1^{ère} ligne, 2^{ème} colonne), 7 (2^{ème} ligne, 1^{ère} colonne) et 9 (2^{ème} ligne, 2^{ème} colonne). La ligne du haut présente les rangs dans l'ensemble a priori et celle du bas dans l'ensemble analysé. La fréquence d'occurrence est exprimée en pourcentage (sur le nombre total d'observations). La ligne en pointillés bleus précise la fréquence d'occurrence d'un histogramme plat.

Comme on pouvait s'y attendre à la vue des résultats précédents, on constate une amélioration globale de la cohérence statistique de l'ensemble analysé. En considérant simultanément toutes les analyses, le nombre d'observations en dehors de l'ensemble est partout largement réduit. Excepté pour la région 9, cette réduction s'accompagne d'une amélioration très nette de la fiabilité. Si la tendance à surestimer la chlorophylle dans la province 6 peut être identifiée par ces histogrammes, la faible dispersion n'est toutefois pas perceptible. Cela met en avant la nécessité de considérer différents diagnostics pour évaluer convenablement les caractéristiques d'une simulation probabiliste.

Afin d'affiner l'évaluation de la fiabilité dans ces différentes régions, la figure 7.13 présente les rangs des observations MERIS dans l'ensemble analysé, calculés pour les différentes saisons. En dépit des qualités statistiques de l'ensemble analysé, cette figure souligne toutefois certaines imperfections. Par exemple, nous pouvons constater que l'ensemble analysé est légèrement sous-dispersif en période estivale dans la province 4. On ne s'étendra pas plus en détail sur cette figure. L'essentiel des conclusions ont, en effet, déjà été préalablement évoquées. L'objectif de cette figure est que le lecteur puisse disposer d'une vision étendue des caractéristiques spatiales et temporelles de la fiabilité de l'ensemble analysé.

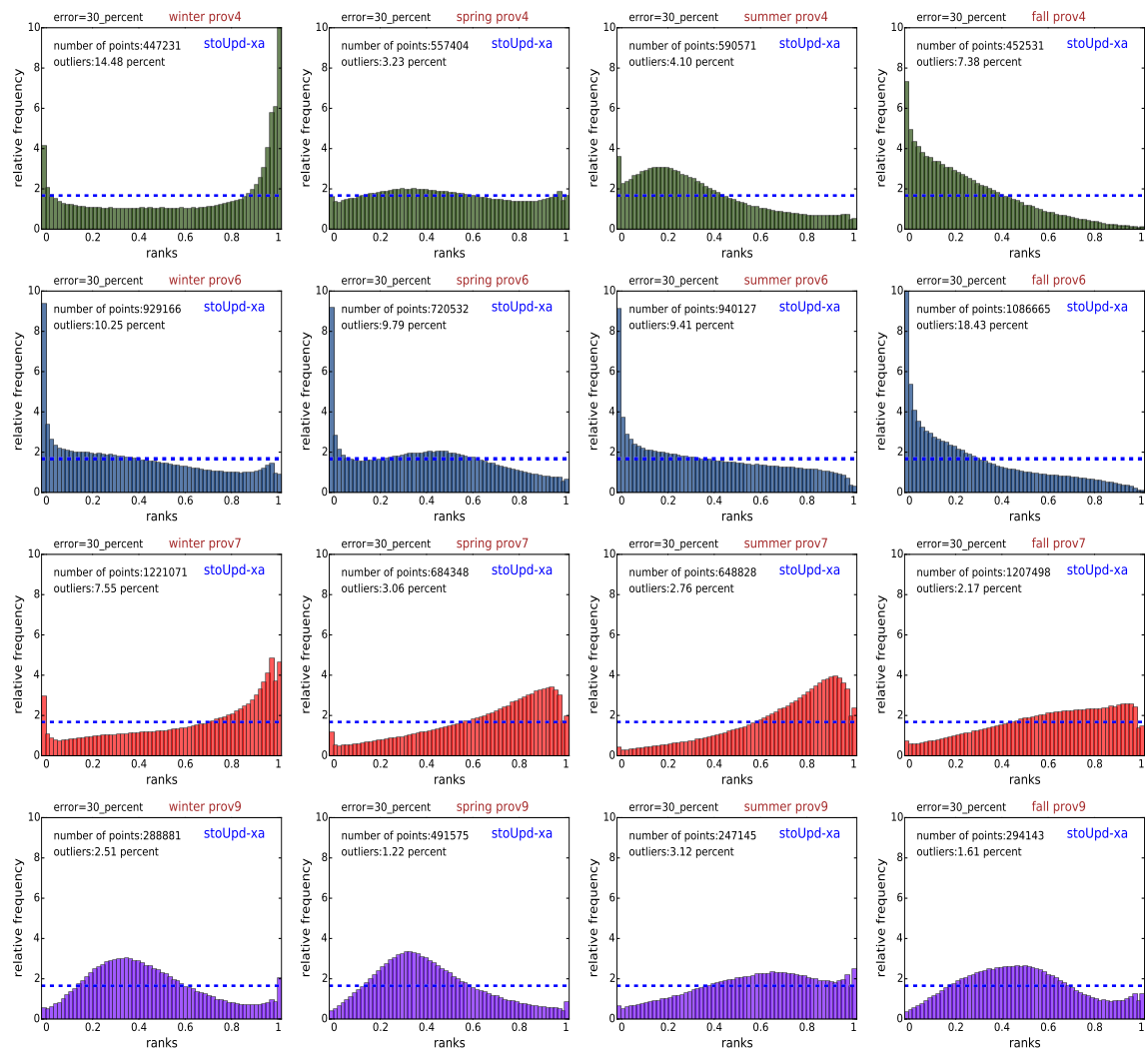


FIGURE 7.13 – Histogrammes de rangs des observations MERIS dans l'ensemble analysé présentés par saison. La 1^{ère} colonne correspond aux mois d'hiver (de janvier à mars), la 2^{ème} colonne au printemps (d'avril à juin), la 3^{ème} colonne à l'été (de juillet à septembre) et la dernière colonne à l'automne (de octobre à décembre). Les rangs sont calculés avec une erreur d'observation de 30% dans les provinces de Longhurst 4 (1^{ère} ligne), 6 (2^{ème} ligne), 7 (3^{ème} ligne) et 9 (4^{ème} ligne). La fréquence d'occurrence est exprimée en pourcentage (sur le nombre total d'observations). La ligne en pointillés bleus précise la fréquence d'occurrence d'un histogramme plat.

Bilan

Si l'amélioration de la représentation moyenne de la chlorophylle de surface était attendu (cf. section précédente), celle de la cohérence statistique entre les membres de l'ensemble analysé et les observations est un résultat fort. Légitimement, on pouvait même s'attendre à ce qu'une réduction de la dispersion ait tendance à provoquer une sous-dispersion de l'ensemble (comme en province 9). Ce n'est globalement pas le cas ici. Il est important de concevoir ce résultat comme provenant d'une définition adéquate des distributions de probabilité de l'ensemble a priori, issue des paramétrisations stochastiques et des erreurs d'observation. Grâce à cette estimation "dynamique" des incertitudes du modèle, l'information apportée par les observations ne perturbe que faiblement les distributions grande échelle de la chlorophylle. En définitive, le point important est ici encore que les performances de l'analyse dépendent très largement de la pertinence des paramétrisations stochastiques.

7.3.2 Score probabiliste CRPS

Jusqu'à présent, les comparaisons statistiques de l'ensemble avec les observations ont été uniquement réalisées à partir du calcul des rangs des observations. Évidemment, de nombreux autres diagnostics existent. Dans cette section, on introduit le CRPS (Continuous Ranked Probability Score), initialement proposé par Brown (1974). Il se présente comme une généralisation à plusieurs événements continus du score de Brier (Brier, 1950), qui évalue la qualité d'une prévision probabiliste d'un événement binaire. De manière simplifiée, le CRPS correspond à la distance entre la densité de probabilité cumulée $F_p(x)$ prédite d'une variable x (ici la chlorophylle) et l'observation y_o de cette même variable, exprimée sous la forme d'une distribution de probabilité $F_o(x)$. Il est souvent perçu comme l'équivalent probabiliste de la moyenne de la valeur absolue des écarts aux observations, diagnostic régulièrement utilisé pour caractériser les prévisions déterministes. Il s'exprime de la façon suivante (Candille and Talagrand, 2005) :

$$CRPS = E \left[\int_{\mathbb{R}} (F_p(x) - F_o(x))^2 dx \right] \quad (7.19)$$

où $E[.]$ représente la moyenne de cette intégrale sur l'ensemble des vérifications que l'on choisit de considérer (ensemble spatio-temporel d'observations). Dans notre cas, on ne considèrera pas d'erreur d'observation. En un point de grille donné, $F_o(x)$ est donc une simple fonction de Heaviside explicité par l'équation 7.20.

$$H(x) = \begin{cases} 0 & \text{si } x < y_o \\ 1 & \text{si } x \geq y_o \end{cases} \quad (7.20)$$

Pour une description approfondie des différentes caractéristiques mathématiques du CRPS, le lecteur peut se référer à Candille and Talagrand, 2005; Gneiting et al., 2005 ou Candille et al., 2015. Pour cette section, il est surtout important de retenir que le CRPS donne une quantification comprise entre 0 et 1 de la cohérence statistique globale d'un système probabiliste. Plus sa valeur est proche de 0, meilleur est le score. L'unité du CRPS

étant la même que celle de la variable qu'il quantifie, il est dépendant de la gamme de valeur considérée. Une valeur seule de ce score ne signifie donc pas qu'il est bon ou mauvais. En ce sens, le CRPS sera surtout adapté à la comparaison entre différentes estimations statistiques d'un même système. Dans cette section, le CRPS de l'ensemble analysé est ainsi comparé avec celui de l'ensemble a priori. Comme dans la partie précédente, les scores sont calculés en utilisant les observations MERIS.

7.3.2.1 Variabilité temporelle

Pour évaluer l'impact de l'analyse, la valeur du CRPS est calculée tous les 5 jours à partir de toutes les observations disponibles sur le domaine ($E[.]$ est une moyenne spatiale). La figure 7.14 présente ces résultats sous forme de séries temporelles de l'ensemble a priori (courbe en pointillés bleus) et de l'ensemble analysé (courbe rouge). À chaque pas de temps, cette figure montre que l'analyse améliore nettement (d'environ 10-15%) les valeurs de CRPS. Elle est maximale entre mai et septembre et légèrement plus faible pendant l'hiver

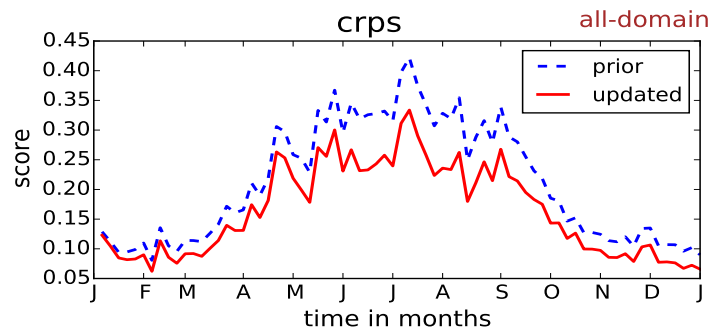


FIGURE 7.14 – Séries temporelles, sur l'année 2005, des valeurs de CRPS calculées pour la simulation d'ensemble a priori *stoUpd-xf* (courbe en pointillés bleus) et la simulation de l'ensemble analysé *stoUpd-xa* (courbe rouge). À chaque pas de temps le CRPS est calculé à partir de toutes les observations disponibles sur le domaine.

Puisque le CRPS possède l'unité de la variable considérée, il est nécessaire de pouvoir différencier les différents comportements biogéochimiques (les scores ont tendance à être plus élevés lorsque les concentrations en chlorophylle sont importantes). À chaque pas de temps, la moyenne est ainsi calculée en utilisant toutes les observations disponibles dans une province donnée. En figure 7.15, les valeurs de CRPS sont calculées pour les provinces de Longhurst 4, 6, 7 et 9.

Comme précédemment, on constate que l'analyse améliore toujours les valeurs du CRPS. Dans les provinces 7 et 9, cette amélioration est relativement constante, sans véritable distinction saisonnière. Dans les plus hautes latitudes (province 4), le CRPS de l'ensemble analysé est uniquement amélioré pour la période allant de mai à octobre. Dans la province 6, l'importante diminution des quantités de chlorophylle, provoquée par l'analyse, provoque une forte amélioration du CRPS. Comme pour l'évaluation des histogrammes de rangs, ce résultat positif masque cependant la trop faible dispersion de l'ensemble analysé dans la région du Gulf Stream.

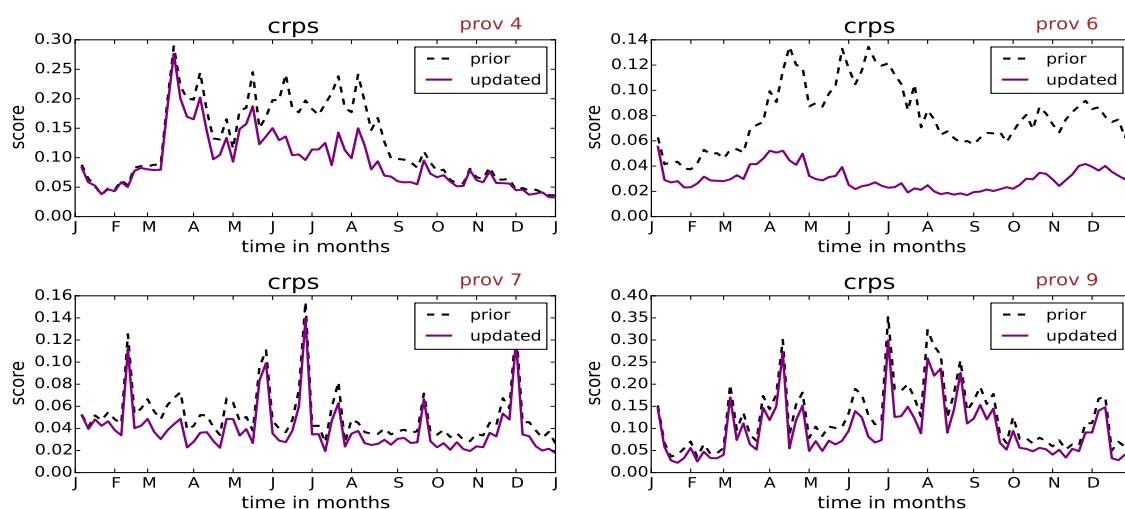


FIGURE 7.15 – Séries temporelles, sur l'année 2005, des valeurs de CRPS calculées dans les provinces de Longhurst 4, 6, 7 et 9. Les courbes en pointillés noirs présentent les résultats de l'ensemble a priori *stoUpd-xf* et les courbes violettes ceux de l'ensemble analysé *stoUpd-xa*. À chaque pas de temps, le CRPS est calculé à partir de toutes les observations disponibles dans les provinces de Longhurst considérées.

7.3.2.2 Décomposition du score

Afin de différencier les caractéristiques statistiques de l'ensemble, il est possible de décomposer le CRPS (équation 7.19) en une somme de 2 termes : $CRPS = RELI + RESOL$. Le 1^{er} terme, RELI, exprime la fiabilité de l'ensemble et le terme, RESOL, est un terme de résolution potentielle, c'est-à-dire la résolution qu'aurait l'ensemble si la fiabilité était parfaite. Cette procédure est détaillée dans Hersbach (2000).

La résolution statistique qui est examinée ici exprime la capacité du système probabiliste à distinguer différentes situations observées. Pour l'illustrer, reprenons l'exemple du chapitre précédent. Imaginons que 10% des données d'une climatologie suggèrent des concentrations en chlorophylle supérieures à 4 mg/m^3 dans l'Atlantique Nord. Une simulation d'ensemble dont 10% des membres vérifient cette estimation en chaque point de grille sera considérée comme fiable. Cependant, aucune information n'est donnée quand aux variations spatiales de cet événement. La résolution sera donc faible. Par contre, si une simulation est capable de différencier spatialement ces 10% en estimant, par exemple, que 15% des membres de la partie nord et que 5% des membres de la partie sud possèdent des concentrations en chlorophylle supérieures à 4 mg/m^3 , la résolution de l'ensemble sera améliorée. En résumé, la résolution quantifie en quelque sorte l'information complémentaire apportée par le modèle par rapport à des observations climatologiques. Elle peut également être perçue comme une mesure de l'étendue des incertitudes associées à la distribution de probabilité.

Les 2 critères (RESO et RELI) sont présentés en figure 7.14 sous forme de séries temporelles, établies à partir de toutes les observations disponibles sur le domaine à chaque pas de temps. Cette figure est comparable à la figure 7.14 pour le CRPS total.

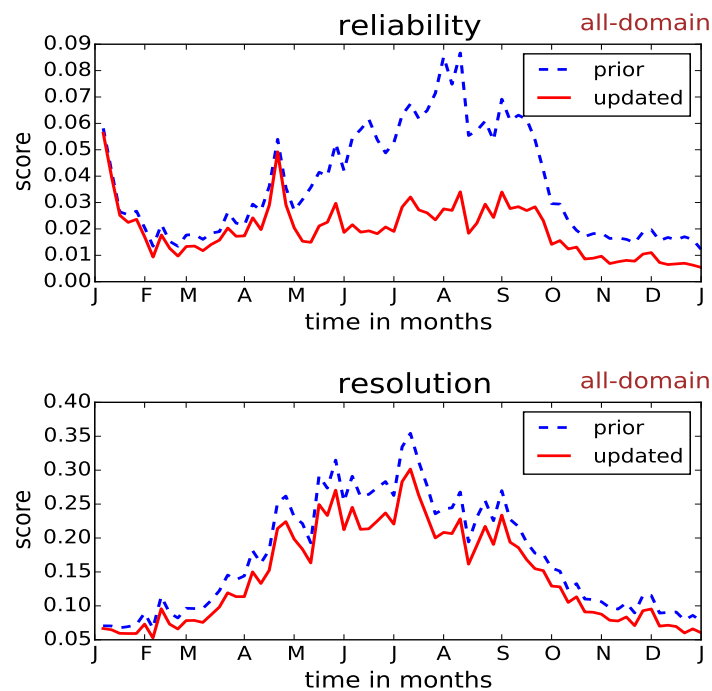


FIGURE 7.16 – Séries temporelles, sur l'année 2005, des valeurs de la fiabilité (figure du haut) et de la résolution (figure du bas) de la simulation d'ensemble a priori *stoUpd-xf* (courbe en pointillés bleus) et de l'ensemble analysé *stoUpd-xa* (courbe rouge), calculées à partir de la décomposition du CRPS. Toutes les observations disponibles sur le domaine à chaque pas de temps sont utilisées.

Comme nous avons pu l'observer à partir des histogrammes de rangs, cette figure montre que l'analyse permet d'améliorer systématiquement la fiabilité de l'ensemble. Si ce résultat est relativement modéré en début d'année, toute la dispersion de l'ensemble n'étant pas encore générée par les paramétrisations stochastiques, la fiabilité est significativement améliorée pendant la période de printemps-été.

En ce qui concerne la résolution, l'amélioration du score due à l'analyse reste toujours relativement constante. Ce constat peut s'expliquer par une diminution de la probabilité d'existence des événements extrêmes (« queues » des distributions de probabilité), ce qui provoque un resserrement de l'étendue des distributions de probabilité. Puisque cette diminution s'accompagne d'une meilleure fiabilité, les résultats obtenus sur la résolution indiquent que l'analyse permet de discriminer plus précisément les différents états biogéochimiques identifiables dans les observations.

Pour rester cohérent avec la démarche utilisée en section précédente, la figure 7.17 présente les séries temporelles des 2 termes du CRPS (RESO et RELI), calculés pour les provinces de Longhurst 4, 6, 7 et 9. Pour la fiabilité, on constate des améliorations principalement au niveau de la province 6. Dans les autres provinces, la fiabilité de l'ensemble analysé n'est que légèrement améliorée. L'essentiel des conclusions qui auraient pu être ajoutées ayant d'ores et déjà été évoquées à plusieurs reprises, nous n'entrerons pas ici plus dans le détail.

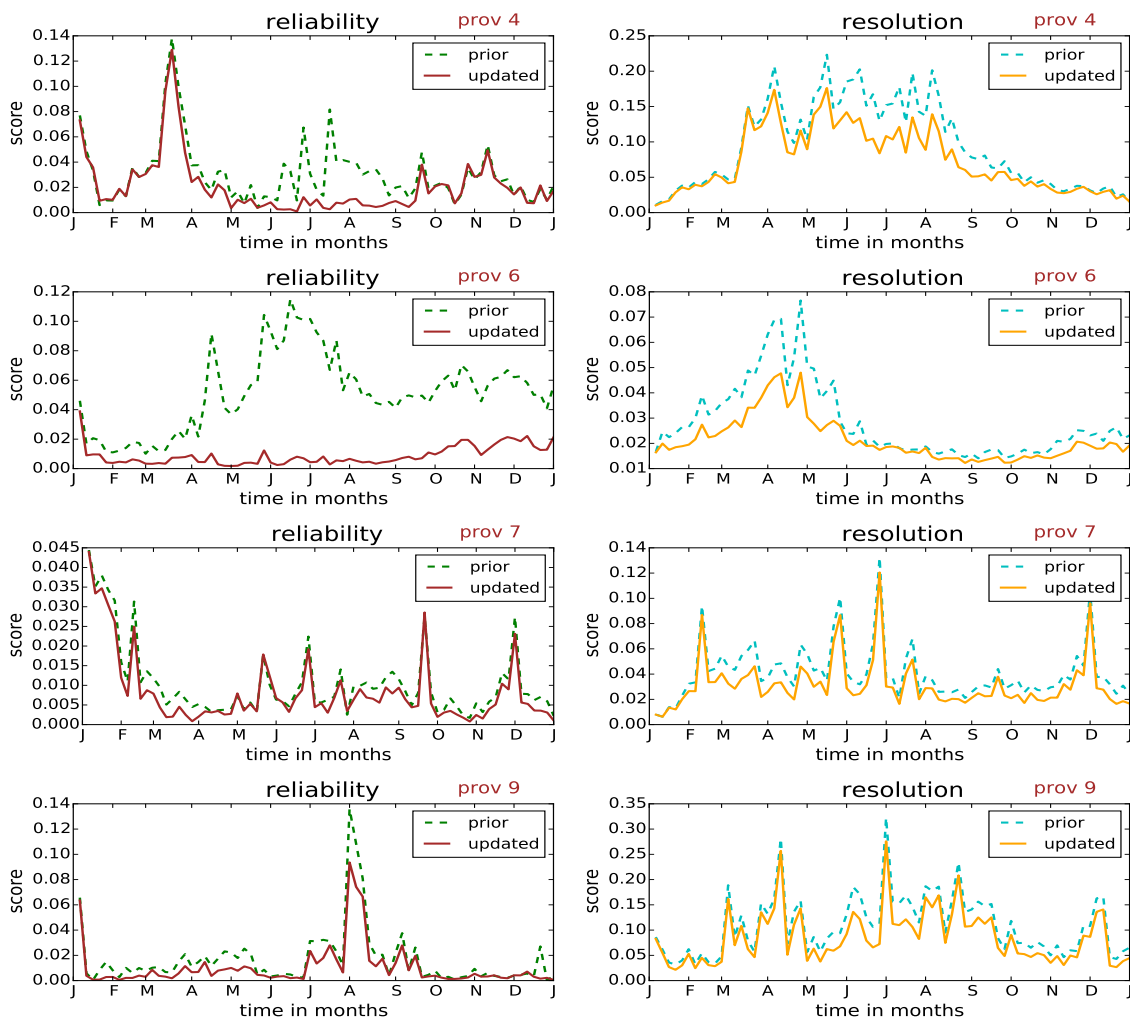


FIGURE 7.17 – Séries temporelles sur l'année 2005 des valeurs de de fiabilité (1^{ère} colonne) et de résolution (2^{ème} colonne) obtenues pour les provinces de Longhurst 4, 6, 7 et 9. Les courbes en pointillés affichent les résultats pour l'ensemble a priori *stoUpd-xf* et les courbes pleines de l'ensemble analysé *stoUpd-xa* obtenus à partir de la décomposition du CRPS. La moyenne utilisée est spatiale, réalisée avec toutes les observations disponibles dans les différentes provinces de Longhurst.

Bilan

Dans cette section, nous avons évalué statistiquement l'impact des analyses sur la simulation d'ensemble. De manière générale, tous les diagnostics ont mis en avant l'impact positif des corrections sur les distributions de la chlorophylle de surface. Un des résultats importants de cette thèse est donc que la démarche utilisée pour réaliser les analyses (qui sous-entend la prise en compte de l'incertitude du modèle) paraît appropriée à l'assimilation de données de la couleur de l'eau.

Additionné aux histogrammes de rangs, le CRPS a été un moyen efficace pour quantifier l'impact des analyses en conservant une vision statistique. Toutefois, il évalue principalement la qualité de la prévision par rapport aux observations. Même avec une dispersion

bien trop faible dans la région du Gulf Stream, la cohérence statistique de l'ensemble analysé est en effet toujours améliorée. Ces scores sont donc nécessaires mais pas suffisants. Ce constat souligne une des difficultés majeures de l'approche probabiliste. Comment définir des diagnostics adaptés? Une partie de la réponse est probablement de ne pas se limiter à un seul diagnostic et d'évaluer de manière variée les différentes caractéristiques d'un ensemble.

Pour conclure, je précise une nouvelle fois que les résultats obtenus sont particulièrement encourageants, compte tenu du caractère exploratoire des expériences réalisées ici. Ils constituent un pas de plus permettant de valoriser le potentiel de l'approche probabiliste pour l'étude de la biogéochimie marine.

7.4 Correction des variables non observées

Jusqu'ici, l'étude de la simulation d'ensemble et l'évaluation de l'impact des analyses se sont uniquement focalisées sur la variable chlorophylle. Telles qu'établies, les paramétrisations stochastiques avaient, en effet, pour objectif de simuler des sources d'incertitude biogéochimiques agissant principalement sur la production primaire. À terme, l'objectif sera de prolonger ces travaux de thèse à travers des expériences d'assimilation de données qui seront réalisées dans un contexte opérationnel (projet AtlantOS, cf. conclusion). Un schéma d'analyse multivariée a donc été implémenté. Cela permet de pouvoir déjà examiner le potentiel des corrélations entre les variables obtenues à partir des paramétrisations stochastiques. L'impact des corrections d'une seule variable sur l'ensemble du système reste en effet un verrou important à l'assimilation de données dans les modèles biogéochimiques.

En plus de la chlorophylle totale, les corrections sur la chlorophylle de surface impactent ici les 5 nutriments : les nitrates (NO_3), l'ammonium (NH_4), le phosphate (PO_4), le Fer (Fe) et le silicate (Si), la chlorophylle des 2 classes d'espèces : les diatomées (CHL^D) et le nanophytoplancton (CHL^N) ainsi que les concentrations en microzooplancton (Z_m , notées ici ZOO). À noter que nous ne présenterons pas dans cette section l'impact des analyses sur les concentrations en phytoplancton (C^P), les résultats étant comparables à ceux obtenus sur la chlorophylle.

Dans la première partie de cette section, nous évaluerons l'impact des corrections sur les distributions de surface des variables non observées, à partir de séries temporelles des membres de l'ensemble. Ce point permettra une nouvelle fois de mettre en évidence l'importance de la définition initiale des paramétrisations stochastiques. Dans la seconde partie, nous examinerons l'impact des analyses sur les variables non observées dans la colonne d'eau, à partir de profils verticaux des membres de l'ensemble.

7.4.1 Dispersion des séries temporelles

Pour évaluer l'impact des corrections sur les variables non observées, la figure 7.18 présente les séries temporelles de surface des membres de l'ensemble a priori (courbes vertes) et de l'ensemble analysé (courbes bleues) aux points de grille stations STAT_D et STAT_E. À noter que le silicate et le fer sont uniquement représentés à la station STAT_E.

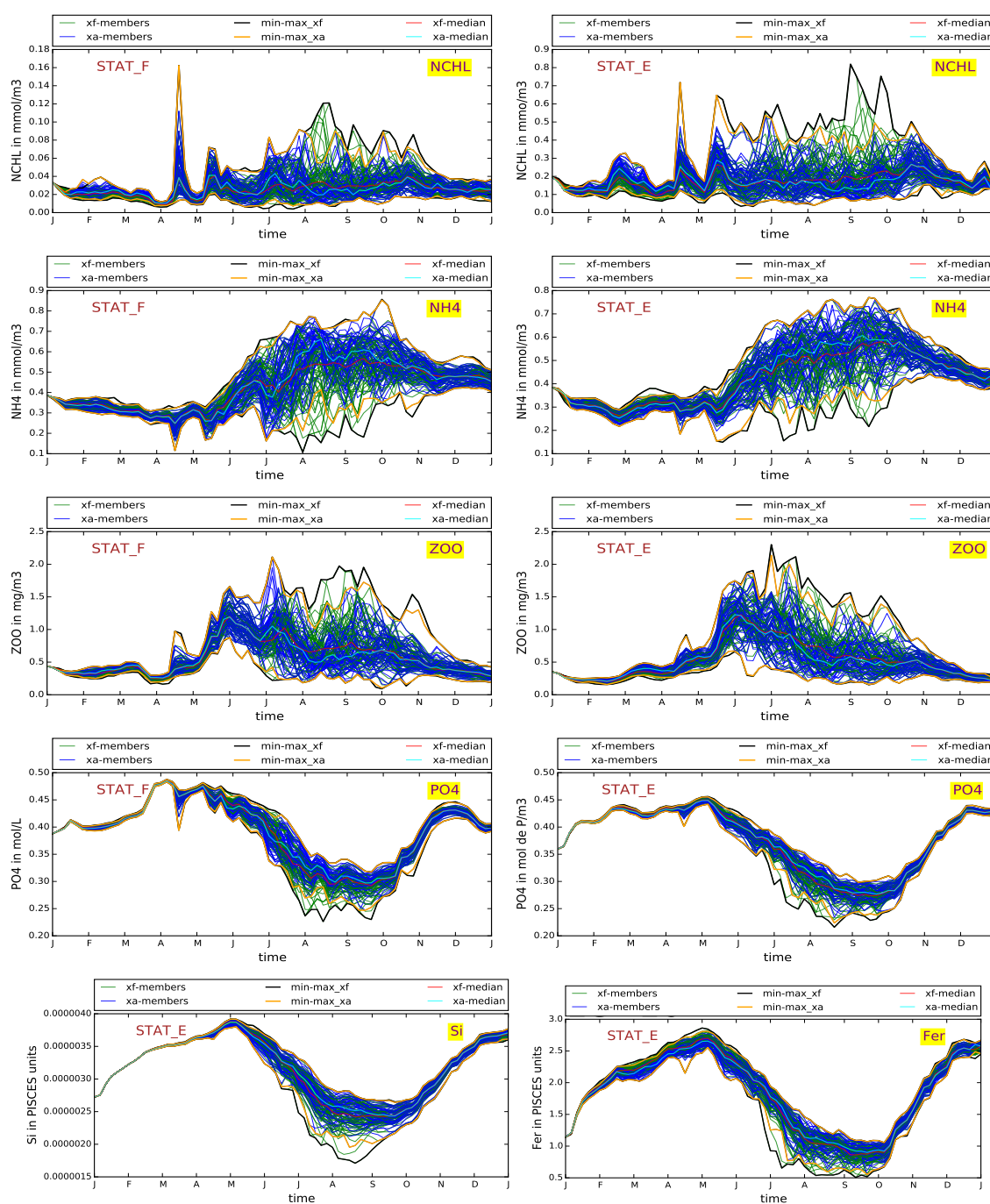


FIGURE 7.18 – Séries temporelles de surface de l'ensemble a priori (courbes vertes) et de l'ensemble analysé (courbes bleues) présentées aux points de grille stations STAT_F (1^{ère} colonne et STAT_E (2^{ème} colonne) pour 5 variables non observées impactées par les analyses. De haut en bas sont représentés : la chlorophylle du nanophytoplancton (NCHL), l'ammonium (NH₄), le microzooplancton (ZOO), et le phosphate (PO₄). La dernière ligne présente les séries temporelles des silicates (1^{ère} colonne) et du fer (2^{ème} colonne), uniquement à la station STAT_E.

Tout d'abord, la figure 7.18 souligne la faculté de l'ensemble à présenter de la dispersion sur des variables indirectement concernées par les paramétrisations stochastiques. Cette dispersion présente, de plus, des caractéristiques temporelles cohérentes avec celles que nous avons observées pour la chlorophylle de surface (cf. figure 7.6 et 6.4).

Les variables qui affichent les niveaux de dispersion les plus élevés sont le microzooplancton et le nanophytoplancton. En ce qui concerne le zooplancton, ce résultat provient certainement des perturbations affectant la sensibilité du broutage à la température b'_Z . Pour les nutriments, la dispersion est globalement plus faible. Cela peut s'expliquer par le fait qu'aucun paramètre du modèle agissant au premier ordre sur les nutriments n'ait été perturbé (cf. section 5.1.2). À noter toutefois qu'une dispersion plutôt importante peut être observée sur l'ammonium.

Sur les séries temporelles de l'ensemble analysé, il est satisfaisant de voir que les corrections impactent les distributions statistiques de toutes les variables biogéochimiques présentées. Ces effets sont particulièrement visibles sur les courbes de l'évolution temporelle des médianes. Dans tous les cas, la figure 7.18 montre que l'analyse provoque une réduction significative de la dispersion d'ensemble. Elle est principalement perceptible par la diminution de l'étendue de l'enveloppe des distributions des membres de l'ensemble.

Assez logiquement, nous pouvons constater que l'effet des analyses sur les variables non observées est d'autant plus grand que la dispersion de l'ensemble a priori est importante. Les corrections respectent donc bien le niveau d'incertitude a priori défini par les paramétrisations stochastiques. Il reste cependant difficile de juger du bien-fondé des distributions de probabilité a priori, comme des corrections, sans effectuer de comparaisons avec des observations. Il sera important d'aborder ce point afin de poursuivre ces travaux.

La figure 7.18 montre également que les membres de l'ensemble analysé sont répartis de manière homogène à l'intérieur de l'enveloppe de l'ensemble a priori. Ce point, déjà observé lorsque nous avons examiné l'impact des analyses sur les distributions verticales de la chlorophylle (cf. figure 7.7), montre que la méthode d'anamorphose ne contraint pas les membres de l'ensemble en proximité des extrema.

Bilan

En définitive, les caractéristiques de l'ensemble a priori sont toujours conservées lors de l'analyse. Ces résultats soulignent l'importance d'une définition initiale pertinente des statistiques d'erreur du modèle. Dans un objectif d'assimilation de données multivariée, il sera donc probablement nécessaire de mettre en place des paramétrisations stochastiques plus élaborées. Le choix initial des incertitudes simulées devra ainsi être établi en fonction : des observations qui seront assimilées et des variables que l'on souhaite corriger par l'assimilation de données.

7.4.2 Dispersion des profils verticaux

Sur la figure 7.19, nous présentons pour plusieurs variables (non observées) les profils verticaux des membres de l'ensemble a priori (courbes vertes) et de l'ensemble analysé (courbes bleues) aux points de grille stations STAT_D et STAT_E. Les profils verticaux des silicates sont uniquement présentés au point de grille STAT_E.

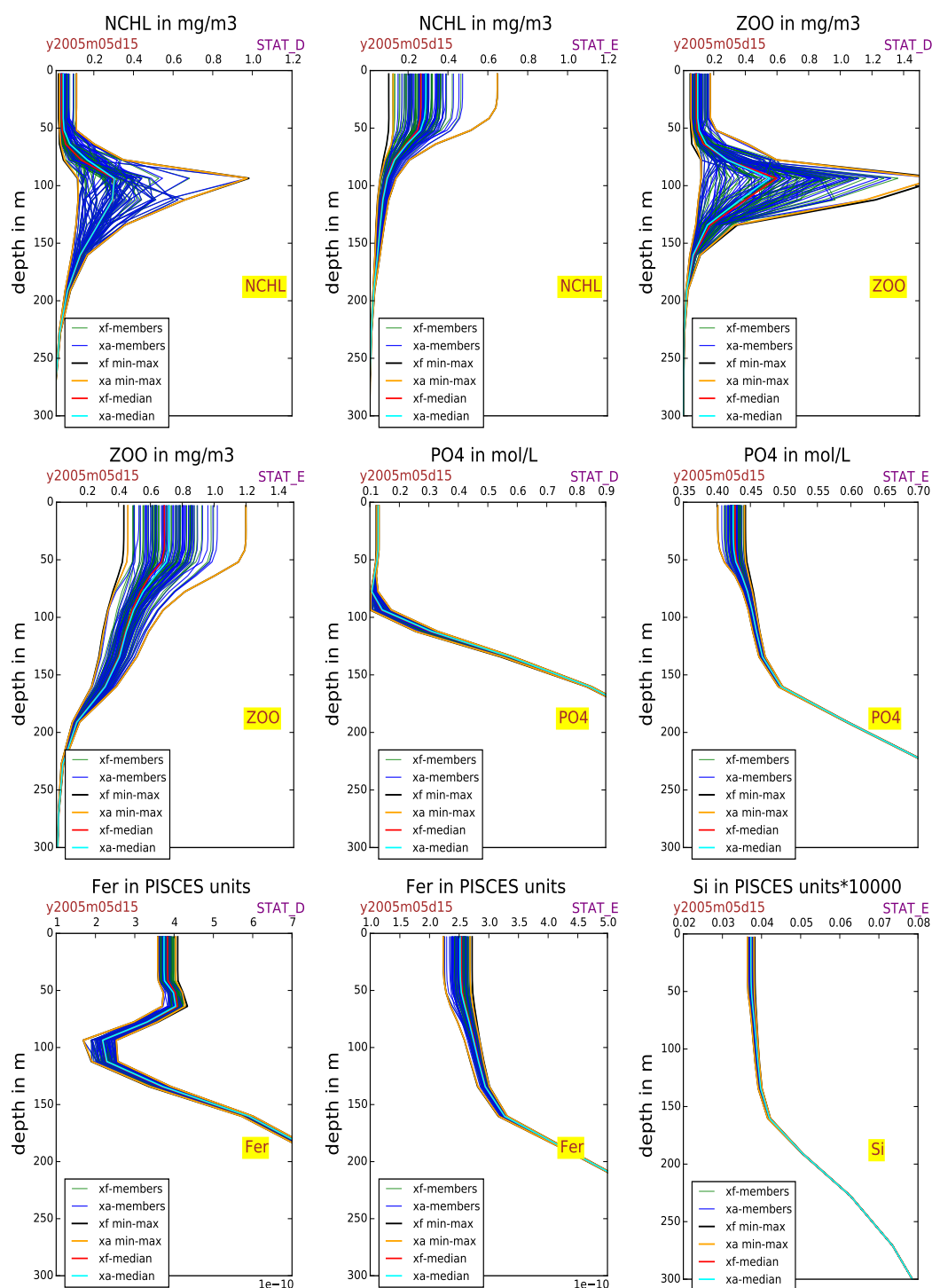


FIGURE 7.19 – Profils verticaux de l'ensemble a priori (courbes vertes) et de l'ensemble analysé (courbes colorées) aux points de grille stations STAT_D et STAT_E pour 5 variables non observées impactées par les analyses. Les informations nécessaires à l'identification des différentes variables sont indiquées sur les figures.

Mis à part pour le nanophytoplancton et le microzooplancton, dont certains paramètres régissant les concentrations ont été considérés comme incertains, la dispersion verticale est globalement faible. Par conséquent, l'impact des analyses est le plus souvent quasiment imperceptible. À noter que ce constat est également valable pour le microzooplancton et le nanophytoplancton. Le point positif reste toutefois que l'analyse ne semble pas induire d'effets négatifs sur les distributions verticales.

Une explication éventuelle aux faibles effets de l'analyse est peut être l'invariance des perturbations stochastiques suivant la verticale (cf. section 5.1.2.2). Si cette propriété nous a permis de conserver des structures verticales cohérentes avec la simulation déterministe, il est possible qu'elle réduise également l'impact des paramétrisations stochastiques. Des expériences supplémentaires seraient toutefois nécessaires pour vérifier ce point. Dans le cas présent, le constat est qu'excepté pour la chlorophylle, les données de la couleur de l'eau n'apportent presque aucune information sur la verticale. Ce résultat reste néanmoins important et pourra être mis en avant pour favoriser le développement des réseaux d'observation comme les bouées BIO-ARGO.

7.5 Conclusion

Dans ce chapitre, nous avons présenté une séquence d'analyses d'ensemble multivariées, réalisée en utilisant les distributions de probabilité définies par la simulation d'ensemble a priori *stoUpd*. Globalement, les résultats ont permis de justifier la pertinence de notre approche probabiliste, ainsi que des différents paramètres d'analyse (anamorphose, corrélations des erreurs d'observation, etc...). Nous avons, en particulier, pu montrer que l'analyse améliore toujours les caractéristiques statistiques de fiabilité et de résolution de l'ensemble.

Le point que nous avons abordé qui est certainement le plus déterminant est celui de la définition initiale des paramétrisations stochastiques. Elles sont à l'origine de la représentation des erreurs de prévision, dont dépendent fortement les performances de l'assimilation de données. Ainsi, plutôt que de se focaliser principalement sur l'établissement de méthodes avancées, affiner la mise en place des paramétrisations stochastiques paraît essentiel à la démarche d'ensemble en assimilation de données. Vérifier que l'ensemble a priori possédait des caractéristiques statistiques en accord avec les observations SeaWIFS (cf. chapitre 6) a ainsi certainement été à la base des bonnes performances du système présenté dans ce chapitre.

En effet, lorsque les observations sont incluses dans l'ensemble a priori, l'état corrigé l'est également. L'état biogéochimique de l'ensemble actualisé restera à l'intérieur d'une dynamique caractérisée par le modèle. À l'opposé, si l'analyse contraint l'état biogéochimique dans un espace des solutions qui n'a pas été au préalable exploré par le modèle, il n'est pas certain que les corrélations d'erreur restent valables et que la dynamique biogéochimique du modèle puisse être stable. L'importance de ce dernier point est, de plus, renforcé par l'obligation de conserver l'information apportée par la couleur de l'eau afin de pouvoir répéter les cycles d'assimilation.

Les résultats présentés dans ce chapitre constituent une première expérience d'analyse d'ensemble avec le modèle PISCES. Dans ce contexte, l'objectif a donc été de mettre en avant le potentiel d'une approche stochastique pour l'assimilation de données biogéochimiques. Certains paramètres, notamment en ce qui concerne la prise en compte des erreurs d'observation, pourront bien sûr être améliorés. Il sera par exemple nécessaire de leur fixer un seuil limite minimal afin d'éviter l'effondrement de l'ensemble. Je précise que l'un des avantages de ce système d'analyse est qu'il est peu dépendant du type d'observation utilisé.

Dans la perspective BIO-ARGO, il sera tout à fait concevable d'assimiler ces données de subsurface de la même manière que les données de la couleur de l'eau. Le seul impératif sera de s'assurer que les paramétrisations stochastiques permettent bien, dans la colonne d'eau, de définir un niveau d'incertitude en accord avec ces observations. En ce qui concerne la chlorophylle, ce devrait vraisemblablement être le cas.

On rappelle les principaux résultats de ce chapitre.

- La méthode d'analyse d'ensemble qui a été implémentée permet de corriger les sorties du modèle PISCES à partir d'observations de la couleur de l'eau.
- L'utilisation des distributions de probabilité de l'ensemble a priori, définies à partir des paramétrisations stochastiques, améliore l'état moyen de la chlorophylle tout en réduisant l'incertitude des prévisions.
- L'extrapolation verticale des corrections sur la chlorophylle de surface impacte l'ensemble de la colonne d'eau sans altérer les caractéristiques de grande échelle de l'ensemble a priori.
- Les analyses améliorent toujours les caractéristiques statistiques de fiabilité et de résolution de l'ensemble.
- Quelle que soit la variable considérée, l'ensemble analysé reste toujours inclus dans l'enveloppe de la distribution de probabilité a priori.

Chapitre 8

Conclusion

En définitive, prévoir «l’océan qu’il fera», pour reprendre des propos déjà existants, est encore aujourd’hui un véritable challenge. Il repose à la fois sur notre faculté à comprendre le fonctionnement de multiples processus en interaction et sur notre capacité à pouvoir modéliser l’étendue de leur large spectre spatio-temporel. Le moins que l’on puisse dire, c’est que la tâche s’avère aussi passionnante que compliquée.

Dans cette thèse, nous avons abordé le sujet selon une démarche orientée par l’assimilation de données, c’est-à-dire avec une approche intégrée modèle/observation. C’est l’estimation de la production primaire de l’océan Atlantique Nord, élément essentiel du cycle du carbone, dont il a principalement été question. La partie modélisation a été réalisée à partir de la plate-forme couplée NEMO/PISCES et les observations qui ont été utilisées sont les données satellites de la couleur de l’océan. Dès le départ, nous avons pu constater toute la difficulté pour confronter ces 2 différentes sources d’information.

L’analyse des formulations du modèle PISCES a rapidement mis en évidence leur forte dépendance vis à vis d’un grand nombre de paramètres, généralement établis de manière approximative. Les équations des modèles biogéochimiques sont en effet basées sur des lois empiriques. Elles ne peuvent donc que partiellement rendre compte de la complexité des processus qui régissent le fonctionnement des écosystèmes marins. Or, nous ne pouvons espérer mieux d’un modèle que ce qui y est formellement représenté.

Un effort particulier a ainsi été porté autour du concept d’incertitude. Le manque de connaissances du comportement biogéochimique de l’océan nous a conduits à considérer explicitement l’incertitude des formulations biogéochimiques du modèle. Évidemment, la multitude des sources potentielles d’incertitude nous a contraints à nous focaliser sur seulement quelques sources d’incertitude, jugées capitales pour la production primaire. L’utilisation d’une approche probabiliste qui s’est traduite par une simulation d’ensemble, a ensuite permis d’analyser statistiquement l’impact de ces incertitudes. Un aspect essentiel de cette thèse a ainsi été de “démystifier” l’approche probabiliste. l’objectif était de montrer que l’information dont nous disposons sur un système peut être optimisée en modélisant explicitement son incertitude. L’idée sous-jacente était d’arriver à présenter cette approche comme une alternative crédible à la modélisation biogéochimique. En ce qui concerne l’assimilation de données, nous avons ensuite pu effectuer une séquence d’analyses à partir de l’erreur de prévision basée sur l’incertitude du modèle.

Nous rappelons les objectifs de départ de cette thèse, tels qu'ils ont été énoncés en introduction :

- Mettre en place une configuration réaliste au $1/4^\circ$ du modèle couplé NEMO/PISCES adaptée à l'océan Atlantique Nord et évaluer sa pertinence pour la représentation de la production primaire.
- Faire évoluer cette configuration vers un système d'ensemble capable de présenter un niveau d'incertitude cohérent avec les écarts entre le modèle et les observations satellites.
- Examiner les aptitudes de ce système dans un contexte d'assimilation de données de la couleur de l'océan.

En premier lieu, il convient de mentionner que ces 3 objectifs ont pu être atteints. Nous reviendrons donc d'abord ici sur la démarche que nous avons suivie pour y parvenir. Nous présenterons ensuite un bilan récapitulatif, décrivant les 3 principaux développements techniques mis en place durant cette thèse et les résultats importants auxquels ils ont permis d'aboutir. Dans la dernière partie de cette conclusion, nous discuterons de quelques perspectives issues des travaux exposés dans ce manuscrit.

8.1 Démarche, développements et résultats

Démarche

Dans la première partie de ce manuscrit, nous avons présenté le contexte qui a justifié la mise en place du modèle couplé NEMO/PISCES au $1/4^\circ$ dans l'océan Atlantique Nord. La configuration NATL025, issue du projet DRAKKAR, a été reprise pour assurer une représentation réaliste de la physique. Nous avons ensuite réalisé une simulation déterministe de 6 années (2005-2010), de ce modèle couplé physique-biogéochimie. Sa comparaison avec une simulation du modèle LOBSTER, construite sur l'exacte même configuration (NATL025), a permis de démontrer : 1) la capacité du modèle PISCES à convenablement simuler les principales caractéristiques de grande échelle présentes dans les observations de la couleur de l'eau SeaWIFS et 2) que l'augmentation du niveau de complexité des formulations biogéochimiques du modèle PISCES permet une meilleure estimation de la chlorophylle de surface. Malgré ces améliorations, la confrontation avec les observations a toutefois mis en évidence une estimation encore imparfaite de la chlorophylle.

À partir de ces considérations, l'identification des sources d'incertitudes à l'origine de ces imperfections a mobilisé toute notre attention. L'analyse de la simulation déterministe a souligné une contribution en général localisée et réduite des termes de forçages physiques à la variabilité interannuelle. Dans le même temps, les incertitudes sur les formulations du modèle, notamment en ce qui concerne la diversité des comportements biogéochimiques, paraissent influencer fortement l'estimation de la production primaire. Nous avons ainsi choisi de focaliser notre étude sur cette source particulière d'incertitude. Il existe bien évidemment d'autres sources d'incertitude dont l'impact aurait également pu être examiné.

Dans la seconde partie de ce document, nous avons proposé une méthode permettant de simuler des incertitudes biogéochimiques du modèle PISCES, c'est-à-dire de tenir explicitement compte de l'erreur modèle. En utilisant des perturbations aléatoires gaussiennes, générées à partir de processus autorégressifs AR(1), nous avons pour cela mis en place des paramétrisations stochastiques pour simuler 2 sources d'incertitude : 1) les incertitudes sur les paramètres biogéochimiques du modèle et 2) les incertitudes des échelles non résolues issues de la non-linéarité des équations biogéochimiques.

Les premières simulations réalisées avec ces paramétrisations stochastiques ont montré qu'elles avaient tendance à stimuler la production en chlorophylle du modèle. Par conséquent, la chlorophylle de surface était fortement surestimée par rapport aux observations SeaWiFS. Cet effet a pu être identifié comme étant principalement le résultat d'un effet moyen, due à la non-linéarité des équations. Pour compenser cet effet, c'est-à-dire accorder le modèle avec l'approche probabiliste, un réajustement de la valeur initiale de certains paramètres (déterministes) du modèle PISCES a été proposé. **Nous avons ensuite pu établir une simulation d'ensemble de 60 membres dont la structure moyenne ainsi que le niveau d'incertitude étaient cohérents avec la simulation déterministe et les données SeaWiFS.**

Initialement, l'objectif des paramétrisations stochastiques était d'améliorer la représentation des statistiques d'erreur de prévision pour l'assimilation de données de la couleur de l'eau. Afin de justifier cette approche, il a été au préalable nécessaire de démontrer l'accord entre la distribution de probabilité de la simulation d'ensemble et la distribution statistique de ces données satellites. Je précise que l'utilisation de ce type de diagnostic est très peu répandue en océanographie. Il est donc essentiel de retenir que, dans cette thèse, **nous sommes parvenus à démontrer la possibilité de réaliser une simulation d'ensemble statistiquement en accord (fiable) avec les observations de la couleur de l'eau SeaWiFS.** Ce point constitue une étape importante pour pouvoir confronter des prévisions de modèles biogéochimiques avec les observations.

En complément, nous avons mis un point d'honneur à considérer la simulation d'ensemble sous la forme d'une distribution de probabilité. De la sorte, nous avons pu constater que l'information de chacun des membres de l'ensemble est nécessaire pour caractériser les distributions de chlorophylle pouvant être observées. **Chaque réalisation de la simulation d'ensemble affiche donc une représentation possible de l'état biogéochimique de l'océan.** Cet aspect est lui aussi particulièrement important. Il montre que l'approche stochastique que nous avons utilisée n'est pas uniquement un moyen pour générer de la dispersion. Elle permet de définir un ensemble d'états possibles (avec une probabilité d'existence) représentatif de différents comportements pouvant satisfaire à la dynamique des processus du modèle PISCES. Nous avons, par exemple, été capables de simuler plusieurs types de profils verticaux pouvant être perçus comme la signature de différentes espèces, non explicitement simulées par le modèle déterministe. C'est donc l'approche probabiliste, associée à une simulation stochastique d'incertitudes du modèle, qui révèle l'existence possible de comportements biogéochimiques différents.

Dans la dernière partie de chapitre, nous avons présenté une séquence d'analyse de la couleur de l'eau, réalisée à partir de l'erreur de prévision définie par les distributions de probabilité de l'ensemble a priori. Les résultats présentés dans ce manuscrit ont pu mettre en évidence le bon fonctionnement de la méthode d'analyse mise en place. Ils ont permis de montrer que la prise en compte de l'incertitude du modèle permettait de corriger la simulation d'ensemble (réduire l'incertitude), en conservant les caractéristiques dynamiques a priori définie par la simulation d'ensemble. Nous avons également pu vérifier que les analyses améliorent systématiquement les caractéristiques statistiques (fiabilité, résolution) de l'ensemble. **La prise en compte de statistiques d'erreur de prévision basées sur l'incertitude du modèle permet d'améliorer significativement les distributions de la chlorophylle.** Les résultats présentés ont également souligné l'importance de la définition initiale des paramétrisations stochastiques.

Réalisations techniques

1. Mise en place d'une plate-forme de modélisation couplée au 1/4° NEMO/PISCES sur l'océan Atlantique Nord à partir de la configuration physique NATL025 développée au sein du projet DRAKKAR.

↔ *Réalisation d'une simulation déterministe « online » de 6 ans entre le 1^{er} janvier 2005 et le 31 décembre 2010. L'état biogéochimique de l'océan, défini par les 24 variables du modèle PISCES, a été évalué tous les 5 jours.*

2. Mise en place d'une version probabiliste du modèle PISCES, établie à partir de paramétrisations stochastiques simulant l'incertitude sur les paramètres biogéochimiques du modèle et sur les échelles biogéochimiques non résolues (non-linéarité des équations). Le niveau d'incertitude peut être défini par l'utilisateur.

↔ *Réalisation d'une simulation d'ensemble de 60 membres sur l'année 2005. Comme pour la simulation déterministe, les résultats sont évalués tous les 5 jours pour les 24 variables biogéochimiques.*

3. Mise en place du schéma d'analyse d'ensemble multivariée de l'ETKF (Filtre d'Ensemble de Kalman Transformé) permettant de corriger les variables biogéochimiques à partir de données de la couleur de l'eau. Le schéma inclut une localisation des covariances d'erreur. Les non-gaussianités sont prises en compte par l'intermédiaire d'une transformation anamorphique.

↔ *Réalisation d'une séquence d'analyse d'ensemble multivariée tous les 5 jours de l'année 2005. Les observations SeaWiFS-HR au 1/24° sont utilisées pour corriger l'état prédit par le modèle.*

Principaux enseignements

issus du développement 1.

- La complexité biogéochimique du modèle PISCES a permis de réaliser une simulation déterministe dont l'estimation de la chlorophylle est nettement plus proche des observations que celle du modèle LOBSTER.
- L'analyse de la variabilité interannuelle, déduite des 6 années de simulation, n'explique que partiellement l'amplitude des écarts entre le modèle et les données satellites SeaWiFS dans l'océan Atlantique Nord.
- Les formulations des processus à l'origine de la production primaire dans le modèle PISCES présentent un niveau élevé d'incertitude. La prise en compte de 2 classes de phytoplancton apparaît insuffisante pour englober la diversité des comportements biogéochimiques de l'océan Atlantique Nord.

issus du développement 2.

- L'utilisation de paramétrisations stochastiques permet de simuler explicitement des sources d'incertitudes inhérentes à la formulation des processus biogéochimiques du modèle PISCES.
- L'introduction de processus aléatoires dans les équations du modèle impose de redéfinir le paradigme déterministe de modélisation. L'approche devient probabiliste. L'état du système doit être caractérisé par des distributions de probabilité, définies à partir des membres d'une simulation d'ensemble.
- La prise en compte d'incertitudes sur les formulations biogéochimiques du modèle modifie le niveau global d'incertitude du modèle → un réajustement des paramètres biogéochimiques du modèle, initialement calibrés pour une approche déterministe, a été nécessaire.
- Les paramétrisations stochastiques génèrent une importante dispersion sur la chlorophylle (en surface et dans la colonne d'eau), tout en permettant de conserver les grands traits biogéochimiques de la simulation déterministe.
- L'enveloppe statistique des membres de l'ensemble permet d'inclure la majorité des observations de la couleur de l'eau. Les distributions de probabilité de la simulation d'ensemble sont en accord statistique avec les observations : l'ensemble a priori est fiable.
- L'utilisation de paramétrisations stochastiques permet de simuler une variété de comportements biologiques, pouvant être associée à une diversité d'espèces, qui n'est pas explicitement représentée dans la version déterministe du modèle.

issus du développement 3.

- L'utilisation de statistiques d'erreur basées sur l'incertitude du modèle permet de significativement réduire les incertitudes associées aux états analysés.
- La correction de l'ensemble faite à partir des observations de la couleur de l'eau améliore les caractéristiques statistiques a priori de fiabilité et de résolution.
- Les performances des analyses sont en grande partie dépendantes de la définition initiale des paramétrisations stochastiques.

Je précise que les travaux de cette thèse vont être utilisés dans le cadre du projet AtlantOS¹, dont l'objectif est le développement de systèmes intégrés pour améliorer la compréhension des processus océaniques de l'océan Atlantique. J'ajoute qu'une publication, présentant la séquence d'analyse d'ensemble (cf. chapitre 7), est actuellement en préparation.

Recommandations

-À l'issue de cette thèse, nous pouvons dégager quelques "recommandations", pouvant être utiles aux travaux à venir dans le contexte de modélisation/assimilation biogéochimique.

- En ce qui concerne la modélisation, nous restons convaincus que l'approche probabiliste est un élément essentiel à la prévision biogéochimique opérationnelle.
- En ce qui concerne le développement des modèles, ajouter des classes de phytoplancton semble important pour espérer pouvoir rendre compte de la diversité des comportements biologiques.
- En ce qui concerne les réseaux d'observation, une vérification de grande échelle est indispensable pour vérifier la cohérence des structures verticales des modèles. Le développement des bouées BIO-ARGO devrait être un atout considérable. Toutefois, l'impact de la densité des observations sera un point à considérer. Pour l'assimilation de données, il faudra en effet s'assurer que l'introduction des profils verticaux puisse impacter la représentation de plus grande échelle du modèle.
- Que ce soit en modélisation ou en assimilation de données, il est crucial de tenir compte de l'incertitude des modèles.

8.2 Perspectives

Persistence de l'apport des données de la couleur de l'eau

Dans le dernier chapitre de cette thèse, nous avons présenté une séquence d'analyse 3D de la couleur de l'eau. Si la nécessité de cette expérience a déjà été évoquée à plusieurs reprises, une seconde étape, préliminaire à l'assimilation de données, devra toutefois être

1. Optimising and Enhancing the Integrated Atlantic Ocean Observing Systems

réalisée. En l'état actuel, la disponibilité des observations nous oblige en effet à répéter les cycles d'assimilation tous les 5 à 10 jours. Pour envisager une expérience d'assimilation de données, il faut pouvoir s'assurer que l'information apportée par les observations puisse rester en mémoire durant cette fenêtre temporelle. Il est donc indispensable de réaliser une (ou plusieurs) simulation d'ensemble, initialisée à partir de l'état corrigé par les observations de la couleur de l'eau. L'objectif sera d'examiner la persistance de l'information apportée par les observations, c'est-à-dire déterminer le temps au bout duquel l'incertitude de la prévision redevient équivalente à celle de la simulation a priori. En d'autres termes, il faudra répondre à la question : au bout de combien de temps la dispersion d'un ensemble initialisé par l'état analysé est égale à celle de l'ensemble a priori ? Si l'information se conserve moins longtemps que la durée envisagée d'une fenêtre d'assimilation, l'impact de l'assimilation de donnée serait insignifiant.

Ce point est de plus essentiel pour permettre le développement de réseaux d'observations adaptés aux besoins des modèles. Il est tout à fait possible que, dans le cas de la biogéochimie, l'incertitude des modèles soit telle que l'assimilation de données nécessite d'être considérée différemment. Dans ce cadre, l'apport des observations des satellites géostationnaires (e.g missions GOCI² et OCAPI³) pourrait être déterminant. Ces données permettent en effet l'accès à une information, certes moins globale, mais de résolution temporelle nettement plus élevée. Grâce à elles, il sera alors possible à l'assimilation de données de procéder par fenêtres temporelles beaucoup plus courtes (quelques heures). Les résultats énoncés dans ce manuscrit ne permettent pas de savoir si l'information apportée par les observations satellites sera conservée suffisamment longtemps pour envisager de répéter des cycles d'assimilation de données. Dans le cadre du projet ATLantOS, dans lequel les travaux initiés dans cette thèse seront prolongés, cette question sera vraisemblablement abordée.

Assimilation de données optiques

Tout au long de thèse, nous avons focalisé notre attention sur la représentation des erreurs de prévision. Elles jouent en effet un rôle clé en contrôlant la nature des corrections apportées lors de l'analyse. L'amplitude de la correction reste toutefois contrainte par le niveau d'incertitude des observations. Dans la cas de la couleur de l'océan, les algorithmes de traitement, nécessaires pour passer d'un signal de réflectance à un produit de chlorophylle, constituent une importante source d'incertitude. Une solution pour réduire l'impact des incertitudes des observations serait d'utiliser directement l'information des propriétés optiques mesurées par les capteurs satellites (Fujii et al., 2007). Des travaux récents ont d'ailleurs déjà permis de démontrer la pertinence de l'assimilation de données optiques sur l'estimation des caractéristiques biogéochimiques de l'océan (e.g Shulman et al., 2013; Ciavatta et al., 2014).

Reste que la majorité des modèles biogéochimiques ne propose qu'une formulation très simplifiée des propriétés optiques. Or, la réalisation d'expériences d'assimilation de données optiques nécessite d'inclure des caractéristiques telles que l'irradiance ou l'absorption de la lumière par le phytoplancton en tant que variables d'état du modèle. Entre autres pour répondre à ce besoin, la mise en place de modèles d'écosystème proposant une description

2. Geostationary Ocean Color Imager

3. Ocean Color Advanced Permanent Imager

appropriée des propriétés optiques est depuis peu en cours de développement (e.g Dutkiewicz et al., 2015). En outre, l'avantage principal d'une telle approche (dans le cadre de notre approche d'assimilation de données) serait de disposer d'une chaîne de traitement homogène, permettant de se concentrer uniquement la représentation de l'erreur modèle. Nul doute que ce type de démarche, combinée à une modélisation stochastique des incertitudes, permettrait d'aboutir à des résultats particulièrement prometteurs.

Probabiliste versus complexité

Dans le second chapitre de cette thèse, nous avons brièvement abordé la notion de "course à la complexité" qui existe aujourd'hui dans la modélisation du climat. Ainsi, la démarche généralement adoptée consiste à améliorer les prévisions en utilisant des modèles de plus en plus complexes, c'est-à-dire qui incluent de plus en plus de processus avec une résolution de plus en plus élevée. Toutefois, compte tenu du niveau d'incertitude des processus à représenter, utiliser une approche probabiliste comparable à celle présentée dans ce manuscrit paraîtrait tout autant justifié ; l'intérêt étant de considérer explicitement l'imperfection de notre connaissance d'un système et de pouvoir évaluer la qualité des prévisions par une meilleure quantification des incertitudes. L'approche probabiliste possède en plus l'avantage de pouvoir facilement être adaptable à différentes problématiques. Selon les besoins, ce qu'il est en définitive important de se demander est si, à coût informatique égal, une simulation déterministe, par exemple de haute résolution, permet d'apporter plus d'information qu'une simulation probabiliste de plus basse résolution.

Dans les faits, il me paraît compliqué de répondre à cette question de manière catégorique. J'imagine pourtant qu'après la lecture de ce manuscrit, le type d'approche que j'aurais tendance à préconiser ne doit pas vraiment prêter à confusion. Mais même si ma position est bien évidemment discutable, il n'en reste pas moins qu'encourager l'utilisation parallèle de différentes approches ne pourra que renforcer notre capacité à pouvoir prédire : « le climat qu'il fera demain ».

Bibliographie

- Abbott, B., Abbott, R., Abbott, T., Abernathy, M., Acernese, F., Ackley, K., Adams, C., Adams, T., Addesso, P., Adhikari, R., et al., 2016. Observation of gravitational waves from a binary black hole merger. *Physical Review Letters* vol.116.
- Alvain, S., Moulin, C., Dandonneau, Y., Bréon, F., 2005. Remote sensing of phytoplankton groups in case 1 waters from global seawifs imagery. *Deep Sea Research.I* vol.52.
- Anderson, J., 1996. A method for producing and evaluating probabilistic forecasts from ensemble model integrations. *Journal of Climate* vol.9, 1518–1530.
- Anderson, T., 2005. Plankton functional type modelling : running before we can walk ? *Journal of Plankton Research* vol.27, 1073–1081, doi :10.1093/plankt/fbi099.
- Antoine, D., 1998. Apports de la télédétection spatiale de la «couleur de l’océan» à l’océanographie. *Océanis* vol.24, 81–150.
- Arakawa, A., Lamb, V., 1980. A potential enstrophy and energy conserving scheme for the shallow water equations. *Monthly Weather Review* vol.109.
- Arhonditsis, G. B., Perhar, G., Zhang, W., Massos, E., Shi, M., Das, A., 2008. Addressing equifinality and uncertainty in eutrophication models. *Water Resources Research* vol.44 (1).
- Armstrong, R., Sarmiento, J., Slater, R., 1995. Monitoring ocean productivity by assimilating satellite chlorophyll into ecosystem models. In : Powell, T., Steele, J. (Eds.), *Ecological Time Series*. Springer US, pp. 371–390.
URL http://dx.doi.org/10.1007/978-1-4615-1769-6_17
- Asselin, R., 1980. Frequency filter for time integrations. *Monthly Weather Review* vol.100, 487–490.
- Aumont, O., Ethé, C., Tagliabue, A., Bopp, L., Gehlen, M., 2015. Pisces-v2 : An ocean biogeochemical model for carbon and ecosystem studies. *Geosciences Model Development. Discussion* vol.8, 1–135, doi :10.5194/gmdd-8-1-2015.
- Aumont, O., Maier-Reimer, E., Blain, S., Monfray, P., 2003. An ecosystem model of the global ocean including fe, si, p colimitations. *Global Biogeochemical Cycles* vol.17, doi :10.1029/2001GB001745.

- Barnier, B., Blaker, A., Biastoch, A., Böning, C. W., Coward, A., Deshayes, J., Ducez, A., Hirschi, J., Le Sommer, J., Madec, G., et al., 2015. Drakkar : developing high resolution ocean components for european earth system models. *CLIVAR Exchanges* vol.65, 18–21.
- Barnier, B., Madec, G., Penduff, T., Molines, J.-M., Treguier, A.-M., Le Sommer, J., Beckmann, A., Biastoch, A., Böning, C., Dengg, J., Derval, C., Durand, E., Gulev, S., Remy, E., Talandier, C., Theetten, S., Maltrud, M., Mc Clean, J., De Cuevas, B., 2006. Impact of partial steps and momentum advection schemes in a global ocean circulation model at eddy permitting resolution. *Ocean dynamics* vol.56, doi :10.1007/s10236-006-0082-1.
- Behrenfeld, M. J., Boss, E. S., 2014. Resurrecting the ecological underpinnings of ocean plankton blooms. *Annual Review of Marine Science* vol.6, 167–194.
- Berline, L., Brankart, J.-M., Brasseur, P., Ourmières, Y., Verron, J., 2007. Improving the physics of a coupled physical-biogeochemical model of the north atlantic through data assimilation : Impact on the ecosystem. *Journal of Marine Systems* vol.64, 153–172, doi :10.1016/j.jmarsys.2006.03.007.
- Berrisford, P., Källberg, P., Kobayashi, S., Dee, D., Uppala, S., Simmons, A., Poli, P., Sato, H., 2011. Atmospheric conservation properties in era-interim. *Quarterly Journal of the Royal Meteorological Society* vol.137, 1381–1399.
- Bishop, C. H., Etherton, B. J., Majumdar, S. J., 2001. Adaptive sampling with the ensemble transform kalman filter. part i : Theoretical aspects. *Monthly weather review* vol.129 (3), 420–436.
- Blanke, B., Delecluse, P., 1993. Variability of the tropical ocean simulated by a general circulation model with two different miwer-layer physics. *American Meteorological Society* vol.23, 1363–1388.
- Bopp, L., Aumont, O., Cadule, P., Alvain, S., Gehlen, M., 2005. Response of diatoms distribution to globalm warming and potential implications : A global model study. *Geophysical Research Letters* vol.32, doi :10.1029/2005GL023653.
- Bopp, L., Bowler, C., Guidi, L., Karsenti, E., Jeandel, C., 2015. L’océan, pompe à carbone. *Océan et climat fiches scientifiques*, 13–19.
URL www.ocean-climate.org
- Bourdallé-Badie, R., Treguier, A. M., 2006. A climatology of runoff for the global ocean-ice model orca025. Mercator-Ocean, Toulouse, France.
- Brankart, J., Candille, G., Garnier, F., Calone, C., Melet, A., Bouttier, P., Brasseur, P., Verron, J., 2015. A generic approach to explicit simulation of uncertainty in the nemo ocean model. *Geoscientific Model Development* vol.8 (5), 1285–1297, doi :10.5194/gmd-8-1285-2015.
- Brankart, J.-M., 2013. Impact of uncertainties in the horizontal density gradient upon low resolution global ocean modelling. *Ocean Modelling* vol.66, 64–76, doi :10.1016/j.ocemod.2013.02.004.

- Brankart, J.-M., Testut, C.-E., Béal, D., Doron, M., Fontana, C., Meinvielle, M., Brasseur, P., Verron, J., 2012. Towards an improved description of ocean uncertainties : effect of local anamorphic transformations on spatial correlations. *Ocean Science* 8 (2), 121–142.
- Brankart, J.-M., Ubelmann, C., Testut, C.-E., Cosme, E., Brasseur, P., Verron, J., 2009. Efficient parameterization of the observation error covariance matrix for square root or ensemble kalman filters : application to ocean altimetry. *Monthly Weather Review* vol.137 (6), 1908–1927.
- Brasseur, P., Gruber, N., Barciela, K., Brander, M., Doron, M., El Moussaoui, A., Hobday, A., Huret, M., Krémeur, A.-S., Lehodey, P., Matear, R., Moulin, C., Murtugudde, R., Senina, I., Svendsen, E., 2009. Integrating biogeochemistry and ecology into ocean data assimilation systems. *Oceanography* vol.22(3), 206–215, doi :10.5670/oceanog.2009.80.
- Brasseur, P., Verron, J., 2006. The seek filter method for data assimilation in oceanography : a synthesis. *Ocean Dynamics* vol.56, doi :10.1007/s10236-006-0080-3.
- Brentnall, S. J., Richards, K. J., Brindley, J., Murphy, E., 2003. Plankton patchiness and its effect on larger-scale productivity. *Journal of Plankton Research* 25, 121–140.
- Brier, G., 1950. Verification of forecasts expressed in terms of probability. *Monthly weather review* vol.78 (1), 1–3.
- Brown, T. A., 1974. Admissible scoring systems for continuous distributions.
- Buitenhuis, E. T., Geider, R. J., 2010. A model of phytoplankton acclimation to iron-light colimitation. *Limnology and Oceanography* vol.55, 714.
- Buizza, R., Miller, M., Palmer, T., 1999. Stochastic representation of model uncertainties in the ecmwf ensemble prediction system. *Quarterly Journal of the Royal Meteorological Society* vol.125, 2887–2908.
- Burgers, G., Jan van Leeuwen, P., Evensen, G., 1998. Analysis scheme in the ensemble kalman filter. *Monthly weather review* vol.126 (6), 1719–1724.
- Béal, D., Brasseur, P., Brankart, J.-M. AND, O. Y., Verron, J., 2010. Characterization of mixing errors in a coupled physical biogeochemical model of the north atlantic : implications for nonlinear estimation using gaussian anamorphosis. *Ocean Science* vol.6, 247–262.
- Campbell, J., 1995. The lognormal distribution as a model for bio-optical variability in the sea. *Journal of Geophysical research* vol.100, 13,237–13,254, doi :10.1029.
- Candille, G., Brankart, J.-M., Brasseur, P., 2015. Assessment of an ensemble system that assimilates jason-1/envisat altimeter data in a probabilistic model of the north atlantic ocean circulation. *Ocean Science* vol.11, 425–438, doi :10.5194/os-11-425-2015.
- Candille, G., Talagrand, O., 2005. Evaluation of probabilistic prediction systems for a scalar variable. *Quarterly Journal of the Royal Meteorological Society* vol.131, 2131–2150, doi :10.1256/qj.04.71.

- Ciavatta, S., Torres, R., Martinez-Vicente, V., Smyth, T., Dall'Olmo, G., Polimene, L., Allen, J. I., 2014. Assimilation of remotely-sensed optical properties to improve marine biogeochemistry modelling. *Progress in Oceanography* vol.127, 74–95.
- Ciavatta, S., Torres, R., Saux-Picart, S., Allen, J., 2011. Can ocean color assimilation improve biogeochemical hindcasts in shelf seas? *Journal of Geophysical Research* vol.116, doi :10.1029/2011JC007219.
- Debreu, L., Vouland, C., Blayo, E., 2008. Agrif :adaptive grid refinement in fortran. *Computers and Geosciences* vol.34, 8–13, doi :10.1016/j.cageo.2007.01.009.
- Dijkstra, H. A., Frankcombe, L., Von der Heydt, A., 2008. A stochastic dynamical systems view of the atlantic multidecadal oscillation. *Philosophical Transactions of the Royal Society of London A : Mathematical, Physical and Engineering Sciences* 366 (1875), 2543–2558.
- Doney, S., 1999. Major challenges confronting marine biogeochemical modeling. *Global Biogeochemical Cycles* vol.13, 705–714.
- Doron, M., Brasseur, P., Brankart, J.-M., 2011. Stochastic estimation of biogeochemical parameters of a 3d ocean coupled physical-biogeochemical model :twin experiments. *Journal of Marine Systems* vol.87, 194–207, doi :10.1016/j.jmarsys.2011.04.001.
- Doron, M., Brasseur, P., Brankart, J.-M., Losa, S. N., Melet, A., 2013. Stochastic estimation of biogeochemical parameters from globcolour ocean colour satellite data in a north atlantic 3d ocean coupled physical-biogeochemical model. *Journal of Marine Systems* vol.117-118, 81–85, doi :10.1016/j.jmarsys.2013.02.007.
- Dowd, M., 2011. Estimating parameters for a stochastic dynamic marine ecological system. *Environmetrics* vol.22, 501–515.
- Drange, H., 1996. A 3-dimensional isopycnic coordinate model of the seasonal cycling of carbon and nitrogen in the atlantic ocean. *Physics and Chemistry of the Earth* vol.21, 503–509, doi :10.1016/S0079-1946(97)81149-4.
- Droop, M., 1983. 25 years of algal growth kinetics, a personal view. *Botanica Marina* vol.26, 99–112.
- Dubos, T., Babiano, A., 2002. Two dimensional cascades and mixing : a physical space approach. *Journal of Fluid Mechanics* vol.467, 81–100, doi :10.1017/S0022112002001416.
- Dufour, C. O., Le Sommer, J., Gehlen, M., Orr, J. C., Molines, J.-M., Simeon, J., Barnier, B., 2013. Eddy compensation and controls of the enhanced sea-to-air co2 flux during positive phases of the southern annular mode. *Global Biogeochemical Cycles* vol.27, 950–961, doi :10.1002/gbc.20090,.
- Dutkiewicz, S., Follows, M., Marshall, J., Gregg, W. W., 2001. Interannual variability of phytoplankton abundances in the north atlantic. *Deep Sea Research Part II : Topical Studies in Oceanography* vol.48, 2323–2344.

- Dutkiewicz, S., Hickman, A., Jahn, O., Gregg, W., Mouw, C., Follows, M., 2015. Capturing optically important constituents and properties in a marine biogeochemical and ecosystem model. *Biogeosciences* vol.12, 4447–4481.
- Eppley, R., 1972. Temperature and phytoplakton growth in the sea. *Fishery Bulletin* vol.70, 1063–1085.
- Epstein, E., 1969. Stochastic dynamic prediction. *Tellus* vol.21, 739–759.
- Estrada, M., 1980. Phytoplankton biomass and production in the upwelling region of nw africa. relationship with hydrographic parameters. *Marine Biology* vol.60, 63–71.
- Evensen, G., 1994. Sequential data assimilation with a nonlinear quasi-geostrophic model using monte carlo methods to forecast error statistics. *Journal of Geophysical Research* vol.99, 10,143–10,162.
- Evensen, G., 2003. The ensemble kalman filter : Theoretical formulation and practical implementation. *Ocean dynamics* vol.53 (4), 343–367.
- Fasham, M., Ducklow, H., McKelvie, S., 1990. A nitrogen-based model of plankton dynamics in the oceanic mixed layer. *Journal of Marine Research* vol.48, 591–639.
- Fasham, M., Evans, G., Kiefer, D., Creasey, M., Leach, H., 1995. Testing a marine ecosystem model : sensitivity analysis and parameter optimization. *Philosophical Transactions of The Royal Society B Biological Sciences* vol.348, 203–209, doi :10.1098/rstb.1995.0062.
- Fennel, K., Losch, M., Schröter, J., Wenzel, M., 2001. Testing a marine ecosystem model : sensitivity analysis and parameter optimization. *Journal of Marine Systems* vol.28, 45–63.
- Fennel, K., Wilkin, J., Levin, J., Moisan, J., O'Reilly, J., Haidvogel, D., 2006. Nitrogen cycling in the middle atlantic bight : Results from a three-dimensional model and implications for the north atlantic nitrogen budget. *Global Biogeochemical Cycles* vol.20.
- Fichefet, T., Morales Maqueda, M., 1997. Sensivity of a global sea ice model to the treatment of ice thermodynamics and dynamics. *Journal of Geophysical Research* vol.102, doi :10.1029/97/JC00480.
- Flynn, K. J., Hipkin, C. R., 1999. Interactions between iron, light, ammonium, and nitrate : insights from the construction of a dynamic model of algal physiology. *Journal of Phycology* vol.35, 1171–1190.
- Follows, J., Dutkiewicz, S., 2011. Modeling diverse communities of marine microbes. *Annual Review of Marine Science* vol.3, 427–451, doi :10.1146/annurev-marine-120709-142848.
- Follows, M., Dutkiewicz, S., Grant, S., Chisholm, S., 2007. Emergent biogeography of microbial communities in a model ocean. *Science* vol.315, 1843–1846, doi :10.1126/science.1138544.
- Fontana, C., Brasseur, P., Brankart, J.-M., 2013. Toward a multivariate reanalysis of the north atlantic biogeochemistry during 1998–2006 based on the assimilation of seawifs chlorophyll data. *Ocean Science* vol.9, 37–56, doi :10.5194/os-9-1-2013.

- Ford, D., Edwards, K., Lea, D., Barciela, R., Martin, M., Demaria, J., 2012. Assimilating globcolour ocean colour data into a pre-operational physical biogeochemical model. *Ocean Science* vol.8, 751–771, doi :10.5194/os-8-751-2012.
- Frederiksen, J. S., O’Kane, T. J., Zidikheri, M. J., 2012. Stochastic subgrid parametrizations for atmospheric and oceanic flows. *Physica Scripta* vol.85.
- Friedrichs, M., 2001. Assimilation of jgofs eqpac and seawifs data into a marine ecosystem of the central equatorial pacific ocean. *Deep Sea Research Part II : Topical studies in Oceanography* vol.49, 289–319, doi :10.1016/S0967-0645(01)00104-7.
- Friedrichs, M. A., Dusenberry, J., Anderson, L., Armstrong, R. A., Chai, F., Christian, J., Doney, S., Dunne, J., Fujii, M., Hood, R., McGillicuddy Jr, D., Moore, J., Schartau, M., Sptiz, Y., Wiggert, J., 2007. Assessment of skill and portability in regional marine biogeochemical models : Role of multiple planktonic groups. *Journal of Geophysical Research* vol.112, doi :10.1029/2006JC003852.
- Fujii, M., Boss, E., Chai, F., 2007. The value of adding optics to ecosystem models : a case study. *Biogeosciences Discussions* 4 (3), 1585–1631.
- Garcia-Gorriz, E., Hoepffner, N., Ouberdous, M., 2003. Assimilation of seawifs data in a coupled physical-biological model of the adriatic sea. *Journal of Marine Systems* vol.40-41, 233–252, doi :10.1016/S0924-7963(03)00020-4.
- Garnier, F., Brankart, J.-M., Brasseur, P., Cosme, E., 2016. Stochastic parameterizations of biogeochemical uncertainties in a $1/4^\circ$ nemo/pisces model for probabilistic comparisons with ocean color data. *Journal of Marine Systems* vol.155, 59–72, doi :10.1016/j.jmarsys.2015.10.012.
- Gehlen, M., Bopp, L., Emprin, N., Aumont, O., Heinze, C., Ragueneau, O., 2006. Reconciling surface ocean productivity, export fluxes and sediment composition in a global biogeochemical ocean model. *Biogeosciences* vol.3, 521–537.
- Geider, R., MacIntyre, H., Kana, T., 1997. Dynamic model of phytoplankton growth and acclimation : responses of the balanced growth rate and the chlorophyll a : carbon ratio to light, nutrient-limitation and temperature. *Oceanographic Literature Review* vol.9, 974.
- Gneiting, T., Raftery, A. E., Westveld III, A. H., Goldman, T., 2005. Calibrated probabilistic forecasting using ensemble model output statistics and minimum crps estimation. *Monthly Weather Review* 133 (5), 1098–1118.
- Goodwin, P., Follows, M., Williams, R., 2008. Analytical relationship between atmospheric carbon dioxide, carbon emissions and ocean processes. *Global Biogeochemical Cycles* vol.22, doi :10.1029/2008GB003184.
- Gregg, W.W, F. M., Robinson, A., Rose, K., Schlitzer, R., Thompson, K., Doney, S., 2009. Skill assessment in ocean biological data assimilation. *Journal of Marine Systems* vol.76, 16–33, doi :10.1016/j.jmarsys.2008.05.006.

- Gregg, W., 2008. Assimilation of seawifs ocean chlorophyll data into a three-dimensional global ocean model. *Journal of Marine Systems* vol.69, 205–225, doi :10.1016/j.jmarsys.2006.02.015.
- Gregg, W., Casey, N., 2004. Global and regional evaluation of the seawifs chlorophyll data set. *Remote Sensing of Environment* vol.93, 463–479, doi :org/10.1016/j.rse.2003.12.012.
- Grodsky, S. A., Carton, J. A., McClain, C. R., 2008. Variability of upwelling and chlorophyll in the equatorial atlantic. *Geophysical Research Letters* 35.
- Hamill, T., 2001. Interpretation of rank histograms for verifying ensemble forecasts. *Monthly Weather Review* vol.129, 550–560.
- Hasselmann, K., 1976. Stochastic climate models part i. theory. *Tellus* vol.28 (6), 473–485. URL <http://dx.doi.org/10.1111/j.2153-3490.1976.tb00696.x>
- Hemmings, J., Srokosz, M., Challenor, P., Fasham, M., 2003. Assimilating satellite ocean-colour observations into oceanic ecosystem models. *Philosophical Transactions of the Royal Society of London A : Mathematical, Physical and Engineering Sciences* vol.361, 33–39.
- Hemmings, J., Srokosz, M., Challenor, P., Fasham, M., 2004. Split-domain calibration of an ecosystem model using satellite ocean colour data. *Journal of Marine Systems* vol.50, 141–179. URL <http://eprints.soton.ac.uk/217/>
- Henson, S. A., Dunne, J. P., Sarmiento, J. L., 2009. Decadal variability in north atlantic phytoplankton blooms. *Journal of Geophysical Research : Oceans* (1978–2012) 114.
- Hersbach, H., 2000. Decomposition of the continuous ranked probability score for ensemble prediction systems. *Weather and Forecasting* vol.15 (5), 559–570.
- Hitchcock, G. L., Mariano, A. J., Rossby, T., 1993. Mesoscale pigment fields in the gulf stream : observations in a meander crest and trough. *Journal of Geophysical Research : Oceans* (1978–2012) vol.98, 8425–8445.
- Hoteit, L., Triantafyllou, G., Petihakis, G., 2004. Towards a data assimilation system for the cretan sea ecosystem using a simplified kalman filter. *Journal of marine Systems* vol.45, 159–171, doi :10.1016/j.jmarsys.2003.11.003.
- Hoteit, L., Triantafyllou, G., Petihakis, G., Allen, J., 2003. A singular evolutive extended kalman filter to assimilate real in-situ data in a 1d marine ecosystem. *Annales Geophysicae* vol.21, 389–397.
- Hu, J., Fennel, K., Mattern, J., Wilkin, J., 2012. Data assimilation with a local ensemble kalman filter applied to a three-dimensional biological model of the middle atlantic bight. *Journal of Marine Systems* vol.94, 145–156, doi :10.1016/j.jmarsys.2011.11.016.
- Ishizaka, J., 1990. Coupling of coastal zone color scanner data due to a physical biological model of the southeastern us continental shelf ecosystem :1. czcs data description and lagrangian particle experiments. *Journal of Geophysical Research* vol.95, 20167–20181, doi :10.1029/JC095iC11p20167.

- Juricke, S., Jung, T., 2014. Influence of stochastic sea ice parametrization on climate and the role of atmosphere–sea ice–ocean interaction. *Philosophical Transactions of the Royal Society of London A : Mathematical, Physical and Engineering Sciences* vol.372.
- Juricke, S., Lemke, P., Timmermann, R., Rackow, T., 2013. Effects of stochastic ice strength perturbation on arctic finite element sea ice modeling. *Journal of Climate* vol.26, 3785–3802.
- Kahru, M., Brotas, V., MANZANO-SARABIA, M., Mitchell, B., 2011. Are phytoplankton blooms occurring earlier in the arctic? *Global Change Biology* 17, 1733–1739.
- Kalman, R., 1960. A new approach to linear filtering and prediction problems. *Transaction of the American Society of Mechanical Engineers-Journal of Basic Engineering* vol.82 (Series D), 35–45.
- Killworth, P. D., Cipollini, P., Uz, B. M., Blundell, J. R., 2004. Physical and biological mechanisms for planetary waves observed in satellite-derived chlorophyll. *Journal of Geophysical Research* vol.109.
- Kleeman, R., 2008. Stochastic theories for the irregularity of ENSO. *Philosophical Transactions of the Royal Society of London A : Mathematical, Physical and Engineering Sciences* 366 (1875), 2509–2524.
- Klein, P., Lapeyre, G., 2009. The oceanic vertical pump induced by mesoscale and sub-mesoscale turbulence. *Annual Review of Marine Science* vol.1 :3, 51–75, doi :10.1146/annurev.marine.010908.163704.
- Kutzbach, J. E., 1967. Empirical eigenvectors of sea level pressure, surface temperature and precipitation complexes over north america. *American Meteorological Society*, 791–802.
- Kwiatkowski, L., Yool, A., Allen, J., Anderson, T., Barciela, R., Buitenhuis, E., Butenschön, M., Enright, C., Halloran, P., Le Quéré, C., De Mora, L., Racault, M.-F., Sinha, B., Totterdell, I., Cox, P., 2014. imarnet : an ocean biogeochemistry model intercomparison project within a common physical ocean modelling framework. *Biogosciences* vol.11, 7291–7304, doi :10.5194/bg-11-7291-2014.
- Lacour, L., Claustre, H., Prieur, L., D’Ortenzio, F., 2015. Phytoplankton biomass cycles in the north atlantic subpolar gyre : A similar mechanism for two different blooms in the labrador sea. *Geophysical Research Letters* 42, 5403–5410.
- Large, W. G., Yeager, S. G., 2004. Diurnal to decadal global forcing for ocean and sea-ice models : the data sets and flux climatologies. *Citeseer*.
- Legal, C., Klein, P., Treguier, A.-M., Paillet, J., 2006. Diagnosis of the vertical motions in a mesoscale stirring region. *Journal of Geophysical Oceanography* vol.37, 1413–1424, doi :10.1175/JPO3053.1.
- Lehodey, P., Senina, I., Sibert, J., Hampton, J., 2008. Seapodym.v2 : A spatial ecosystem and population dynamics model with parameter optimization providing a new tool for tuna management. *Tech. rep., wCPFC-SC4-2008/EB-WP-10*.

- Leith, C., 1974. Theoretical skill of monte-carlo forecasts. *Monthly Weather Review* vol.102, 409–418.
- LeVeque, R. J., 1978. Finite volume methods for hyperbolic problems. *METODOLOGÍA DOCENTE*.
- LeVeque, R. J., 2002. Finite volume methods for hyperbolic problems. Vol. vol.31. Cambridge university press.
- Levin, S. A., Segel, L. A., 1976. Hypothesis for origin of planktonic patchiness.
- Levitus, S., Boyer, T. P., Conkright, M. E., O'Brien, T., Antonov, J., Stephen, C., Stathopoulos, L., Johnson, D., Gelfeld, R., 1998. Noaa atlas nesdis 18.
- Li, W., Harrison, W., 2001. Chlorophyll, bacteria and picophytoplankton in ecological provinces of the north atlantic. *Deep Sea Research Part II : Topical Studies in Oceanography* vol.48, 2271–2293.
- Lima, I. D., Doney, S. C., 2004. A three-dimensional, multnutrient, and size-structured ecosystem model for the north atlantic. *Global Biogeochemical Cycles* vol.18.
- Llort, J., 2015. Bloom phenology, mechanisms and future change in the southern ocean. Ph.D. thesis, Université Pierre et Marie Curie-Paris 6.
- Longhurst, A., 1993. Seasonal cooling and blooming in tropical oceans. *Deep Sea Research Part I : Oceanographic Research Papers* 40, 2145–2165.
- Longhurst, A., 1995. Seasonal cycles of pelagic production and consumption. *Progress in Oceanography* vol.36, doi :10.1016/0079-6611(95)00015-1.
- Lorenc, A., 1986. Analysis methods for numerical weather prediction. *Quartely Journal of the Royal Meteorological Society* vol.112, 1177–1194, doi :10.1002/qj.49711247414.
- Lorenzo, M., Taboada, J., Iglesias, I., 2012. The role of stochastic forcing in climate models : The case of thermohaline circulation. *CLIMATE MODELS*, 231.
- Losa, S., Kivman, G., Schröter, J., Wenzel, M., 2004. Weak constraint parameter estimation for a simple ocean ecosystem model : what can we learn about the model and the data ? *Journal of Marine Systems* vol.45, 1–20, doi :10.1016/j.jmarsys.2003.08.005.
- Ludwig, W., Probst, J.-L., Kempe, S., 1996. Predicting the oceanic input of organic carbon by continental erosion. *Global Biogeochemical Cycles* vol.10, 23–41.
- Lévy, M., Bopp, L., 2007. Turbulences dans l'océan 414.
- Lévy, M., Gavart, L., Mémerly, L., Caniaux, G., Paci, A., 2005. A four-dimensional mesoscale map of the spring bloom in the northeast atlantic (pomme experiment) : Results of a pronostic model. *Journal of Geophysical research*.
- Lévy, M., Kremer, A.-S., Memery, L., 2004. Description of the LOBSTER biogeochemical model implemented in the OPA system. Laboratoire d'Océanographie Dynamique et de Climatologie (LODYC), Institut Pierre Simon Laplace (IPSL), technical note.

- Lévy, M., Lovino, D., Resplandy, L., Klein, P., Madec, G., Tréguier, A.-M., Masson, S., Takahashi, K., 2011. Large scale impact of submesoscale dynamics on phytoplankton : Local and remote effects. *Ocean Modelling* vol.43-44, 77–93, doi :10.1016/j.ocemod.2011.12.003.
- Lévy, M., Martin, A., 2013. The influence of mesoscale and submesoscale heterogeneity on ocean biogeochemical reactions. *Global Biogeochemical cycles* vol.27, 1–12, doi :10.1002/2012GB004518.
- Lévy, M., Mémery, L., Madec, G., 1998. The onset of a bloom after deep winter convection in the northwestern mediterranean sea : mesoscale process study with a primitive equation model. *Journal of Marine Systems* vol.16, 7–21.
- Madec, G., , the NEMO team, 2012. NEMO ocean engine. version 3.4 Edition.
URL <http://www.nemo-ocean.eu>
- Madec, G., Delecluse, P., Imbard, M., Lévy, M., 1998. Opa 8 ocean general circulation model. reference manuel.tech.rep, lodyc/ipsl note 11 Edition.
- Magri, S., Brasseur, P., Lacroix, G., 2005. Data assimilation in a marine ecosystem coupled to a mixed layer model of the upper ocean. *C.R. Géosciences* 337, 1065–1074.
URL <https://hal.archives-ouvertes.fr/hal-00212270>
- Mahowald, N., Jickells, T. D., Baker, A. R., Artaxo, P., Benitez-Nelson, C. R., Bergametti, G., Bond, T. C., Chen, Y., Cohen, D. D., Herut, B., et al., 2008. Global distribution of atmospheric phosphorus sources, concentrations and deposition rates, and anthropogenic impacts. *Global Biogeochemical Cycles* vol.22.
- Maier-Reimer, E., Mikolajewicz, U., Winguth, A., 1996. Future ocean uptake of co2 : interaction between ocean circulation and biology. *Climate Dynamics* vol.12 (10), 711–722.
- Maranon, E., Holligan, P., Varela, M., Mourino, B., Bale, A., 1999. Basin scale variability of phytoplankton biomass, production and growth in the atlantic ocean. *Deep-Sea Research PART I* vol.47, 825–857.
- Martin, J. H., Gordon, R. M., Fitzwater, S. E., 1990. Iron in antarctic waters.
- Mattern, J., Edwards, C., Fiechter, J., Song, H., Moore, A., 2015. Comparing the optimization and data assimilation capabilities of ocean biogeochemical ocean models of different complexities. MEOPAR/MEAP-TT workshop.
- Mattern, J. P., Dowd, M., Fennel, K., 2013. Particle filter-based data assimilation for a three-dimensional biological ocean model and satellite observations. *Journal of Geophysical Research : Oceans* 118 (5), 2746–2760.
- Mattern, J. P., Prestin, J., Langemann, D., Fennel, K., Dowd, M., 2008. Ensemble-based data assimilation for a physical–biological ocean model near bermuda. Master’s thesis, Universität zu Lübeck.

- Mayorga, E., Seitzinger, S. P., Harrison, J. A., Dumont, E., Beusen, A. H., Bouwman, A., Fekete, B. M., Kroeze, C., Van Drecht, G., 2010. Global nutrient export from watersheds 2 (news 2) : model development and implementation. *Environmental Modelling & Software* vol.25, 837–853.
- McGillicuddy, D. J., Anderson, L. A., Bates, N. R., Bibby, T., Buesseler, K. O., Carlson, C. A., Davis, C. S., Ewart, C., Falkowski, P. G., Goldthwait, S. A., et al., 2007. Eddy/wind interactions stimulate extraordinary mid-ocean plankton blooms. *Science* 316, 1021–1026.
- Mollo, P., Noury, A., 2013. *Le manuel du plancton*, editions charles léopold mayer Edition. ISBN :978-2-84377-173-6.
- Molteni, F., Buizza, R., Palmer, T., Petroliagis, T., 1996. The ecmwf ensemble prediction system : methodology and validation. *Quarterly Journal of the Royal Meteorological Society* vol.122, 73–119, doi :10.1002/qj.49712252905.
- Monahan, A. H., Alexander, J., Weaver, A. J., 2008. Stochastic models of the meridional overturning circulation : time scales and patterns of variability. *Philosophical Transactions of the Royal Society of London A : Mathematical, Physical and Engineering Sciences* vol.366, 2525–2542.
- Monod, J., 1942. *Recherches sur la croissance des cultures bactériennes*. Hermann Mayenne.
- Moore, J. K., Doney, S. C., Kleypas, J. A., Glover, D. M., Fung, I. Y., 2001. An intermediate complexity marine ecosystem model for the global domain. *Deep Sea Research Part II : Topical Studies in Oceanography* vol.49, 403–462.
- Moore, T. S., Campbell, J. W., Dowell, M. D., 2009. A class-based approach to characterizing and mapping the uncertainty of the modis ocean chlorophyll product. *Remote Sensing of Environment* vol.113, 2424–2430.
- Morel, A., Berthon, J., 1989. Surface pigments, algal biomass profiles, and potential production of the euphotic layer. relationship reinvestigated in view of remote sensing applications. *Limnology and Oceanography* vol.34, 1545–1562.
- Morel, A., Prieur, L., 1977. Analysis of variations in ocean color. *Limnology and oceanography* vol.22 (4), 709–722.
- Morel, J., 1988. Optical modeling of the upper ocean in relation to its biogenous matter content. *Journal of Geophysical Research* vol.93, 10,749–10,768.
- Murphy, J. M., Sexton, D. M., Barnett, D. N., Jones, G. S., Webb, M. J., Collins, M., Stainforth, D. A., 2004. Quantification of modelling uncertainties in a large ensemble of climate change simulations. *Nature* vol.430 (7001), 768–772.
- Najjar, R., Jin, X., Louanchi, F., Aumont, O., Caldeira, K., Doney, S., Dutay, J.-C., Follows, M., Gruber, N., Joos, F., Lindsay, K., Maier-Reimer, E., Matear, R., Matsumoto, K., Monfray, P., Mouchet, A., Orr, J., Plattner, G.-K., Sarmiento, J., Schlitzer, R., Slater, R., Weirig, M.-F., Yamanaka, Y., Yool, A., 2007. Impact of circulation on export production,

- dissolved organic matter, and dissolved oxygen in the ocean : Results from phase ii of the ocean carbon-cycle model intercomparison project (ocmip-2). *Global Biogeochemical Cycles* vol.21, doi :10.1029/2006GB002857.
- Natvik, L.-J., Evensen, G., 2003a. Assimilation of ocean colour data into a biogeochemical model of the north atlantic. part 1. data assimilation experiments. *Journal of marine Systems* vol.40-41, 127–153, doi :10.1016/S0924-7963(03)00016-2.
- Natvik, L.-J., Evensen, G., 2003b. Assimilation of ocean colour data into a biogeochemical model of the north atlantic. part 2. statistical analysis. *Journal of marine Systems* vol.40-41, 155–169, doi :10.1016/S0924-7963(03)00017-4.
- Nerger, L., Gregg, W., 2007. Assimilation of seawifs data into a global ocean-biogeochemical model using a local seik filter. *Journal of Marine Systems* vol.68 (1), 237–254.
- Nerger, L., Gregg, W., 2008. Improving assimilation of seawifs data by the application of bias correction with a local seik filter. *Journal of Marine Systems* vol.73, 87–102, doi :10.1016/j.jmarsys.2007.09.007.
- Nerger, L., Hiller, W., Jens, S., 2005. A comparison of error subspace kalman filter. *Tellus* vol.57A, 715–735.
- Oschlies, A., Garçon, V., 1998a. Eddy-induced enhancement of primary production in a model of the north atlantic. *Nature* vol.394, 266–269.
- Oschlies, A., Garçon, V., 1998b. Eddy-induced enhancements of primary production in a model of the north atlantic ocean. *Nature* vol.394, 266–269, doi :10.1038/28373.
- Oschlies, A., Schartau, M., 2005. Basin-scale performance of a locally optimized marine ecosystem model. *Journal of Marine Research* vol.63, 335–358, doi :10.1098/rstb.1995.0062.
- O'Reilly, J. E., Maritorena, S., Siegel, D. A., O'Brien, M. C., Toole, D., Mitchell, B. G., Kahru, M., Chavez, F. P., Strutton, P., Cota, G. F., et al., 2000. Ocean color chlorophyll a algorithms for seawifs, oc2, and oc4 : Version 4. *SeaWiFS postlaunch calibration and validation analyses, Part* vol.3, 9–23.
- Paci, A., Caniaux, G., Giordani, H., Lévy, M., Prieur, L., Reverdin, G., 2007. A high resolution simulation of the ocean during the pomme experiment : Mesoscale variability and near surface processes. *Journal of Geophysical Research* vol.112, doi :10.1029/2005JC003389.
- Palmer, T., 2012. Towards the probabilistic earth-system simulator : a vision for the future of climate and weather prediction. *Quarterly Journal of the Royal Meteorological Society* vol.138, 841–861, doi :10.1002/qj.1923.
- Palmer, T., Molteni, F., Mureau, R., Buizza, R., Chapelet, P., Tribbia, J., 1993. Ensemble prediction. *Vol.* vol.1. pp. 21–66.
- Palmer, T., Shutts, G., Hagedorn, R., Doblas-Reyes, F., Jung, T., Leutbecher, M., 2005. Representing model uncertainty in weather and climate prediction. *Annual Review of Earth and Planetary Sciences* vol.33, 163–193, doi :10.1146/annurev.earth.33.092203.122552.

- Palmer, T. N., 2001. A nonlinear dynamical perspective on model error : A proposal for non-local stochastic-dynamic parametrization in weather and climate prediction models. *Quarterly Journal of the Royal Meteorological Society* vol.127, 279–304.
- Pham, D., Verron, J., Roubaud, M., 1998a. A singular evolutive extended kalman filter for data assimilation in oceanography. *Journal of Marine Systems* vol.16, 323–340.
- Pérez, V., Fernandez, E., Maranon, E., Moran, X., Zubkov, M., 2006. Vertical distribution of phytoplankton biomass, production and growth in the atlantic subtropical gyres. *Deep-Sea Research I* vol.53, 1616–1634.
- Raick, C., Neukermans, G., Vanhellemont, Q., Jolivet, D., 2006. Model complexity and performance : How far can we simplify? *Progress in Oceanography* vol.70, 27–57.
- Redfield, A., 1934. On the proportions of organic derivatives in sea water and their relation to the composition of plankton. *James Johnstone Memorial Volume*, University Press, Liverpool, 177–192.
- Redfield, A., Ketchum, B., Richards, F., 1963. The influence of organisms on the composition of seawater. *The sea*, 26–77.
- Redfield, A. C., 1958. The biological control of chemical factors in the environment. *American scientist*, 230A–221.
- Resplandy, L., Lévy, M., Bopp, L., Echevin, V., Pous, S., V.V.S.S, S., Kumar, D., 2012. Controlling factors of the oxygen balance in the arabian sea's omz. *Biogeosciences* vol.9, 5095–5109, doi :10.5194/bg-9-5095-2012.
- Ridgwell, A., Schmidt, D. N., 2010. Past constraints on the vulnerability of marine calcifiers to massive carbon dioxide release. *Nature Geoscience* 3 (3), 196–200.
- Robinson, I. S., 2004. *Measuring the oceans from space : the principles and methods of satellite oceanography*. Springer Science & Business Media.
- Rodgers, K., Aumont, O., Menkes, C., T, G., 2008. Decadal variations in equatorial pacific ecosystems and ferrocline/pycnocline decoupling. *Global biogeochemical cycles* vol.22, doi :10.1029/2006GB002919.
- Rousset-Regimbeau, F., 2007. *Modélisation des bilans de surface et des débits sur la france, application à la prévision d'ensemble des débits*. Ph.D. thesis, Université Paul Sabatier, Toulouse, France.
- Sabine, C., Feely, R., Gruber, N., Key, R., Lee, K., Bullister, J., Wanninkhof, R., Wong, C., Wallace, D., Tilbrook, B., Millero, F., Peng, T.-H., Kozyr, A., Ono, T., Rios, A., 2004. The oceanic sink for anthropogenic co₂. *Science* vol.16, 367–371, doi :10.1126/science.1097403.
- Saetra, O., Hersbach, H., Bidlot, J., Richardson, D., 2004. Effects of observation errors on statistics for ensemble spread and reliability. *Monthly Weather Review* vol.132, 1487–1501.

- Sarmiento, J., Bender, M., 1994. Carbon biogeochemistry and climate change. *Photosynthesis Research* 39 (3), 209–234.
URL <http://dx.doi.org/10.1007/BF00014585>
- Sarmiento, J. L., 2013. *Ocean biogeochemical dynamics*. Princeton University Press.
- Sathyendranath, S., Longhurst, A., Caverhill, C. M., Platt, T., 1995. Regionally and seasonally differentiated primary production in the north atlantic. *Deep Sea Research Part I : Oceanographic Research Papers* vol.42, 1773–1802.
- Sauzede, R., Claustre, H., Jamet, C., Uitz, J., Ras, J., Mignot, A., D’Ortenzio, F., 2015. Retrieving the vertical distribution of chlorophyll a concentration and phytoplankton community composition from in situ fluorescence profiles : A method based on a neural network with potential for global-scale applications. *Journal of Geophysical Research : Oceans* vol.120, 451–470, doi :10.1002/2014JC010355.
- Schartau, M., Oschlies, A., Willebrand, 1998. Data assimilation and a pelagic ecosystem model : parameterization using time series observations. *Journal of marine Systems* vol.16, 51–68.
- Schmitz, W. J., McCartney, M. S., 1993. On the north atlantic circulation. *Reviews of Geophysics* vol.31, 29–49.
- Semovski, S., Wozniak, B., 1995. Model of the annual phytoplankton cycle in the marine ecosystem-assimilation of monthly satellite chlorophyll data for the north atlantic and baltic. *Oceanologia* vol.37, 3–31, pL ISSN 0078-3234.
- Shulman, I., Frolov, S., Anderson, S., Penta, B., Gould, R., Sakalaukus, P., Ladner, S., 2013. Impact of bio-optical data assimilation on short-term coupled physical, bio-optical model predictions. *Journal of Geophysical Research : Oceans* vol.118, 2215–2230.
- Sigman, D. M., Boyle, E. A., 2000. Glacial/interglacial variations in atmospheric carbon dioxide. *Nature* 407 (6806), 859–869.
- Simon, E., Bertino, L., 2009. Application of the gaussian anamorphosis to assimilation in a 3d coupled physical ecosystem of the north atlantic with the enkf : a twin experiment. *Ocean Science* vol.23, 3171–3185.
- Simon, E., Bertino, L., 2012. Gaussian anamorphosis extension of the denkf for combined state and parameter estimation : application to a 1d ocean ecosystem model. *Journal of Marine Systems* vol.89, 1–18.
- Six, K., Maier-Reimer, E., 1996. Effects of plankton dynamics on seasonal carbon fluxes in an ocean general circulation model. *Global Biogeochemical Cycles* vol.10, 559–583, doi :10.1029/96GB02561.
- Steinacher, M., Joos, F., Frolicher, T., Bopp, L., Cadule, P., Cocco, V., Doney, S., Gehlen, M., Lindsay, K., Moore, J., Schneider, B., Segschneider, J., 2010. Projected 21st century decrease in marine productivity : a multi-model analysis. *Biogeosciences* vol.7, 979–1005, doi :10.5194/bg-7-979-2010.

- Takashi, T., Broecker, W., Langer, S., 1985. Redfield ratio based on chemical data from isopycnal surfaces. *Journal of Geophysical research* vol.90, 6907–6924, doi :10.1029/JC090iC04p06907.
- Tarantola, A., 1987. *Inverse Problem Theory : Methods for data fitting and model parameter estimation*, first edition Edition. Elsevier, iSBN 0-444-42765-1.
- Tegen, I., Fung, I., 1995. Contribution to the atmospheric mineral aerosol load from land surface modification. *Journal of Geophysical Research : Atmospheres* (1984–2012) 100 (D9), 18707–18726.
- Tomczak, M., Godfrey, J. S., 2013. *Regional oceanography : an introduction*. Elsevier.
- Toth, Z., Kalnay, E., 1993. Ensemble forecasting at nmc : the generation of perturbations. *Bulletin of the American Meteorological Society* vol.74, 2317–2330.
- Triantafyllou, G., Hoteit, L., Petihakis, G., Banks, A., 2007. Assimilation of ocean colour data into a biogeochemical flux model of the eastern mediterranean sea. *Ocean Science* vol.3, 397–410.
- Uitz, J., Claustre, H., Morel, A. S. B., 2006. Vertical distribution of phytoplankton communities in open ocean : An assesment based on surface chlorophyll. *Journal of Geophysical research* vol.111, doi :10.1029.
- Uitz, J., Huot, Y., Bruyant, F., Babin, M., Claustre, H., 2008. Relating phytoplankton photophysiological properties to community structure on large scales. *Limnology and Oceanography* 53 (2), 614–630.
- Valcke, S., 2006. *Oasis3 user guide (prism 2-5)*. Tech.rep, prism support initiative report no3, cerfacs, toulouse, france Edition.
- Vidard, A., Bouttier, P.-A., Vigilant, F., 2015. *Nemotam : tangent and adjoint models for the ocean modelling platform nemo*. *Geoscientific Model Development* vol.8, doi :10.5194/gmd-8-1245-2015.
- Voituriez, B., Herbland, A., Leborgne, R., 1982. L'upwelling équatorial de l'atlantique est pendant l'expérience météorologique mondiale (pemg). *Oceanologica Acta* 5 (3), 301–314.
- Vollenweider, R.-A., Marchetti, R., Viviani, R., 1992. *Marine coastal eutrophisation*. Elsevier Science, iSBN : 978-0-444-89990-3.
- Ward, B., Schartau, M., Oschlies, A., Martin, A., Follows, M., Anderson, T., 2013. When is a biogeochemical model too complex? objective model reduction and selection for north atlantic time-series sites. *Progress in Oceanography* vol.116, doi :10.1016/j.pocean.2013.06.002.
- Weir, B., Miller, R. N., Spitz, Y. H., 2013. Implicit estimation of ecological model parameters. *Bulletin of mathematical biology* vol.75, 223–257.
- Wilks, D., 2006. *Statistical Methods in the Atmospheric Sciences*, 2nd Edition. Vol. vol.91. Academic Press.

- Williams, P. D., 2012. Climatic impacts of stochastic fluctuations in air–sea fluxes. *Geophysical Research Letters* 39 (10).
- Williams, R. G., 2011. Ocean eddies and plankton blooms. *Nature Geoscience* vol.4, 739–740.
- Wolf, K., Woods, J., 1988. Lagrangian simulation of primary production in the physical environment—the deep chlorophyll maximum and nutricline. *Toward a theory on Biological-Physical Interactions in the World Ocean*, 51–70Doi :10.1016/j.jmarsys.2010.02.015.
- Woods, J., Onken, R., 1982. Diurnal variation and primary production in the ocean preliminary results of a lagrangian ensemble model. *Journal of Plankton Research*, 735–756Doi :10.1093/plankt/4.3.735.
- Xiao, Y., Friedrichs, M. A., 2014. The assimilation of satellite-derived data into a one-dimensional lower trophic level marine ecosystem model. *Journal of Geophysical Research : Oceans* 119 (4), 2691–2712.
- Yoder, J. A., Doney, S. C., Siegel, D. A., Wilson, C., 2010. Study of marine ecosystems and biogeochemistry now and in the future : examples of the unique contributions from space vol.23, 104–117.

Table des figures

1.1	Représentation simplifiée de la séquestration du carbone atmosphérique par l’océan. Source : www.ocean-climate.org	8
2.1	a) Images satellites présentant des structures en filaments de la chlorophylle de surface (source http://earthobservatory.nasa.gov). a) Bloom de coccolithophoridés observé au large du Groënland avec le capteur MODIS. b) Bloom de phytoplancton dans le Gulf Stream observé le 14 Mai 2015 au large de la Nouvelle Ecosse avec le capteur VIIRS.	22
2.2	Principe d’une simulation d’ensemble tel que présenté dans le blog de Steve Easterbrook ”Tracking down the uncertainties in weather and climate prediction“. Source : http://www.easterbrook.ca/steve/?s=stochastic	24
3.1	Images satellites au 1/12° de la chlorophylle de surface de l’Atlantique Nord obtenues à partir des observations MODIS-Aqua (gauche), MERIS-ENVISAT (centre) et SeaWiFS-SeaStar (droite). Ces observations sont présentées en moyenne de 5 jours pour le 15 mai 2005.	37
3.2	Images satellites au 1/24° de la chlorophylle de surface de l’Atlantique Nord obtenues à partir des données d’observations SEAWIFS-HR. Ces observations sont présentées en moyenne de 5 jours pour le 15 mai 2005.	38
3.3	(a) Schéma de la grille Arakawa C (Arakawa and Lamb, 1980). (b) Profondeur des 46 niveaux verticaux.	43
3.4	Représentation schématique des 4 principaux compartiments de PISCES. À l’exception du carbone inorganique, de l’alcalinité et de l’oxygène, toutes les variables du modèle sont indiquées. Pour conserver une représentation simple, seules les interactions avec le compartiment PHY sont représentées. Quelques-uns des paramètres clés sont également indiqués.	46
3.5	(a) diatomées. (b) synecochoccus (espèce de nanophytoplancton) source John B Waterbury http://genome.jgi.doe.gov , (c) microzooplancton nauplii (d) mesozooplancton calaninus finimarchius source http://arthropoda.southernfriedscience.com	47
3.6	Cartes de surface de la simulation déterministe <i>detIni</i> , présentée dans le chapitre suivant (ligne du haut) et de la simulation Mercator BIOMER, utilisée pour initialiser les variables de PISCES (ligne du bas). Les cartes affichent les moyennes annuelles de l’année 2005 des concentrations en microzooplancton (1 ^{ère} colonne), nitrates (2 ^{ème} colonne) et phosphates (3 ^{ème} colonne).	55

4.1	a) Principaux courants océaniques de l'océan Atlantique Nord (source : Tomczak and Godfrey (2013)). b) Découpage de l'océan Atlantique Nord en provinces biogéochimiques tel que proposé par Sathyendranath et al. (1995).	63
4.2	Cartes de surface des moyennes mensuelles de la chlorophylle des données SeaWIFS (1 ^{ère} colonne), du modèle LOBSTER (simulation <i>detLob</i> , 2 ^{ème} colonne) et du modèle PISCES (simulation <i>detIni</i> , 3 ^{ème} colonne). En partant de la ligne du haut, les résultats sont présentés pour successivement les mois de janvier, mars, mai, juillet, septembre et novembre de l'année 2006.	67
4.3	Provinces biogéochimiques de Longhurst (Longhurst, 1995), telles que présentées dans Doron et al. (2011). Les zones côtières sont exclues de cette représentation.	68
4.4	Profils verticaux de la chlorophylle présentés en moyenne spatiale pour les provinces de Longhurst 4, 6, 7 et 9. La courbe bleue représente une climatologie établie par Sauzede et al. (2015) (notée CHL data), la courbe marron est une climatologie sur 6 ans de la simulation <i>detIni</i> (notée ici <i>detIni-clim</i>). Les courbes vertes et rouges sont des moyennes mensuelles des modèles LOBSTER (simulation <i>detLob</i> , notée ici Lobster) et PISCES (simulation <i>detIni</i>), présentées pour les mois de mars (1 ^{ère} ligne), mai (2 ^{ème} ligne) et novembre (3 ^{ème} ligne) de l'année 2006.	70
4.5	Cartes de surface des écarts-types (sd, 1 ^{ère} ligne) et des écarts-types relatifs (relative sd, 2 ^{ème} ligne) interannuels de la chlorophylle, calculés selon l'équation 4.1 à partir des 6 années de simulation. Les résultats sont présentés pour le 31 mars (1 ^{ère} colonne), le 15 mai (2 ^{ème} colonne) et le 27 octobre (3 ^{ème} colonne).	72
4.6	Séries temporelles des écarts-types (haut) et des écarts-types relatifs (bas) interannuels de la chlorophylle de surface, calculés selon l'équation 4.1 à partir des 6 années de simulation. Les résultats sont présentés en moyenne spatiale pour les provinces de Longhurst 4, 6, 7 et 9.	74
4.7	Structures verticales des écarts-types (sd, 1 ^{ère} colonne) et des écarts-types relatifs interannuels (relative sd, 2 ^{ème} colonne) de la chlorophylle le long d'une coupe longitudinale à 30°W. Les résultats sont calculés selon l'équation 4.1 à partir des 6 années de simulation. Ils sont présentés pour le 31 mars (1 ^{ère} ligne), le 15 mai (2 ^{ème} ligne) et le 27 octobre (3 ^{ème} ligne).	75
4.8	Séries temporelles de la RMSE (haut) et de la RMSE relative (bas) moyennées spatialement dans les provinces de Longhurst 4,6,7 et 9. La RMSE est calculée selon l'équation 4.2 et la RMSE relative est calculée comme le rapport ($RMSE^t/\bar{x}^{t_a}$) entre la RMSE (absolue) et la concentration climatologique de la chlorophylle \bar{x}^{t_a}	78
4.9	Séries temporelles de la RMSE moyennée spatialement puis normalisées par l'écart-type sd^t moyen calculé précédemment selon l'équation 4.1. Les résultats sont présentés pour les provinces de Longhurst 4,6,7 et 9. La ligne en pointillés verts exprime l'égalité entre la RMSE et l'écart-type interannuel.	79
4.10	Séries temporelles des écarts-types de la chlorophylle de surface calculés dans les provinces de Longhurst 4, 6, 7 et 9 pour chacune des 6 années de simulation. Les lignes en pointillés présentent des résultats pour les observations SeaWIFS et les lignes pleines pour la simulation <i>detIni</i>	80

4.11 Figure du haut : séries temporelles du rapport CHL^D/CHL^N entre les concentrations moyennes de la chlorophylle des diatomées et celles du nanophytoplancton obtenues à partir des 6 années de simulation. Les séries temporelles sont présentées en moyenne spatiale dans les provinces 4, 6, 7 et 9. La droite en pointillés noirs indique des concentrations identiques pour ces 2 catégories d'espèces. Pour ces mêmes provinces, la figure du bas présente les mêmes séries temporelles pour la chlorophylle totale $CHL^D + CHL^N$ 83

4.12 Séries temporelles du rapport $sd_{CHL^D}^t/sd_{CHL^N}^t$, entre les écarts-types interannuels des concentrations en chlorophylle des diatomées et celles du nanophytoplancton, moyennées spatialement dans les provinces 4, 6, 7 et 9. La droite en pointillés noirs indique une variabilité identique pour les 2 espèces. 83

4.13 Cartes de prédominance des espèces de phytoplancton, présentées au mois de mai 2005 comme le rapport entre la quantité de chlorophylle de l'espèce considérée et la concentration totale de chlorophylle. L'identification des espèces des observations satellites SeaWIFS est faite en 6 espèces selon la méthode PHYSAT. Les organismes représentés sont les diatomées (1^{ère} ligne, 1^{ère} colonne), la somme des 3 espèce de plus grande taille (1^{ère} ligne, 2^{ème} colonne), les nanoeucaryotes (2^{ème} ligne, 1^{ère} colonne) et la somme des 3 espèces de plus petite taille (2^{ème} ligne, 2^{ème} colonne). La contribution relative des diatomées (1^{ère} ligne, 3^{ème} colonne) et du nanophytoplancton (2^{ème} ligne, 3^{ème} colonne) de la simulation *detIni* est également présentée. 85

5.1 Cartes de perturbations générées avec une corrélation spatiale de 5 points de grille (gauche) et de 12 points de grille (droite). Ces perturbations sont tirées dans une gaussienne de moyenne nulle et d'écart-type 0.3. 93

5.2 Comparaison entre une distribution de probabilité log-normale $log\mathcal{N}(\mu = 0, \sigma = 0.3)$, utilisée pour perturber les paramètres du modèle, et une distribution gaussienne $\mathcal{N}(\mu = 1, \sigma = 0.3)$ 94

5.3 Exemple de fluctuations sous-maille pouvant exister à l'intérieur d'un point de grille d'un modèle. 96

5.4 Cartes du rapport $sd_a/chl_{1/4^\circ}$ entre l'écart-type sd_a des anomalies des concentrations en chlorophylle des données SeaWIFS avant et après filtrage au $1/4^\circ$ et la moyenne de chlorophylle SeaWIFS au $1/4^\circ$ après filtrage $chl_{1/4^\circ}$. La ligne du haut présente les résultats pour les mois de mars (1^{ère} colonne) et de mai (2^{ème} colonne). La ligne du bas présente les résultats pour les mois d'août (1^{ère} colonne) et d'octobre (2^{ème} colonne). 97

5.5 Cartes de chlorophylle de surface des observations SeaWIFS (1^{ère} ligne), de la simulation déterministe *detIni* (2^{ème} ligne) et des 3 membres de la simulation d'ensemble *stoIni* (lignes 3 à 5). Les résultats sont présentés en moyenne de 5 jours pour le 15 mai (1^{ère} colonne) et le 27 octobre (2^{ème} colonne) de l'année 2005. 99

5.6 Séries temporelles présentant les concentrations de la chlorophylle de surface moyennées sur l'ensemble du domaine. Les résultats sont présentés pour la simulation déterministe *detIni* (courbe bleue), les données SeaWIFS (courbe jaune), un membre de la simulation perturbée *stoIni* (courbe verte) et la simulation déterministe *detUpd* établie avec le jeu de paramètres calibré pour l'approche probabiliste (courbe rouge). 100

5.7	Cartes de chlorophylle de surface des observations SeaWIFS (1 ^{ère} ligne), de la simulation déterministe <i>detUpd</i> (2 ^{ème} ligne) et des 3 membres de la simulation d'ensemble <i>stoUpd</i> (lignes 3 à 5). Les résultats sont présentés en moyenne de 5 jours pour le 15 mai (1 ^{ère} colonne) et le 27 octobre (2 ^{ème} colonne) de l'année 2005.	102
5.8	Cartes de chlorophylle de surface de la moyenne d'ensemble de la simulation <i>stoUpd</i> (1 ^{ère} ligne) et de la différence entre la moyenne d'ensemble et la simulation <i>detIni</i> (2 ^{ème} ligne). Ces résultats sont présentés pour le 31 mars (1 ^{ère} colonne), le 15 mai (2 ^{ème} colonne) et le 27 octobre (3 ^{ème} colonne) de l'année 2005.	104
5.9	Cartes de surface de l'écart-type (1 ^{ère} ligne) et de l'écart-type relatif de la simulation d'ensemble <i>stoUpd</i> (2 ^{ème} ligne). Ces résultats sont présentés sur la chlorophylle pour le 31 mars (1 ^{ère} colonne), le 15 mai (2 ^{ème} colonne) et le 27 Octobre (3 ^{ème} colonne) de l'année 2005.	106
5.10	Coupes verticales à 30°W de la chlorophylle pour le 31 mars (1 ^{ère} colonne), le 15 mai (2 ^{ème} colonne). La 1 ^{ère} ligne présente les structures moyennes de la simulation d'ensemble et la 2 ^{ème} ligne celles de la simulation déterministe <i>detIni</i> . La différence entre la moyenne d'ensemble et la simulation déterministe <i>detIni</i> est également présentée pour les 2 dates (3 ^{ème} ligne).	107
5.11	Coupes verticales à 30°W de l'écart-type (1 ^{ère} colonne) et de l'écart-type relatif (2 ^{ème} colonne) de la simulation d'ensemble <i>stoUpd</i> . Les résultats sont présentés sur la chlorophylle pour le 31 mars (1 ^{ère} ligne), le 15 mai 2005 (2 ^{ème} ligne) et le 15 octobre 2005 (3 ^{ème} ligne).	108
6.1	Réponse statistique de la chlorophylle de surface aux paramétrisations stochastiques telle que décrite par la simulation d'ensemble à 60 membres pour le 15 mai 2005. De gauche à droite, la 1 ^{ère} ligne présente les cartes des minima, des 1 ^{ères} quartiles et des médianes de l'ensemble. Les images de la 2 ^{ème} ligne représentent les 3 ^{èmes} quartiles et les maxima de l'ensemble. Les observations satellites SeaWIFS sont également représentées.	113
6.2	Légende identique à la figure 6.1 pour le 27 octobre 2005.	114
6.3	Centiles d'ensemble de la chlorophylle de surface le long de coupes longitudinales à 20°W (2 figures du haut) et 30°W (2 figures du bas) pour le 15 mai et le 27 octobre 2005. Les maxima et minima sont représentés en noir, les 1 ^{ères} et 3 ^{èmes} quartiles en vert et la médiane en bleu foncé. Les points rouges indiquent les observations SeaWIFS. La simulation déterministe <i>detIni</i> est représentée en bleu cyan. À noter que l'incertitude des observations n'est pas considérée ici. Elle sera prise en compte dans la section 6.2.	115
6.4	Séries temporelles de la chlorophylle de surface simulée par les membres de l'ensemble en différents points de grille du domaine. Les observations SeaWIFS sont indiquées par des cercles pleins rouges. L'enveloppe statistique de l'ensemble est caractérisée par les quartiles de l'ensemble.	117
6.5	Localisation des points de grille stations utilisés dans cette étude	118

6.6 Rangs des observations SeaWIFS dans l'ensemble calculés sans erreur d'observation sur la totalité du domaine pour le 31 mars (1^{ère} colonne), le 15 mai (2^{ème} colonne) et le 27 octobre (3^{ème} colonne) 2005. La ligne du haut présente ces rangs sous forme d'histogrammes. La fréquence d'occurrence est exprimée en pourcentage du nombre total d'observations. Les lignes en pointillés bleus précisent la fréquence d'occurrence d'un histogramme plat (parfaite fiabilité de l'ensemble.) Les cartes de la ligne du bas affichent la répartition spatiale de ces rangs. 121

6.7 Rangs des observations SeaWIFS dans l'ensemble calculés avec une erreur d'observation de 30% sur la totalité du domaine pour le 31 mars (1^{ère} colonne), le 15 mai (2^{ème} colonne) et le 27 octobre (3^{ème} colonne) 2005. La ligne du haut présente ces rangs sous forme d'histogrammes. La fréquence d'occurrence est exprimée en pourcentage du nombre total d'observations. Les lignes en pointillés bleus précisent la fréquence d'occurrence d'un histogramme plat (parfaite fiabilité de l'ensemble) Les cartes de la ligne du bas affichent la répartition spatiale de ces rangs. 123

6.8 Rangs des observations SeaWIFS dans l'ensemble calculés avec une erreur d'observation de 30% sur toute l'année 2005 dans les provinces de Longhurst 4 (1^{ère} ligne, 1^{ère} colonne), 6 (1^{ère} ligne, 2^{ème} colonne), 7 (2^{ème} ligne, 1^{ère} colonne) et 9 (2^{ème} ligne, 2^{ème} colonne). Les rangs sont présentés sous forme d'histogrammes. La fréquence d'occurrence est exprimée en pourcentage du nombre total d'observations. Les lignes en pointillés bleus précisent la fréquence d'occurrence d'un histogramme plat (parfaite fiabilité de l'ensemble). 124

6.9 Rangs des observations SeaWIFS présentés par saison, sous forme d'histogrammes. La 1^{ère} colonne correspond aux mois d'hiver (janvier à mars), la 2^{ème} colonne au printemps (avril à juin), la 3^{ème} colonne à l'été (juillet à septembre) et la dernière colonne à l'hiver (octobre à décembre). Les rangs sont calculés avec 30% d'erreur d'observation dans les provinces de Longhurst 4 (1^{ère} ligne), 6 (2^{ème} ligne), 7 (3^{ème} ligne) et 9 (4^{ème} ligne). La fréquence d'occurrence est exprimée en pourcentage du nombre total d'observations. Les lignes en pointillés bleus précisent les occurrences d'un histogramme plat (fiabilité parfaite). 125

6.10 Réponse statistique de la chlorophylle aux paramétrisations stochastiques telle que décrite par la simulation d'ensemble pour le 15 mai (1^{ère} colonne) et le 27 octobre (2^{ème} colonne) 2005. Les quantiles sont identifiés le long de coupes longitudinales à 30°W. De haut en bas, on représente le minimum de l'ensemble, le premier quartile, la médiane, le troisième quartile et le maximum de l'ensemble. 127

6.11 Séries temporelles du minimum (1^{ère} ligne), de la médiane (2^{ème} ligne) et du maximum de l'ensemble (3^{ème} ligne) aux stations *STAT_B* (1^{ème} colonne) et *INDIA* (2^{ème} colonne). Le séries temporelles de la simulation *detIni* sont présentées (dernière ligne). 128

6.12 Profils verticaux en chlorophylle en moyenne spatiale pour les provinces de Longhurst 4, 6, 7 et 9. Les courbes noires représentent les moyennes mensuelles par province de Longhurst des minimums de l'ensemble et des maximums de l'ensemble. La courbe bleue représente les données de chlorophylle climatologiques et la courbe rouge représente les profils verticaux de la simulation déterministe *detIni*. 130

6.13	Profils verticaux des 60 membres de la simulation d'ensemble au point de grille le plus proche de la station BATS. Les observations disponibles à cette station sont représentées en bleu et les membres de l'ensemble en vert. Pour chaque date les réalisations de l'ensemble dont l'allure du profil vertical se rapproche le plus de l'observation sont représentées en noir. Le profil vertical de la simulation déterministe est représenté en rouge.	133
6.14	Profils verticaux des 60 membres de la simulation d'ensemble classés en fonction des concentrations en chlorophylle de surface. Les profils sont rangés comme suit : Pour la ligne du haut : 1 ^{ère} colonne $\text{CHL} \leq 0.15 \text{ mg/m}^3$; 2 ^{ème} colonne $0.15 \text{ mg/m}^3 < \text{CHL} \leq 0.3 \text{ mg/m}^3$, 3 ^{ème} colonne $0.3 \text{ mg/m}^3 < \text{CHL} \leq 0.45 \text{ mg/m}^3$. Pour la ligne du bas : 1 ^{ère} colonne $0.45 \text{ mg/m}^3 < \text{CHL} \leq 0.6 \text{ mg/m}^3$; 2 ^{ème} colonne $0.6 \text{ mg/m}^3 < \text{CHL} \leq 0.75 \text{ mg/m}^3$; 3 ^{ème} colonne $\text{CHL} > 0.75 \text{ mg/m}^3$	134
7.1	Principe simplifié de l'assimilation séquentielle de données.	139
7.2	Distributions de probabilité de l'ensemble a priori (1 ^{ère} ligne) aux points de grille BATS, STAT_E et INDIA (cf. figure 6.5) pour le 15 mai 2005. Les histogrammes de la 2 ^{ème} ligne indiquent ces mêmes distributions de probabilité après anamorphose. La comparaison avec une distribution gaussienne est explicitée par l'intermédiaire des courbes en pointillés rouges.	145
7.3	Cartes de chlorophylle de surface de la moyenne de l'ensemble a priori (1 ^{ère} ligne), de la moyenne de l'ensemble après analyse (2 ^{ème} ligne) et des observations SeaWIFS-HR utilisées pour l'analyse (3 ^{ème} ligne). Ces résultats sont présentés pour le 31 mars (1 ^{ère} colonne), le 15 mai (2 ^{ème} colonne) et le 27 octobre (3 ^{ème} colonne) de l'année 2005.	146
7.4	Histogrammes des moyennes de l'ensemble a priori (vert) et de l'ensemble analysé (violet) de la chlorophylle de surface à tous les points de grille situées à l'intérieur des provinces de Longhurst 4,6 et 18. La superposition des 2 histogrammes est représentée en bleu. Les lignes en pointillés indiquent la concentration moyenne en chlorophylle de surface des différentes provinces. Ces résultats sont présentés pour le 15 mai (1 ^{ère} ligne) et le 27 octobre (2 ^{ème} ligne) de l'année 2005	147
7.5	Cartes de surface de l'écart-type relatif de la chlorophylle de l'ensemble analysé (1 ^{ère} ligne) et du rapport avec l'écart-type relatif de l'ensemble a priori (analysé/a priori, 2 ^{ème} ligne). Ces résultats sont présentés pour le 31 mars (1 ^{ère} colonne), le 15 mai (2 ^{ème} colonne) et le 27 octobre (3 ^{ème} colonne) de l'année 2005.	149
7.6	Séries temporelles de la chlorophylle de surface simulée par les membres de l'ensemble a priori (vert) et de l'ensemble analysé (bleu) en plusieurs points de grille du domaine (cf. figure 6.5). Les observations SeaWIFS-HR sont indiquées par des cercles rouges. Les médianes (courbes rouge et cyan) et les extrema des ensembles (courbes noires et orange) sont également représentés.	150
7.7	Coupes verticales en chlorophylle à 30°W de la moyenne de l'ensemble a priori (1 ^{ère} ligne), de la moyenne de l'ensemble actualisé par filtre d'ensemble ETKF (2 ^{ème} ligne) et du rapport entre les 2 (analysé/a priori, 3 ^{ème} ligne). Ces résultats sont présentés pour le 15 mai (1 ^{ère} colonne) et le 27 octobre (2 ^{ème} colonne) de l'année 2005.	152

7.8	Coupes verticales en chlorophylle à 30°W de l'écart-type relatif de l'ensemble actualisé par filtre d'ensemble ETKF (1 ^{ère} ligne) et du rapport avec l'écart-type relatif de l'ensemble a priori (analysé/a priori, 2 ^{ème} ligne). Ces résultats sont présentés pour le 15 mai (1 ^{ère} colonne) et le 27 octobre (2 ^{ème} colonne) de l'année 2005.	153
7.9	Profils verticaux en chlorophylle des membres de l'ensemble a priori (vert) et analysé (bleu) aux points de grille (cf. figure 6.5) correspondant aux stations BATS (1 ^{ère} ligne), INDIA (2 ^{ème} ligne) et STAT_C (3 ^{ème} ligne). Les observations SeaWIFS-HR sont indiquées par des cercles rouges. Les médianes (courbes rouges et bleu cyan) et les extrema (courbes noires et orange) sont également représentés.	154
7.10	Histogrammes de rangs des observations MERIS, calculés pour la totalité du domaine avec une erreur d'observation de 30% au 31 mars (1 ^{ère} colonne), 15 mai (2 ^{ème} colonne) et 27 octobre (3 ^{ème} colonne) 2005. La ligne du haut présente les rangs dans l'ensemble a priori et celle du bas dans l'ensemble analysé. La fréquence d'occurrence est en pourcentage (sur le nombre total d'observations). Les lignes en pointillés précisent la fréquence d'occurrence d'un histogramme plat.	155
7.11	Cartes des rangs des observations MERIS, calculés pour la totalité du domaine avec une erreur d'observation de 30% au 31 mars (1 ^{ère} colonne), 15 mai (2 ^{ème} colonne) et 27 octobre (3 ^{ème} colonne) 2005. La ligne du haut présente la répartition spatiale des rangs dans l'ensemble a priori et celle du bas dans l'ensemble analysé.	156
7.12	Histogrammes de rangs des observations MERIS, calculés sur toute l'année 2005 avec une erreur d'observation de 30% dans les provinces de Longhurst 4 (1 ^{ère} ligne, 1 ^{ère} colonne), 6 (1 ^{ère} ligne, 2 ^{ème} colonne), 7 (2 ^{ème} ligne, 1 ^{ère} colonne) et 9 (2 ^{ème} ligne, 2 ^{ème} colonne). La ligne du haut présente les rangs dans l'ensemble a priori et celle du bas dans l'ensemble analysé. La fréquence d'occurrence est exprimée en pourcentage (sur le nombre total d'observations). La ligne en pointillés bleus précise la fréquence d'occurrence d'un histogramme plat.	157
7.13	Histogrammes de rangs des observations MERIS dans l'ensemble analysé présentés par saison. La 1 ^{ère} colonne correspond aux mois d'hiver (de janvier à mars), la 2 ^{ème} colonne au printemps (d'avril à juin), la 3 ^{ème} colonne à l'été (de juillet à septembre) et la dernière colonne à l'automne (de octobre à décembre). Les rangs sont calculés avec une erreur d'observation de 30% dans les provinces de Longhurst 4 (1 ^{ère} ligne), 6 (2 ^{ème} ligne), 7 (3 ^{ème} ligne) et 9 (4 ^{ème} ligne). La fréquence d'occurrence est exprimée en pourcentage (sur le nombre total d'observations). La ligne en pointillés bleus précise la fréquence d'occurrence d'un histogramme plat.	158
7.14	Séries temporelles, sur l'année 2005, des valeurs de CRPS calculées pour la simulation d'ensemble a priori <i>stoUpd-xf</i> (courbe en pointillés bleus) et la simulation de l'ensemble analysé <i>stoUpd-xa</i> (courbe rouge). À chaque pas de temps le CRPS est calculé à partir de toutes les observations disponibles sur le domaine.	160
7.15	Séries temporelles, sur l'année 2005, des valeurs de CRPS calculées dans les provinces de Longhurst 4, 6, 7 et 9. Les courbes en pointillés noirs présentent les résultats de l'ensemble a priori <i>stoUpd-xf</i> et les courbes violettes ceux de l'ensemble analysé <i>stoUpd-xa</i> . À chaque pas de temps, le CRPS est calculé à partir de toutes les observations disponibles dans les provinces de Longhurst considérées.	161

7.16	Séries temporelles, sur l'année 2005, des valeurs de la fiabilité (figure du haut) et de la résolution (figure du bas) de la simulation d'ensemble a priori <i>stoUpd-xf</i> (courbe en pointillés bleus) et de l'ensemble analysé <i>stoUpd-xa</i> (courbe rouge), calculées à partir de la décomposition du CRPS. Toutes les observations disponibles sur le domaine à chaque pas de temps sont utilisées.	162
7.17	Séries temporelles sur l'année 2005 des valeurs de de fiabilité (1 ^{ère} colonne) et de résolution (2 ^{ème} colonne) obtenues pour les provinces de Longhurst 4, 6, 7 et 9. Les courbes en pointillés affichent les résultats pour l'ensemble a priori <i>stoUpd-xf</i> et les courbes pleines de l'ensemble analysé <i>stoUpd-xa</i> obtenus à partir de la décomposition du CRPS. La moyenne utilisée est spatiale, réalisée avec toutes les observations disponibles dans les différentes provinces de Longhurst.	163
7.18	Séries temporelles de surface de l'ensemble a priori (courbes vertes) et de l'ensemble analysé (courbes bleues) présentées aux points de grille stations STAT_F (1 ^{ère} colonne et STAT_E (2 ^{ème} colonne) pour 5 variables non observées impactées par les analyses. De haut en bas sont représentés : la chlorophylle du nanophytoplancton (NCHL), l'ammonium (NH4), le microzooplancton (ZOO), et le phosphate (PO4). La dernière ligne présente les séries temporelles des silicates (1 ^{ère} colonne) et du fer (2 ^{ème} colonne), uniquement à la station STAT_E.	165
7.19	Profils verticaux de l'ensemble a priori (courbes vertes) et de l'ensemble analysé (courbes bleues) aux points de grille stations STAT_D et STAT_E pour 5 variables non observées impactées par les analyses. Les informations nécessaires à l'identification des différentes variables sont indiquées sur les figures.	167


```

!-----
nn_bathy = 1 ! compute (=0) or read (=1) the
bathymetry file
nn_closea = 0 ! remove (=0) or keep (=1) closed
seas and lakes (ORCA)
nn_msh = 0 ! create (/=0) a mesh file(s) or
not (=0)
! if not 0 can be in [1 - 6 ] for drakkar usually 6
rn_hmin = -3. ! min depth of the ocean (>0) or
min number of ocean level (<0)
rn_e3zps_min= 25. ! partial step thickness is set
larger than the minimum of
rn_e3zps_rat= 0.2 ! rn_e3zps_min and rn_e3zps_rat*e3t,
with 0<rn_e3zps_rat<1
!
rn_rdt = 2400. ! time step for the dynamics (and
tracer if nn_acc=0)
nn_baro = 64 ! number of barotropic time step
("key_dynspg_ts")
rn_atfp = 0.1 ! asselin time filter parameter
nn_acc = 0 ! acceleration of convergence : =1
used, rdt < rdtrtra(k)
!
rn_rdtmin = 2400. ! minimum time step on tracers
(used if nn_acc=1)
rn_rdtmax = 2400. ! maximum time step on tracers
(used if nn_acc=1)
rn_rdth = 800. ! depth variation of tracer time
step (used if nn_acc=1)
/
!-----
&namtsd ! data : Temperature & Salinity
!-----
! ! file name ! frequency (hours) ! variable !
time interp. ! clim ! 'yearly' or ! weights ! rotation !
! ! ! ! (if <0 months) ! name !
(logical) ! (T/F) ! 'monthly' ! filename ! pairing !
! data used for initial condition (istate)
sn_tem_ini = 'Levitus_p2.1_1m_Tpot_nat1025' , -1,'votemper',
.true. , .true., 'yearly' , , , ,
sn_sal_ini = 'Levitus_p2.1_1m_S_correc_nat1025' ,
-1,'vosaline' , .true. , .true., 'yearly' , , ,
! data used for damping ( tradmp)
sn_tem_dmp = 'Levitus_p2.1_1m_Tpot_nat1025' , -1,'votemper',
.true. , .true., 'yearly' , , , ,
sn_sal_dmp = 'Levitus_p2.1_1m_S_correc_nat1025' ,
-1,'vosaline' , .true. , .true., 'yearly' , , ,
!
cn_dir = './' ! root directory for the location
of the temperature and salinity file
ln_tsd_init = .true. ! Initialisation of ocean T & S
with T & S input data (T) or not (F)
ln_tsd_tradmp = .true. ! damping of ocean T & S toward T
&S input data (T) or not (F)
/
!-----
!! ** Stochastic parameterization namelists
!!-----
!!-----
&namsto ! Stochastic parameterization
!-----
ln_rststo = .RESTART. ! start from mean parameter
(F) or from restart file (T)
cn_storst_in = "restart_sto" ! suffix of stochastic
parameter restart file (input)
cn_storst_out = "restart_sto" ! suffix of stochastic
parameter restart file (output)
ln_sto_ldf = .false. ! stochastic lateral diffusion
rn_ldf_std = 0.0 ! lateral diffusion standard
deviation (in percent)
rn_ldf_tcor = 0.0 ! lateral diffusion time
correlation (in timesteps)
ln_sto_hpg = .false. ! stochastic horizontal
pressure gradient
rn_hpg_std = 0.0 ! horizontal density gradient
standard deviation (in percent)
rn_hpg_tcor = 0.0 ! horizontal density gradient
time correlation (in timesteps)
ln_sto_trd = .true. ! stochastic trend
nn_sto_trd = 7 ! number of matrices
rn_trd_std = 0.3 ! trend standard deviation (in
percent)
rn_trd_tcor = 36 ! trend time correlation (in
timesteps)
nn_trd_ord = 1 ! order of autoregressive processes
nn_trdflt = 4 ! passes of laplacien filter
ln_sto_eos = .false. ! stochastic equation of state
nn_sto_eos = 1 ! number of independent random
walks
rn_eos_stdxy = 1.4 ! random walk horz. standard
deviation (in grid points)
rn_eos_stdz = 0.7 ! random walk vert. standard
deviation (in grid points)
rn_eos_tcor = 1440.0 ! random walk time correlation
(in timesteps)
nn_eos_ord = 1 ! order of autoregressive
processes
nn_eosflt = 0 ! passes of Laplacian filter
rn_eos_lim = 2.0 ! limitation factor (default =
3.0)
ln_sto_trc = .true. ! stochastic tracer dynamics
nn_sto_trc = 1 ! number of independent
fluctuations
rn_trc_std = 0.2 ! fluctuations standard
deviation (in percent)
rn_trc_tcor = 1080 ! fluctuations time
correlation (in timesteps)
nn_trc_ord = 1 ! order of autoregressive
processes
nn_trcflt = 4 ! passes of Laplacian filter
rn_trc_lim = 2 ! limitation factor (default =
3.0)
/
!!-----
!! ** Surface Boundary Condition namelists **
!!-----
!! namsbc surface boundary condition
!! namsbc_ana analytical formulation
!! namsbc_flux flux formulation
!! namsbc_clio CLIO bulk formulae formulation
!! namsbc_core CORE bulk formulae formulation
!! namsbc_mfs MFS bulk formulae formulation
!! namsbc_cpl CouPLed formulation
("key_coupled")
!! namtra_qsr penetrative solar radiation
!! namsbc_rnf river runoffs
!! namsbc_apr Atmospheric Pressure
!! namsbc_ssr sea surface restoring term (for T
and/or S)
!! namsbc_alb albedo parameters
!!-----
&namsbc ! Surface Boundary Condition (surface
module)
!-----
nn_fsbc = 1 ! frequency of surface boundary
condition computation
! (also = the frequency of sea-ice model call)
ln_ana = .false. ! analytical formulation
=> fill namsbc_ana )
ln_flux = .false. ! flux formulation
ln_blk_core = .false. ! CLIO bulk formulation
=> fill namsbc_clio )
ln_blk_core = .true. ! CORE bulk formulation
=> fill namsbc_core )
ln_blk_mfs = .false. ! MFS bulk formulation
=> fill namsbc_mfs )
ln_cpl = .false. ! Coupled formulation
=> fill namsbc_cpl )
ln_apr_dyn = .false. ! Patm gradient added in ocean &
ice Eqs. (T => fill namsbc_apr )
nn_ice = 2 ! =0 no ice boundary condition ,
! =1 use observed ice-cover ,
! =2 ice-model used ("key_lim3" or
"key_lim2)
ln_dm2dc = .false. ! daily mean to diurnal cycle on
short wave
ln_rnf = .true. ! runoffs
=> fill namsbc_rnf )
ln_ssr = .true. ! Sea Surface Restoring on T and/or
S
(T => fill namsbc_ssr)
nn_fwb = 1 ! FreshWater Budget: =0 unchecked
! =1 global mean of e-p-r set to zero at each time step
! =2 annual global mean of e-p-r set to zero
! =3 global emp set to zero and spread out over erp area
ln_cdgw = .false. ! Neutral drag coefficient read
from wave model (T => fill namsbc_wave)
/
!!-----
&namsbc_ana ! analytical surface boundary condition
!!-----
nn_tau000 = 0 ! gently increase the stress over
the first ntau_rst time-steps
rn_utau0 = 0.5 ! uniform value for the i-stress

```

```

    rn_vtau0 = 0.e0 ! uniform value for the j-stress
    rn_qns0 = 0.e0 ! uniform value for the total heat
flux
    rn_qsr0 = 0.e0 ! uniform value for the solar
radiation
    rn_emp0 = 0.e0 ! uniform value for the freshwater
budget (E-P)
/
!-----
&namsbc_flux ! surface boundary condition : flux
formulation
!-----
! ! file name ! frequency (hours) ! variable
! time interp. ! clim ! 'yearly' / ! weights ! rotation !
! ! ! ! (if <0 months) ! name
! (logical) ! (T/F) ! 'monthly' ! filename ! pairing !
sn_utau = 'utau' , 24 , 'utau' ,
.false. , .false. , 'yearly' , " , " ,
sn_vtau = 'vtau' , 24 , 'vtau' ,
.false. , .false. , 'yearly' , " , " ,
sn_qtot = 'qtot' , 24 , 'qtot' ,
.false. , .false. , 'yearly' , " , " ,
sn_qsr = 'qsr' , 24 , 'qsr' ,
.false. , .false. , 'yearly' , " , " ,
sn_emp = 'emp' , 24 , 'emp' ,
.false. , .false. , 'yearly' , " , " ,
cn_dir = './' ! root directory for the location
of the flux files
/
!-----
&namsbc_cli0 ! namsbc_cli0 CLI0 bulk formulae
!-----
! ! file name ! frequency (hours) ! variable
! time interp. ! clim ! 'yearly' / ! weights ! rotation !
! ! ! ! (if <0 months) ! name
! (logical) ! (T/F) ! 'monthly' ! filename ! pairing !
sn_taux = 'taux_lm' , -1 , 'sozotaux' ,
.true. , .true. , 'yearly' , " , " ,
sn_vtau = 'tauy_lm' , -1 , 'sometauy' ,
.true. , .true. , 'yearly' , " , " ,
sn_wndm = 'flx' , -1 , 'socliowi' ,
.true. , .true. , 'yearly' , " , " ,
sn_tair = 'flx' , -1 , 'socliot2' ,
.true. , .true. , 'yearly' , " , " ,
sn_humi = 'flx' , -1 , 'socliohu' ,
.true. , .true. , 'yearly' , " , " ,
sn_ccov = 'flx' , -1 , 'socliocl' ,
.false. , .true. , 'yearly' , " , " ,
sn_prec = 'flx' , -1 , 'socliopl' ,
.false. , .true. , 'yearly' , " , " ,
cn_dir = './' ! root directory for the location
of the bulk files are
/
!-----
&namsbc_core ! namsbc_core CORE bulk formulae
!-----
! ! file name ! frequency (hours) !
variable ! time interp. ! clim ! 'yearly' or ! weights !
rotation !
! ! ! ! (if <0 months) !
name ! (logical) ! (T/F) ! 'monthly' ! filename !
pairing !
sn_wndi = 'drowned_u10 ERAinterim' , 3 , 'u10'
, .true. , .false. , 'yearly' , 'bicubic_weights_NATLO25.nc'
, 'U1'
sn_wndj = 'drowned_v10 ERAinterim' , 3 , 'v10'
, .true. , .false. , 'yearly' , 'bicubic_weights_NATLO25.nc'
, 'V1'
sn_qsr = 'drowned_radsw ERAinterim' , 24 , 'radsw' 'cartesian' , 'eastward-northward' , 'U,V'
, .false. , .false. , 'yearly' , 'bilinear_weights_NATLO25.nc'
, "
sn_qlw = 'drowned_radlw ERAinterim' , 24 , 'radlw'
, .false. , .false. , 'yearly' , 'bilinear_weights_NATLO25.nc'
, "
sn_tair = 'drowned_t2 ERAinterim' , 3 , 't2'
, .true. , .false. , 'yearly' , 'bilinear_weights_NATLO25.nc'
, "
sn_humi = 'drowned_q2 ERAinterim' , 3 , 'q2'
, .true. , .false. , 'yearly' , 'bilinear_weights_NATLO25.nc'
, "
sn_prec = 'drowned_precip ERAinterim' , 24 , 'precip'
, .true. , .false. , 'yearly' , 'bilinear_weights_NATLO25.nc'
, "
sn_snow = 'drowned_snow ERAinterim' , 24 , 'snow'
, .true. , .false. , 'yearly' , 'bilinear_weights_NATLO25.nc'
, "
sn_tdif = 'taudif' , 24 , 'taudif' ,
.true. , .false. , 'yearly' , "
!
cn_dir = './' ! root directory for the location
of the bulk files
/
!-----
&namsbc_mfs ! namsbc_mfs MFS bulk formulae
!-----
! ! file name ! frequency (hours) ! variable
! time interp. ! clim ! 'yearly' / ! weights ! rotation !
! ! ! ! (if <0 months) ! name
! (logical) ! (T/F) ! 'monthly' ! filename ! pairing !
sn_wndi = 'ecmf' , 6 , 'u10' ,
.true. , .false. , 'daily' , 'bicubic.nc' , " , 'v10' ,
sn_wndj = 'ecmf' , 6 , 'v10' ,
.true. , .false. , 'daily' , 'bicubic.nc' , " , 'clc' ,
sn_clc = 'ecmf' , 6 , 'clc' ,
.true. , .false. , 'daily' , 'bilinear.nc' , " , 'msl' ,
sn_msl = 'ecmf' , 6 , 'msl' ,
.true. , .false. , 'daily' , 'bicubic.nc' , " , 't2' ,
sn_tair = 'ecmf' , 6 , 't2' ,
.true. , .false. , 'daily' , 'bicubic.nc' , " , 'rh' ,
sn_rhm = 'ecmf' , 6 , 'rh' ,
.true. , .false. , 'daily' , 'bilinear.nc' , " , 'precip' ,
sn_prec = 'ecmf' , 6 , 'precip' ,
.true. , .true. , 'daily' , 'bicubic.nc' , "
cn_dir = './ECMWF/' ! root directory for the
location of the bulk files
/
!-----
&namsbc_cpl ! coupled ocean/atmosphere model
!-----
! ! description ! multiple !
vector ! vector ! vector !
! ! ! ! categories !
reference ! orientation ! grids !
! send
sn_snd_temp = 'weighted oce and ice' , 'no' ,
" , " , " , "
sn_snd_alb = 'weighted ice' , 'no' ,
" , " , " , "
sn_snd_thick = 'none' , 'no' ,
" , " , " , "
sn_snd_crt = 'none' , 'no' ,
'spherical' , 'eastward-northward' , 'T'
sn_snd_co2 = 'coupled' , 'no' ,
" , " , " , "
! receive
sn_rcv_w10m = 'none' , 'no' ,
" , " , " , "
sn_rcv_taumod = 'coupled' , 'no' ,
" , " , " , "
sn_rcv_tau = 'oce only' , 'no' ,
" , " , " , "
sn_rcv_dqnsdt = 'coupled' , 'no' ,
" , " , " , "
sn_rcv_qsr = 'oce and ice' , 'no' ,
" , " , " , "
sn_rcv_qns = 'oce and ice' , 'no' ,
" , " , " , "
sn_rcv_emp = 'conservative' , 'no' ,
" , " , " , "
sn_rcv_rnf = 'coupled' , 'no' ,
" , " , " , "
sn_rcv_cal = 'coupled' , 'no' ,
" , " , " , "
sn_rcv_co2 = 'coupled' , 'no' ,
" , " , " , "
/
!-----
&namtra_qsr ! penetrative solar radiation
!-----
! ! file name ! frequency (hours) ! variable
! time interp. ! clim ! 'yearly' / ! weights ! rotation !

```

```

!
! (logical) ! (T/F) ! 'monthly' ! filename ! pairing !
sn_chl = 'chlorophyl', -1, 'CHLA',
.true., .true., 'yearly', " , "
cn_dir = './' ! root directory for the location
of the runoff files
ln_traqsr = .true. ! Light penetration (T) or not (F)
ln_qsr_rgb = .false. ! RGB (Red-Green-Blue) light
penetration
ln_qsr_2bd = .true. ! 2 bands light
penetration
ln_qsr_bio = .false. ! bio-model light penetration
nn_chldta = 1 ! RGB : Chl data (=1) or cst value
(=0)
rn_abs = 0.58 ! RGB & 2 bands: fraction of light
(rn_sil)
rn_si0 = 0.35 ! RGB & 2 bands: shortess depth of
extinction
rn_sil = 23.0 ! 2 bands: longest depth of
extinction
/
!-----
&namsbc_rnf ! runoffs namelist surface boundary
condition
!-----
! ! file name ! frequency (hours)
! variable ! time interp. ! clim ! 'yearly' ! weights !
rotation !
! ! ! (if <0 months) !
name ! (logical) ! (T/F) ! 'monthly' ! filename ! pairing
!
sn_rnf = 'runoff_NATL025_from_coastipt_ant3pt_obtaz_1m_ORCA025',
-1, 'sorunoff', .true., .true.,
'yearly', " , "
sn_cnf = 'runoff_NATL025_from_coastipt_ant3pt_obtaz_1m_ORCA025',
0, 'soccefr', .false., .true., 'yearly'
, " , "
sn_s_rnf = 'runoffs', , 24 ,
'rosaline', .true., .true., 'yearly', " , "
sn_t_rnf = 'runoffs', , 24 ,
'rotemper', .true., .true., 'yearly', " , "
sn_dep_rnf = 'runoffs', , 0 ,
'rodepth', .false., .true., 'yearly', " , "
cn_dir = './' ! root directory for the location
of the runoff files
ln_rnf_emp = .false. ! runoffs included into precipitation
field (T) or into a file (F)
ln_rnf_mouth = .true. ! specific treatment at rivers
mouths
rn_hrnf = 10.e0 ! depth over which enhanced
vertical mixing is used
rn_avt_rnf = 2.e-3 ! value of the additional vertical
mixing coef. [m2/s]
rn_rfact = 1.e0 ! multiplicative factor for runoff
ln_rnf_depth = .false. ! read in depth information for
runoff
ln_rnf_tem = .false. ! read in temperature information
for runoff
ln_rnf_sal = .false. ! read in salinity information for
runoff
/
!-----
&namsbc_apr ! Atmospheric pressure used as ocean
forcing or in bulk
!-----
! ! file name ! frequency (hours) ! variable !
! ! ! (if <0 months) ! ! ! rotation !
! ! ! (if <0 months) ! ! ! name !
(logical) ! (T/F) ! 'monthly' ! filename ! pairing !
sn_apr = 'patm', -1, 'soms1pre',
.true., .true., 'yearly', " , "
cn_dir = './' ! root directory for the location
of the bulk files
ln_ref_apr = .false. ! ref. pressure: global mean Patm
(T) or a constant (F)
/
!-----
&namsbc_ssr ! surface boundary condition : sea surface
restoring
!-----
! ! file name ! frequency (hours) ! variable
! time interp. ! clim ! 'yearly' ! weights ! rotation !
! ! ! (if <0 months) ! ! ! name !
! (logical) ! (T/F) ! 'monthly' ! filename ! pairing !
sn_sst = 'sst_data', , 24 , 'sst' ,
.false., .false., 'yearly', " , "
sn_sss = 'Levitus_p2.1_1m_SSS_correc_nat1025',
-1, 'sss', .true., .true., 'yearly', "
, "
cn_dir = './' ! root directory for the location
of the runoff files
nn_sstr = 0 ! add a retroaction term in the
surface heat flux (=1) or not (=0)
nn_ssr = 2 ! add a damping term in the
surface freshwater flux (=2)
! or to SSS only (=1) or no damping term (=0)
rn_dqdt = -40. ! magnitude of the retroaction on
temperature [W/m2/K]
rn_deds = -166.67 ! magnitude of the damping on
salinity [mm/day]
ln_ssr_bnd = .true. ! flag to bound erp term (associated
with nn_ssr=2)
rn_ssr_bnd = 4.e0 ! ABS(Max/Min) value of the damping
erp term [mm/day]
ln_ssrflt = .false. ! use filtering of SSS model for sss
restoring
nn_shap_iter = 100 ! number of iteration of the shapiro
filter
ln_ssr_msk = .false. ! use a mask near the coast
! ! file name ! frequency (hours) ! variable
! time interp. ! clim ! 'yearly' ! weights ! rotation !
! ! ! (if <0 months) ! ! ! name !
! (logical) ! (T/F) ! 'monthly' ! filename ! pairing !
sn_coast = 'distcoast', , 0 , 'tcoast' ,
.false., .true., 'yearly', " , "
rn_dist = 150. ! distance to the coast
/
!-----
&namsbc_alb ! albedo parameters
!-----
ln_rn_cloud = 0.06 ! cloud correction to snow and ice
albedo
rn_albice = 0.53 ! albedo of melting ice in the
arctic and antarctic
rn_alphd = 0.80 ! coefficients for linear
interpolation used to
rn_alphc = 0.65 ! compute albedo between two
extremes values
rn_alphdi = 0.72 ! (Pyane, 1972)
/
!-----
!! *** Lateral boundary condition ***
!-----
!! namlbc lateral momentum boundary condition
!! namcla cross land advection
!! namobc open boundaries parameters ("key_
!! namagrif agrif nested grid ( read by child model
only ) ("key_agrif")
!! nambody Unstructured open boundaries ("key_
!! namtide Tidal forcing at open boundaries
("key_bdy_tides")
!-----
&namlbc ! lateral momentum boundary condition
!-----
rn_shlat = 0. ! shlat = 0 ! 0 < shlat < 2 !
shlat = 2 ! 2 < shlat
! free slip ! partial slip ! no slip ! strong slip
ln_vorlat = .false. ! consistency of vorticity boundary
condition with analytical eqs.
/
!-----
&namlcla ! cross land advection
!-----
nn_cla = 0 ! advection between 2 ocean pts
separates by land
/
!-----
&namobc ! open boundaries parameters
!-----
ln_obc_clim = .false. ! climatological obc data files (T)
or not (F)
ln_vol_cst = .true. ! impose the total volume
conservation (T) or not (F)
ln_obc fla = .false. ! Flather open boundary condition
nn_obc dta = 1 ! = 0 the obc data are equal to the
initial state
! = 1 the obc data are read in 'obc.dta' files
cn_obc dta = 'annual' ! set to annual if obc datafile
hold 1 year of data
! set to monthly if obc datafile hold 1 month of data
rn_dpoin = 1. ! damping time scale for inflow at
east open boundary
rn_dpwin = 1. ! - - -
west - - -
rn_dpnin = 1. ! - - -
north - - -

```

```

rn_dpsin = 1. ! - - -
south - - -
rn_dpeob = 3000. ! time relaxation (days) for the
east open boundary
rn_dpwob = 15. ! - - -
west - - -
rn_dpnob = 3000. ! - - -
north - - -
rn_dpsob = 15. ! - - -
south - - -
rn_volemp = 1. ! = 0 the total volume change with
the surface flux (E-P-R)
! = 1 the total volume remains constant
/
!-----
&namagrif ! AGRIF zoom
!-----
nn_cln_update = 3 ! baroclinic update frequency
ln_spd_dyn = .true. ! use 0 as special value for
dynamics
rn_sponge_tra = 2880. ! coefficient for tracer sponge
layer [m2/s]
rn_sponge_dyn = 2880. ! coefficient for dynamics sponge
layer [m2/s]
/
!-----
&nam_tide ! tide parameters (#ifdef key_tide)
!-----
ln_tide_pot = .true. ! use tidal potential forcing
nb_harmo = 11 ! number of constituents used
c1name(1) = 'M2' ! name of constituent
c1name(2) = 'S2'
c1name(3) = 'N2'
c1name(4) = 'K1'
c1name(5) = 'O1'
c1name(6) = 'Q1'
c1name(7) = 'M4'
c1name(8) = 'K2'
c1name(9) = 'P1'
c1name(10) = 'Mf'
c1name(11) = 'Mm'
/
!-----
&nambdy ! unstructured open boundaries
!-----
nb_bdy = 1 ! number of open
boundary sets
ln_coords_file = .true. ! =T : read bdy
coordinates from file
cn_coords_file = 'coordinates.bdy.nc' ! bdy coordinates
files
ln_mask_file = .false. ! =T : read mask from
file
cn_mask_file = " ! name of mask file
(if ln_mask_file=.TRUE.)
nn_dyn2d = 2 ! boundary conditions
for barotropic fields
nn_dyn2d_dta = 3 ! = 0, bdy data are
equal to the initial state
! = 1, bdy data are read in 'bdydata .nc' files
! = 2, use tidal harmonic forcing data from files
! = 3, use external data AND tidal harmonic forcing
nn_dyn3d = 0 ! boundary conditions
for baroclinic velocities
nn_dyn3d_dta = 0 ! = 0, bdy data are
equal to the initial state
! = 1, bdy data are read in 'bdydata .nc' files
nn_tra = 1 ! boundary conditions
for T and S
nn_tra_dta = 1 ! = 0, bdy data are
equal to the initial state
! = 1, bdy data are read in 'bdydata .nc' files
nn_rimwidth = 10 ! width of the
relaxation zone
ln_vol = .false. ! total volume
correction (see nn_volctl parameter)
nn_volctl = 1 ! = 0, the total
water flux across open boundaries is zero
/
!-----
&nambdy_dta ! open boundaries - external data
("key_bdy")
!-----
! ! file name ! frequency (hours) !
variable ! time interpol. ! clim ! 'yearly' ! weights !
rotation !
! ! ! (if <0 months) !
name ! (logical) ! (T/F) ! 'monthly' ! filename !
pairing !
bn_ssh = 'amm12_bdyT_u2d', 24 ,
'sosshieg', .true., .false., 'daily', " , "
bn_u2d = 'amm12_bdyU_u2d', 24 ,
'vobtrctx', .true., .false., 'daily', " , "
bn_v2d = 'amm12_bdyV_u2d', 24 ,
'vobtrctx', .true., .false., 'daily', " , "
bn_u3d = 'amm12_bdyU_u3d', 24 ,
'vozocrtx', .true., .false., 'daily', " , "
bn_v3d = 'amm12_bdyV_u3d', 24 ,
'vomercrty', .true., .false., 'daily', " , "
bn_tem = 'amm12_bdyT_tra', 24 ,
'votemper', .true., .false., 'daily', " , "
bn_sal = 'amm12_bdyT_tra', 24 ,
'vosaline', .true., .false., 'daily', " , "
cn_dir = 'bdydt/'
ln_key_vagrif = .false.
/
!-----
&nambdy_tide ! tidal forcing at open boundaries
!-----
fildtide = 'bdydt/amm12_bdytide_' ! file name
root of tidal forcing files
tide_cpt(1) = 'Q1' ! names of tidal components used
tide_cpt(2) = 'O1' ! names of tidal components used
tide_cpt(3) = 'P1' ! names of tidal components used
tide_cpt(4) = 'S1' ! names of tidal components used
tide_cpt(5) = 'K1' ! names of tidal components used
tide_cpt(6) = 'N2' ! names of tidal components used
tide_cpt(7) = 'M2' ! names of tidal components used
tide_cpt(8) = 'N2' ! names of tidal components used
tide_cpt(9) = 'N2' ! names of tidal components used
tide_cpt(10) = 'M2' ! names of tidal components used
tide_cpt(11) = 'L2' ! names of tidal components used
tide_cpt(12) = 'T2' ! names of tidal components used
tide_cpt(13) = 'S2' ! names of tidal components used
tide_cpt(14) = 'K2' ! names of tidal components used
tide_cpt(15) = 'M4' ! names of tidal components used
tide_speed(1) = 13.398661 ! phase speeds of tidal
components (deg/hour)
tide_speed(2) = 13.943036 ! phase speeds of tidal
components (deg/hour)
tide_speed(3) = 14.958932 ! phase speeds of tidal
components (deg/hour)
tide_speed(4) = 15.000001 ! phase speeds of tidal
components (deg/hour)
tide_speed(5) = 15.041069 ! phase speeds of tidal
components (deg/hour)
tide_speed(6) = 27.895355 ! phase speeds of tidal
components (deg/hour)
tide_speed(7) = 27.968210 ! phase speeds of tidal
components (deg/hour)
tide_speed(8) = 28.439730 ! phase speeds of tidal
components (deg/hour)
tide_speed(9) = 28.512585 ! phase speeds of tidal
components (deg/hour)
tide_speed(10) = 28.984106 ! phase speeds of tidal
components (deg/hour)
tide_speed(11) = 29.528479 ! phase speeds of tidal
components (deg/hour)
tide_speed(12) = 29.958935 ! phase speeds of tidal
components (deg/hour)
tide_speed(13) = 30.000002 ! phase speeds of tidal
components (deg/hour)
tide_speed(14) = 30.082138 ! phase speeds of tidal
components (deg/hour)
tide_speed(15) = 57.968212 ! phase speeds of tidal
components (deg/hour)
ln_tide_date = .true. ! adjust tidal
harmonics for start date of run
/
!!=====
!! ** Bottom boundary condition **
!!=====
!! nambfr bottom friction
!! nambbc bottom temperature boundary condition
!! namdbl bottom boundary layer scheme ("key_trabbl")
!!=====
!
!-----
&nambfr ! bottom friction
!-----
nn_bfr = 2 ! type of bottom friction : = 0 :
free slip, = 1 : linear friction = 2 : nonlinear friction
!
! rn_bfri1 = 4.e-4 ! bottom drag coefficient (linear
case)
! rn_bfri2 = 1.e-3 ! bottom drag coefficient (non
linear case)

```

```

rn_bfeb2 = 2.5e-3 ! bottom turbulent kinetic energy
background (m2/s2)
ln_bfr2d = .false. ! horizontal variation of the
bottom friction coef (read a 2D mask file )
rn_bfrlen = 50. ! local multiplying factor of bfr
(ln_bfr2d=T)
ln_bfrimp = .true. ! implicit bottom friction
(requires ln_zdfexp = .false. if true)
/
!-----
&nambbc ! bottom temperature boundary condition
!-----
ln_trabbc = .false. ! Apply a geothermal heating at the
ocean bottom
nn_geoflx = 2 ! geothermal heat flux: = 0 no flux
! = 1 constant flux
! = 2 variable flux (read in geothermal_heating.nc in
mW/m2)
rn_geoflx_cst = 86.4e-3 ! Constant value of geothermal heat
flux [W/m2]
/
!-----
&namtbl ! bottom boundary layer scheme
!-----
nn_bbl_ldf = 1 ! diffusive bbl (=1) or not (=0)
nn_bbl_adv = 1 ! advective bbl (=1/2) or not (=0)
rn_ahtbbl = 1000. ! lateral mixing coefficient in the
bbl [m2/s]
rn_gambbl = 10. ! advective bbl coefficient
[s]
/
!!-----
!! Tracer ( T & S ) namelists
!!-----
!! nameos equation of state
!! namtra_adv advection scheme
!! namtra_ldf lateral diffusion scheme
!! namtra_dmp T & S newtonian damping
!!-----
&nameos ! ocean physical parameters
!-----
nn_eos = 0 ! type of equation of state and
Brunt-Vaisala frequency
! = 0, UNESCO (formulation of Jackett and McDougall
(1994) and of McDougall (1987) )
! = 1, linear: rho(T) = rau0 * ( 1.028 - ralpha * T )
! = 2, linear: rho(T,S) = rau0 * ( rbeta * S - ralpha *
T )
rn_alpha = 2.0e-4 ! thermal expansion coefficient
(nn_eos= 1 or 2)
rn_beta = 7.7e-4 ! saline expansion coefficient
(nn_eos= 2)
/
!-----
&namtra_adv ! advection scheme for tracer
!-----
ln_traadv_cen2 = .false. ! 2nd order centered scheme
ln_traadv_tv2 = .true. ! TVD scheme
ln_traadv_muscl = .false. ! MUSCL scheme
ln_traadv_muscl2 = .false. ! MUSCL2 scheme + cen2 at
boundaries
ln_traadv_ubs = .false. ! UBS scheme
ln_traadv_qck = .false. ! QUICKEST scheme
/
!-----
&namtra_ldf ! lateral diffusion scheme for tracer
!-----
! ! Type of the operator :
ln_traldf_lap = .true. ! laplacian operator
ln_traldf_bilap = .false. ! bilaplacian operator
! ! Direction of action :
ln_traldf_level = .false. ! iso-level
ln_traldf_hor = .false. ! horizontal (geopotential)
(require "key_ldfslp" when ln_sco=T)
ln_traldf_iso = .true. ! iso-neutral
("key_ldfslp")
ln_traldf_grif = .false. ! griffies skew flux
formulation (require "key_ldfslp")
ln_traldf_gdia = .false. ! griffies operator strfn
diagnostics (require "key_ldfslp")
ln_triad_iso = .false. ! griffies operator calculates
triads twice => pure lateral mixing in ML (require "key_ldfslp")
ln_botmix_grif = .false. ! griffies operator with
lateral mixing on bottom (require "key_ldfslp")
! Coefficient
rn_aht_0 = 300. ! horizontal eddy diffusivity
for tracers [m2/s]
rn_ahtb_0 = 0. ! background eddy diffusivity
for ldf_iso [m2/s]
rn_aeiv_0 = 0. ! eddy induced velocity
coefficient [m2/s] (require "key_traldf_eiv")
/
!-----
&namtra_dmp ! tracer: T & S newtonian damping
!-----
ln_tradmp = .true. ! add a damping term (T) or not
(F)
nn_hdmp = -2 ! horizontal shape =-1, damping in
Med and Red Seas only
! =XX, damping poleward of XX degrees
(XX>0)
! + F(distance-to-coast) + Red and Med
Seas
nn_zdmp = 0 ! vertical shape =0 damping
throughout the water column
! =1 no damping in the mixing layer (kz
criteria)
! =2 no damping in the mixed layer (rho
criteria)
rn_surf = 50. ! surface time scale of damping
[days]
rn_bot = 360. ! bottom time scale of damping
[days]
rn_dep = 800. ! depth of transition between
rn_surf and rn_bot [meters]
nn_file = 0 ! create a damping.coef NetCDF
file (=1) or not (=0)
ln_dmpmask = .false. ! Read dmp_mask.nc file when T
(between 0 and 1 )
rn_timsk = 730. ! Time scale used for dmp_mask
/
!!-----
!! *** Dynamics namelists ***
!!-----
!! namdyn_adv formulation of the momentum advection
!! namdyn_vor advection scheme
!! namdyn_hpg hydrostatic pressure gradient
!! namdyn_spg surface pressure gradient (CPP
key only)
!! namdyn_ldf lateral diffusion scheme
!!-----
&namdyn_adv ! formulation of the momentum advection
!-----
ln_dynadv_vec = .true. ! vector form (T) or flux form (F)
ln_dynadv_cen2 = .false. ! flux form - 2nd order centered
scheme
ln_dynadv_ubs = .false. ! flux form - 3rd order UBS
scheme
/
!-----
&namdyn_vor ! option of physics/algorithm (not control
by CPP keys)
!-----
ln_dynvor_ene = .false. ! enstrophy conserving scheme
ln_dynvor_ens = .false. ! energy conserving scheme
ln_dynvor_mix = .false. ! mixed scheme
ln_dynvor_ee = .true. ! energy & enstrophy scheme
/
!-----
&namdyn_hpg ! Hydrostatic pressure gradient option
!-----
ln_hpg_zco = .false. ! z-coordinate - full steps
ln_hpg_zps = .true. ! z-coordinate - partial steps
(interpolation)
ln_hpg_sco = .false. ! s-coordinate (standard jacobian
formulation)
ln_hpg_djc = .false. ! s-coordinate (Density Jacobian
with Cubic polynomial)
ln_hpg_prj = .false. ! s-coordinate (Pressure Jacobian
scheme)
ln_dynhpg_imp = .true. ! time stepping: semi-implicit time
scheme(implicit)
! centered time scheme (F)
/
!-----
&namdyn_spg ! surface pressure gradient (CPP key
only)
!-----
! ! explicit free surface
("key_dynspg_exp")
! ! filtered free surface
("key_dynspgflt")
! ! split-explicit free surface
("key_dynspgts")

```

```

!-----
&namdyn_ldf ! lateral diffusion on momentum
!-----
!
! Type of the operator :
ln_dynldf_lap = .false. ! laplacian operator
ln_dynldf_bilap = .true. ! bilaplacian operator
!
! Direction of action :
ln_dynldf_level = .false. ! iso-level
ln_dynldf_hor = .true. ! horizontal (geopotential)
(require "key_ldfslp" in s-coord.)
ln_dynldf_iso = .false. ! iso-neutral
"key_ldfslp")
!
! Coefficient
rn_ahm_0_lap = 0. ! horizontal laplacian eddy
viscosity [m2/s]
rn_ahmb_0 = 0. ! background eddy viscosity
for ldf_iso [m2/s]
rn_ahm_0_blp = -1.5e11 ! horizontal bilaplacian eddy
viscosity [m4/s]
/
!-----
!! Tracers & Dynamics vertical physics namelists
!-----
!! namzdf vertical physics
!! namzdf_ric richardson number dependent vertical
mixing ("key_zdfric")
!! namzdf_tke TKE dependent vertical mixing
!! namzdf_kpp KPP dependent vertical mixing
!! namzdf_ddm double diffusive mixing parameterization
("key_zdfddm")
!! namzdf_tmh tidal mixing parameterization
!-----
&namzdf ! vertical physics
!-----
rn_avm0 = 1.e-4 ! vertical eddy viscosity [m2/s]
(background Kz if not "key_zdfcst")
rn_avt0 = 1.e-5 ! vertical eddy diffusivity [m2/s]
(background Kz if not "key_zdfcst")
nn_avb = 0 ! profile for background avt & avm
(=1) or not (=0)
nn_havtb = 1 ! horizontal shape for avtb (=1) or
not (=0)
ln_zdfevd = .true. ! enhanced vertical diffusion (evd)
(T) or not (F)
nn_evdm = 1 ! evd apply on tracer (=0) or on
tracer and momentum (=1)
rn_avevd = 10. ! evd mixing coefficient [m2/s]
ln_zdfnpc = .false. ! Non-Penetrative Convective
algorithm (T) or not (F)
nn_npc = 1 ! frequency of application of
npc
nn_npcp = 365 ! npc control print frequency
ln_zdfexp = .false. ! time-stepping: split-explicit (T)
or implicit (F) time stepping
nn_zdfexp = 3 ! number of sub-timestep for
ln_zdfexp=T
/
!-----
&namzdf_ric ! richardson number dependent vertical
diffusion ("key_zdfric")
!-----
rn_avmri = 100.e-4 ! maximum value of the vertical
viscosity
rn_alp = 5. ! coefficient of the parameterization
nn_ric = 2 ! coefficient of the parameterization
rn_ekmfc = 0.7 ! Factor in the Ekman depth
Equation
rn_mldmin = 1.0 ! minimum allowable mixed-layer
depth estimate (m)
rn_mldmax = 1000.0 ! maximum allowable mixed-layer
depth estimate (m)
rn_wtmix = 10.0 ! vertical eddy viscosity coeff
[m2/s] in the mixed-layer
rn_vvmix = 10.0 ! vertical eddy diffusion coeff
[m2/s] in the mixed-layer
ln_mldw = .true. ! Flag to use or not the mixed
layer depth param.
/
!-----
&namzdf_tke ! turbulent eddy kinetic dependent vertical
diffusion ("key_zdf_tke")
!-----
rn_ediff = 0.1 ! coef. for vertical eddy coef.
(avt=rn_ediff*mxl*sqrt(e))
rn_ediss = 0.7 ! coef. of the Kolmogoroff
dissipation
rn_ebb = 60 ! coef. of the surface input of the
(67.83 now usual value)
rn_ebbice = 3.75 ! coef. of the surface input of the
under ice
nn_havti = 1 ! horizontal shape for avtb (=1) or
not (=0) under ice
rn_emin = 1.e-6 ! minimum value of the [m2/s2]
rn_emin0 = 1.e-4 ! surface minimum value of the
[m2/s2]
rn_bshear = 1.e-20 ! background shear (>0)
nn_nmix = 3 ! mixing length: = 0 bounded by the
distance to surface and bottom
!
! = 1 bounded by the local vertical scale
factor
!
! = 2 first vertical derivative of mixing
length bounded by 1
!
! = 3 as =2 with distinct dissipative an
mixing length scale
nn_pdl = 1 ! Prandtl number function of
richarson number (=1, avt=pd1(Ri)*avm) or not (=0, avt=avm)
ln_mx10 = .true. ! surface mixing length scale =
F(wind stress) (T) or not (F)
rn_mx10 = 0.01 ! surface buoyancy length scale
minimum value (0.04)
ln_lc = .true. ! Langmuir cell parameterisation
(Axell 2002)
rn_kkey_zdf_tke = 0.15 ! coef. associated to Langmuir
cells ("key_zdfkpp")
nn_etau = 1 ! penetration of tke below the
mixed layer (ML) due to internal & inertial waves
!
! ("key_zdf_tke")
!
! = 1 add a tke source below the ML
!
! = 2 add a tke source just at the base of the ML
!
! = 3 as = 1 applied on HF part of the stress
("key_coupled")
rn_efr = 0.05 ! fraction of surface tke value
which penetrates below the ML (nn_etau=1 or 2)
nn_htau = 1 ! type of exponential decrease of
tke penetration below the ML
!
! = 0 constant 10 m length scale
!
! = 1 0.5m at the equator to 30m poleward of 40
degrees
/
!-----
&namzdf_kpp ! K-Profile Parameterization dependent
vertical mixing ("key_zdfkpp", and optionally:
!-----
"key_kppcustom" or
"key_kpplktb")
ln_kpprimix = .true. ! shear instability mixing
rn_difmiw = 1.0e-04 ! constant internal wave viscosity
[m2/s]
rn_difsiw = 0.1e-04 ! constant internal wave diffusivity
[m2/s]
rn_rinfty = 0.8 ! local Richardson Number limit for
shear instability
rn_difri = 0.0050 ! maximum shear mixing at Rig = 0
[m2/s]
rn_bvsqcon = -0.01e-07 ! Brunt-Vaisala squared for maximum
convection [1/s2]
rn_difcon = 1. ! maximum mixing in interior
convection [m2/s]
nn_avb = 0 ! horizontal averaged (=1) or not
(=0) on avt and avm
nn_ave = 1 ! constant (=0) or profile (=1)
background on avt
/
!-----
&namzdf_gls ! GLS vertical diffusion
("key_zdfgl")
!-----
rn_emin = 1.e-6 ! minimum value of e [m2/s2]
rn_epsmin = 1.e-12 ! minimum value of eps [m2/s3]
ln_length_lim = .true. ! limit on the dissipation rate
under stable stratification (Galperin et al., 1988)
rn_clim_galp = 0.53 ! galperin limit
ln_crban = .true. ! Use Craig & Banner (1994) surface
wave mixing parameterisation
ln_sigpsi = .true. ! Activate or not Burchard 2001
mods on psi schmidt number in the wb case
rn_crban = 100. ! Craig and Banner 1994 constant
for wb tke flux
rn_charn = 70000. ! Charnock constant for wb induced
roughness length
nn_tkebc_surf = 1 ! surface tke condition (0/1=Dir/Neum/Dir
Mellor-Blumberg)
nn_tkebc_bot = 1 ! bottom tke condition (0/1=Dir/Neum)
nn_psbcb_surf = 1 ! surface psi condition (0/1=Dir/Neum/Dir
Mellor-Blumberg)
nn_psbcb_bot = 1 ! bottom psi condition (0/1=Dir/Neum)

```



```

ln_diaptr = .false. ! Poleward heat and salt transport
(T) or not (F)
ln_diaznl = .false. ! Add zonal means and meridional
stream functions
ln_subbas = .false. ! Atlantic/Pacific/Indian basins
computation (T) or not
! (orca configuration only, need input basins mask file
named "subbasins.nc"
ln_ptrcomp = .false. ! Add decomposition : overturning
nn_fptr = 1 ! Frequency of ptr computation
[time step]
nn_fwri = 15 ! Frequency of ptr outputs [time
step]
/
!-----
&namhsb ! Heat and salt budgets
!-----
ln_diahsb = .false. ! check the heat and salt budgets
(T) or not (F)
/
!-----
&nam_diaharm ! Harmonic analysis of tidal constituents
('key_diaharm')
!-----
nit000_han = 1 ! First time step used for harmonic
analysis
nitend_han = 75 ! Last time step used for harmonic
analysis
nstep_han = 15 ! Time step frequency for harmonic
analysis
tname(1) = 'M2' ! Name of tidal constituents
tname(2) = 'K1'
/
!-----
&namdct ! transports through sections
!-----
nn_dct = 15 ! time step frequency for transports
computing
nn_dctwri = 15 ! time step frequency for transports
writing
nn_secdebug = 112 ! 0 : no section to debug
! -1 : debug all section
! 0 < n : debug section number n
/
!!=====
!! *** Observation & Assimilation namelists ***
!!=====
!! namobs observation and model comparison
('key_diaobs')
!! nam_asminc assimilation increments
!!=====
!
&namobs ! observation usage switch
!-----
ln_t3d = .false. ! Logical switch for T profile
observations
ln_s3d = .false. ! Logical switch for S profile
observations
ln_ena = .false. ! Logical switch for ENACT insitu
data set
! ln_cor Logical
switch for Coriolis insitu data set
ln_prof = .false. ! Logical switch for feedback insitu
data set
ln_sla = .false. ! Logical switch for SLA observations
ln_sladt = .false. ! Logical switch for AVISO SLA data
ln_slafb = .false. ! Logical switch for feedback SLA
data
! ln_ssh Logical switch for SSH
observations
ln_sst = .false. ! Logical switch for SST observations
! ln_reysst Logical switch for Reynolds
observations
! ln_ghrsst Logical switch for GHRSSST
observations
ln_sstfb = .false. ! Logical switch for feedback SST
data
! ln_sss Logical switch for SSS
observations
! ln_seaice Logical switch for Sea Ice
observations
! ln_vel3d Logical switch for velocity
observations
! ln_velavcur Logical switch for velocity
daily av. cur.
! ln_velhrcur Logical switch for velocity
high freq. cur.
! ln_velavadcp Logical switch for velocity
daily av. ADCP
! ln_velhradcp Logical switch for velocity
high freq. ADCP
! ln_velfb Logical switch for feedback
velocity data
! ln_grid_global Global distribution of
observations
! ln_grid_search_lookup Logical switch for obs grid
search w/lookup table
! grid_search_file Grid search lookup file header
! enactfiles ENACT input observation file
names
! coriofiles Coriolis input observation
file name
! profbfiles: Profile feedback input
observation file name
profbfiles = 'profiles_01.nc'
! ln_prof_enatim Enact feedback input time
setting switch
! slafilesact Active SLA input observation
file name
! slafilespas Passive SLA input observation
file name
! slafbfiles: Feedback SLA input
observation file name
slafbfiles = 'sla_01.nc'
! sstfiles GHRSSST input observation file
name
! sstfbfiles: Feedback SST input
observation file name
sstfbfiles = 'sst_01.nc' 'sst_02.nc' 'sst_03.nc' 'sst_04.nc'
'sst_05.nc'
! seaicefiles Sea Ice input observation file
name
! velavcurfiles Vel. cur. daily av. input file
name
! velhvcurfiles Vel. cur. high freq. input
file name
! velavadcpfiles Vel. ADCP daily av. input file
name
! velhvadcpcfiles Vel. ADCP high freq. input
file name
! velfbfiles Vel. feedback input observation
file name
! dobsini Initial date in window
YYYYMMDD.HHMMSS
! dobsend Final date in window
YYYYMMDD.HHMMSS
! nldint Type of vertical interpolation
method ('key_asminc')
! n2dint Type of horizontal interpolation
method
! ln_nea Rejection of observations near
('key_diaobs')
land switch
nmsshc = 0 ! MSSH correction scheme
! mdtcorr MDT correction
! mdtcutoff MDT cutoff for computed
correction
ln_altbias = .false. ! Logical switch for alt bias
ln_ignmis = .true. ! Logical switch for ignoring
missing files
! endailyavtypes ENACT daily average types
ln_grid_global = .true.
ln_grid_search_lookup = .false.
/
!-----
&nam_asminc ! assimilation increments ('key_asminc')
!-----
ln_bkgwri = .false. ! Logical switch for writing out
background state
ln_trjwri = .false. ! Logical switch for writing out
state trajectory
ln_trainc = .false. ! Logical switch for applying tracer
increments
ln_dyninc = .false. ! Logical switch for applying
velocity increments
ln_sshinc = .false. ! Logical switch for applying SSH
increments
ln_asminc = .false. ! Logical switch for Direct
Initialization (DI)
ln_asmiau = .false. ! Logical switch for Incremental
Analysis Updating (IAU)
nitbkg = 0 ! Timestep of background in
[0,nitend-nit000-1]
nitdin = 0 ! Timestep of background for DI in
[0,nitend-nit000-1]
nitiaustr = 1 ! Timestep of start of IAU interval
in [0,nitend-nit000-1]

```

```

nitiaufin = 15          ! Timestep of end of IAU interval in
[0,nitend-nit000-1]
niaufn   = 0           ! Type of IAU weighting function
nittrjfrq = 0          ! Frequency of trajectory output for
4D-VAR
ln_salfix = .false.    ! Logical switch for ensuring that
the sa > salfixmin
salfixmin = -9999      ! Minimum salinity after applying
the increments
nn_divdmp = 0          ! Number of iterations of divergence
damping operator
/
!-----
&namsbc_wave ! External fields from wave model
!-----
!           ! file name ! frequency (hours) ! variable
! time interp. ! clim ! 'yearly' ! weights ! rotation !
!           !           ! (if <0 months) ! name
! (logical) ! (T/F) ! 'monthly' ! filename ! pairing !
sn_cdg      = 'cdg_wave',      1      , 'drag_coeff'
, .true.    , .false. , 'daily' , "      , "
!
cn_dir_cdg = './' ! root directory for the location of
drag coefficient files
/
/-----
&namdyn_nept ! Neptune effect (simplified: lateral and
vertical diffusions removed)
!-----
! Suggested lengthscale values are those of Eby & Holloway
(1994) for a coarse model
ln_neptsimp = .false. ! yes/no use simplified neptune
ln_smooth_neptvel = .false. ! yes/no smooth zunep, zvnep
rn_tslse     = 1.2e4 ! value of lengthscale L at the
equator
rn_tslsp     = 3.0e3 ! value of lengthscale L at the
pole
! Specify whether to ramp down the Neptune velocity in
shallow
! water, and if so the depth range controlling such ramping
down
ln_neptramp = .true. ! ramp down Neptune velocity in
shallow water
rn_htrmin   = 100.0 ! min. depth of transition
range
rn_htrmax   = 200.0 ! max. depth of transition
range
/

```



```

&nampisprod      ! parameters for phytoplankton growth
!.....
pislope = 2.      ! P-I slope
pislope2 = 2.    ! P-I slope for diatoms
excret = 0.05    ! excretion ratio of phytoplankton
excret2 = 0.05   ! excretion ratio of diatoms
ln_newprod = .true. ! Enable new parame. of production (T/F)
bresp = 0.00333 ! Basal respiration rate
chlcnm = 0.033  ! Minimum Chl/C in nanophytoplankton
chldm = 0.05    ! Minimum Chl/C in diatoms
chlcmn = 0.0033 ! Maximum Chl/c in phytoplankton
fecnm = 40E-6   ! Maximum Fe/C in nanophytoplankton
fecdm = 40E-6   ! Minimum Fe/C in diatoms
grosip = 0.151  ! mean Si/C ratio
/
!.....
&nampismort      ! parameters for phytoplankton sinks
!.....
wchl = 0.001    ! quadratic mortality of phytoplankton
wchld = 0.02    ! maximum quadratic mortality of diatoms
mprat = 0.01    ! phytoplankton mortality rate
mprat2 = 0.01   ! Diatoms mortality rate
mpratm = 0.01   ! Phytoplankton minimum mortality rate
/
!.....
&nampismes       ! parameters for mesozooplankton
!.....
part2 = 0.75    ! part of calcite not dissolved in mesozoo guts
grazrat2 = 0.7  ! maximal mesozoo grazing rate
resrat2 = 0.005 ! exsudation rate of mesozooplankton
mzrat2 = 0.03   ! mesozooplankton mortality rate
xprefc = 1.     ! zoo preference for phyto
xprefp = 0.3    ! zoo preference for POC
xprefz = 1.     ! zoo preference for zoo
xprefpoc = 0.3  ! zoo preference for poc
xthresh2zoo = 1E-8 ! zoo feeding threshold for mesozooplankton
xthresh2dia = 1E-8 ! diatoms feeding threshold for mesozooplankton
xthresh2phy = 1E-8 ! nanophyto feeding threshold for mesozooplankton
xthresh2poc = 1E-8 ! poc feeding threshold for mesozooplankton
xthresh2 = 2E-7  ! Food threshold for grazing
xkgraz2 = 20.E-6 ! half sturation constant for meso grazing
epshe2 = 0.3    ! Efficiency of Mesozoo growth
sigma2 = 0.6    ! Fraction of mesozoo excretion as DOM
unass2 = 0.3    ! non assimilated fraction of P by mesozoo
grazflux = 2.e3 ! flux-feeding rate
/
!.....
&nampiszoo       ! parameters for microzooplankton
!.....
part = 0.5      ! part of calcite not dissolved in microzoo gutsa
grazrat = 3.0   ! maximal zoo grazing rate
resrat = 0.03   ! exsudation rate of zooplankton
mzrat = 0.001   ! zooplankton mortality rate
xpref2c = 0.1   ! Microzoo preference for POM
xpref2p = 1.    ! Microzoo preference for Nanophyto
xpref2d = 0.5   ! Microzoo preference for Diatoms
xthreshdia = 1.E-8 ! Diatoms feeding threshold for microzooplankton
xthreshphy = 1.E-8 ! Nanophyto feeding threshold for microzooplankton
xthreshpoc = 1.E-8 ! POC feeding threshold for microzooplankton
xthresh = 2.E-7  ! Food threshold for feeding
xkgraz = 20.E-6 ! half sturation constant for grazing
epshe = 0.3     ! Efficiency of microzoo growth
sigma1 = 0.6    ! Fraction of microzoo excretion as DDM
unass = 0.3     ! non assimilated fraction of phyto by zoo
/
!.....
&nampisrem       ! parameters for remineralization
!.....
xremik = 0.25   ! remineralization rate of DDC
xremip = 0.025 ! remineralisation rate of POC
nitrif = 0.05  ! NH4 nitrification rate
xsirem = 0.003 ! remineralization rate of Si
xsiremlab = 0.025 ! fast remineralization rate of Si
xsilab = 0.31  ! Fraction of labile biogenic silica
xlam1 = 0.005 ! scavenging rate of Iron
oxymin = 1.E-6 ! Half-saturation constant for anoxia
ligand = 0.6E-9 ! Ligands concentration
/
!.....
&nampiscal       ! parameters for Calcite chemistry
!.....
kdca = 6.       ! calcite dissolution rate constant (1/time)
nca = 1.        ! order of dissolution reaction (dimensionless)
/
!.....
&nampissed       ! parameters for inputs deposition
!.....
! ! file name ! frequency (hours) ! variable ! time interp. ! clim ! 'yearly' / ! weights ! rotation !
! ! ! (if <0 months) ! name ! (logical) ! (T/F) ! 'monthly' ! filename ! pairing !

```

```

sn_dust      = 'dust.orca'      , -1      , 'dust'      , .true.      , .true.      , 'yearly'    , "      , "
sn_riverdic  = 'river.orca'    , -12     , 'riverdic'  , .false.     , .true.     , 'yearly'    , "      , "
sn_riverdoc  = 'river.orca'    , -12     , 'riverdoc'  , .false.     , .true.     , 'yearly'    , "      , "
sn_ndepo     = 'ndeposition.orca', -12     , 'ndep'      , .false.     , .true.     , 'yearly'    , "      , "
sn_ironseed  = 'bathy.orca'    , -12     , 'bathy'     , .false.     , .true.     , 'yearly'    , "      , "
!
cn_dir       = './'            ! root directory for the location of the dynamical files
ln_dust      = .true.          ! boolean for dust input from the atmosphere
ln_river     = .true.          ! boolean for river input of nutrients
ln_ndepo     = .false.         ! boolean for atmospheric deposition of N
ln_ironseed  = .true.          ! boolean for Fe input from sediments
sedfeinput   = 1E-9           ! Coastal release of Iron
dustsolub    = 0.02           ! Solubility of the dust
wdust        = 2.0            ! Dust sinking speed
nitrfix      = 1E-7           ! Nitrogen fixation rate
diazolight   = 50.            ! Diazotrophs sensitivity to light (W/m2)
concfediaz   = 1.E-10         ! Diazotrophs half-saturation Cste for Iron
/
!=====
&namiskrpk   ! Kriest parameterization : parameters "key_kriest"
!=====
xkr_eta      = 1.17           ! Sinking exponent
xkr_zeta     = 2.28           ! N content exponent
xkr_mass_min = 0.0002        ! Minimum mass for Aggregates
xkr_mass_max = 1.            ! Maximum mass for Aggregates
/
!=====
&namiskrks   ! Kriest parameterization : size classes "key_kriest"
!=====
xkr_sfact    = 942.          ! Sinking factor
xkr_stick    = 0.5           ! Stickiness
xkr_nnano    = 2.337         ! Nbr of cell in nano size class
xkr_ndiat    = 3.718         ! Nbr of cell in diatoms size class
xkr_nmeso    = 7.147         ! Nbr of cell in mesozoo size class
xkr_naggr    = 9.877         ! Nbr of cell in aggregates size class
/
!=====
&namisdia    ! additional 2D/3D tracers diagnostics
!=====
! name ! title of the field ! units !
! ! ! ! !
pisdia2d(1) = 'Cflx' , 'DIC flux' , 'molC/m2/s' ,
pisdia2d(2) = 'Oflx' , 'Oxygen flux' , 'molC/m2/s' ,
pisdia2d(3) = 'Kg' , 'Gas transfer' , 'mol/m2/s/uatm' ,
pisdia2d(4) = 'Delc' , 'Delta CO2' , 'uatm' ,
pisdia2d(5) = 'PMO' , 'POC export' , 'molC/m2/s' ,
pisdia2d(6) = 'PMO2' , 'GOC export' , 'molC/m2/s' ,
pisdia2d(7) = 'ExpFe1' , 'Nano iron export' , 'molFe/m2/s' ,
pisdia2d(8) = 'ExpFe2' , 'Diatoms iron export' , 'molFe/m2/s' ,
pisdia2d(9) = 'ExpSi' , 'Silicate export' , 'molSi/m2/s' ,
pisdia2d(10) = 'ExpCaCO3' , 'Calcite export' , 'molC/m2/s' ,
pisdia2d(11) = 'heup' , 'euphotic layer depth' , 'm' ,
pisdia2d(12) = 'Fedep' , 'Iron dep' , 'molFe/m2/s' ,
pisdia2d(13) = 'Nfix' , 'Nitrogen Fixation' , 'molN/m2/s' ,
pisdia3d(1) = 'PH' , 'PH' , '-' ,
pisdia3d(2) = 'CO3' , 'Bicarbonates' , 'mol/l' ,
pisdia3d(3) = 'CO3sat' , 'CO3 saturation' , 'mol/l' ,
pisdia3d(4) = 'PAR' , 'light penetration' , 'W/m2' ,
pisdia3d(5) = 'PPPHY' , 'Primary production of nanophyto' , 'molC/m3/s' ,
pisdia3d(6) = 'PPPHY2' , 'Primary production of diatoms' , 'molC/m3/s' ,
pisdia3d(7) = 'PPNEWN' , 'New Primary production of nano' , 'molC/m3/s' ,
pisdia3d(8) = 'PPNEWD' , 'New Primary production of diat' , 'molC/m3/s' ,
pisdia3d(9) = 'PBSi' , 'Primary production of Si diatoms' , 'molSi/m3/s' ,
pisdia3d(10) = 'PFen' , 'Primary production of nano iron' , 'molFe/m3/s' ,
pisdia3d(11) = 'PFED' , 'Primary production of diatoms iron' , 'molFe/m3/s' ,
/
!=====
&namisdmp    ! Damping
!=====
ln_pisdmp    = .false.       ! Relaxation fo some tracers to a mean value
nn_pisdmp    = 2400         ! Frequency of Relaxation
/
!=====
&namismass   ! Mass conservation
!=====
ln_check_mass = .true.      ! Check mass conservation
/

```



```

sn_trcdta(6) = 'data_CaCO3_PGDOC_BSi' , -1 , 'POC' , .false. , .true. , 'yearly' , " , "
sn_trcdta(7) = 'data_DIC_alka_O2_Si' , -1 , 'Si' , .false. , .true. , 'yearly' , " , "
sn_trcdta(8) = 'data_PHY_PHY2_Z00_Z002' , -1 , 'PHY' , .false. , .true. , 'yearly' , " , "
sn_trcdta(9) = 'data_PHY_PHY2_Z00_Z002' , -1 , 'Z00' , .false. , .true. , 'yearly' , " , "
sn_trcdta(10) = 'data_CaCO3_PGDOC_BSi' , -1 , 'DOC' , .false. , .true. , 'yearly' , " , "
sn_trcdta(11) = 'data_PHY_PHY2_Z00_Z002' , -1 , 'PHY2' , .false. , .true. , 'yearly' , " , "
sn_trcdta(12) = 'data_PHY_PHY2_Z00_Z002' , -1 , 'Z002' , .false. , .true. , 'yearly' , " , "
sn_trcdta(13) = 'data_BSDNFe_DSi' , -1 , 'DSi' , .false. , .true. , 'yearly' , " , "
sn_trcdta(14) = 'data_P04_N03_NH4_Fer' , -1 , 'Fer' , .false. , .true. , 'yearly' , " , "
sn_trcdta(15) = 'data_BSDNFe_DSi' , -1 , 'BFe' , .false. , .true. , 'yearly' , " , "
sn_trcdta(16) = 'data_CaCO3_PGDOC_BSi' , -1 , 'GOC' , .false. , .true. , 'yearly' , " , "
sn_trcdta(17) = 'data_BSDNFe_DSi' , -1 , 'SFe' , .false. , .true. , 'yearly' , " , "
sn_trcdta(18) = 'data_BSDNFe_DSi' , -1 , 'DFe' , .false. , .true. , 'yearly' , " , "
sn_trcdta(19) = 'data_CaCO3_PGDOC_BSi' , -1 , 'BSi' , .false. , .true. , 'yearly' , " , "
sn_trcdta(20) = 'data_BSDNFe_DSi' , -1 , 'NFe' , .false. , .true. , 'yearly' , " , "
sn_trcdta(21) = 'data_NCHL_DCHL' , -1 , 'NCHL' , .false. , .true. , 'yearly' , " , "
sn_trcdta(22) = 'data_NCHL_DCHL' , -1 , 'DCHL' , .false. , .true. , 'yearly' , " , "
sn_trcdta(23) = 'data_P04_N03_NH4_Fer' , -1 , 'N03' , .false. , .true. , 'yearly' , " , "
sn_trcdta(24) = 'data_P04_N03_NH4_Fer' , -1 , 'NH4' , .false. , .true. , 'yearly' , " , "
!
cn_dir = './' ! root directory for the location of the data files
rn_trfac(1) = 1.0 ! multiplicative factor
rn_trfac(2) = 1.0 ! - - - -
rn_trfac(3) = 1.0 ! - - - -
rn_trfac(4) = 1.0 ! - - - -
rn_trfac(5) = 1.0 ! - - - -
rn_trfac(6) = 1.0 ! - - - -
rn_trfac(7) = 1.0 ! - - - -
rn_trfac(8) = 1.0 ! - - - -
rn_trfac(9) = 1.0 ! - - - -
rn_trfac(10) = 1.0 ! - - - -
rn_trfac(11) = 1.0 ! - - - -
rn_trfac(12) = 1.0 ! - - - -
rn_trfac(13) = 1.0 ! - - - -
rn_trfac(14) = 1.0 ! - - - -
rn_trfac(15) = 1.0 ! - - - -
rn_trfac(16) = 1.0 ! - - - -
rn_trfac(17) = 1.0 ! - - - -
rn_trfac(18) = 1.0 ! - - - -
rn_trfac(19) = 1.0 ! - - - -
rn_trfac(20) = 1.0 ! - - - -
rn_trfac(21) = 1.0 ! - - - -
rn_trfac(22) = 1.0 ! - - - -
rn_trfac(23) = 1.0 ! - - - -
rn_trfac(24) = 1.0 ! - - - -
/
!-----
&namtrc_adv ! advection scheme for passive tracer
!-----
ln_trcadv_cen2 = .false. ! 2nd order centered scheme
ln_trcadv_tvd = .false. ! TVD scheme
ln_trcadv_muscl = .true. ! MUSCL scheme
ln_trcadv_muscl2 = .false. ! MUSCL2 scheme + cen2 at boundaries
ln_trcadv_ubs = .false. ! UBS scheme
ln_trcadv_qck = .false. ! QUICKEST scheme
/
!-----
&namtrc_ldf ! lateral diffusion scheme for passive tracer
!-----
ln_trcldf_diff = .true. ! performs lateral diffusion (T) or not (F)
!
! Type of the operator :
ln_trcldf_lap = .true. ! laplacian operator
ln_trcldf_bilap = .false. ! bilaplacian operator
! Direction of action :
ln_trcldf_level = .false. ! iso-level
ln_trcldf_hor = .false. ! horizontal (geopotential) (require "key_ldfslp" when ln_sco=T)
ln_trcldf_iso = .true. ! iso-neutral (require "key_ldfslp")
! Coefficient
rn_ahtrc_0 = 520. ! horizontal eddy diffusivity for tracers [m2/s]
rn_ahtrb_0 = 0. ! background eddy diffusivity for ldf_iso [m2/s]
/
!-----
&namtrc_zdf ! vertical physics
!-----
ln_trczdf_exp = .false. ! split explicit (T) or implicit (F) time stepping
nn_trczdf_exp = 3 ! number of sub-timestep for ln_trczdfexp=T
/
!-----
&namtrc_rad ! treatment of negative concentrations
!-----
ln_trcrad = .true. ! artificially correct negative concentrations (T) or not (F)
/
!-----
&namtrc_dmp ! passive tracer newtonian damping
!-----
ln_trcdmp = .false. ! add a damping term (T) or not (F)
nn_hdmp_tr = -1 ! horizontal shape =-1, damping in Med and Red Seas only
! =XX, damping poleward of XX degrees (XX>0)
! + F(distance-to-coast) + Red and Med Seas

```



```
obc_tracer(18) = ' DFe      ' ,      .true.      ,      1      ,      .true.
obc_tracer(19) = ' GSi      ' ,      .true.      ,      1      ,      .true.
obc_tracer(20) = ' NFe      ' ,      .true.      ,      1      ,      .true.
obc_tracer(21) = ' NCHL    ' ,      .true.      ,      1      ,      .true.
obc_tracer(22) = ' DCHL    ' ,      .true.      ,      1      ,      .true.
obc_tracer(23) = ' NO3     ' ,      .true.      ,      1      ,      .false.
obc_tracer(24) = ' NH4     ' ,      .true.      ,      1      ,      .true.
/
```

Annexe D

Publication Garnier et al. (2016)



Stochastic parameterizations of biogeochemical uncertainties in a 1/4° NEMO/PISCES model for probabilistic comparisons with ocean color data☆



F. Garnier*, J.-M. Brankart, P. Brasseur, E. Cosme

Univ. Grenoble Alpes, LGGE, F-38000 Grenoble, France
CNRS, LGGE, F38000 Grenoble, France

ARTICLE INFO

Article history:

Received 10 March 2015

Received in revised form 15 October 2015

Accepted 19 October 2015

Available online 2 November 2015

Keywords:

PISCES

Stochastic simulations

Ocean color data

Reliability

Data assimilation

ABSTRACT

In spite of recent advances, biogeochemical models are still unable to represent the full complexity of natural ecosystems. Their formulations are mainly based on empirical laws involving many parameters. Improving biogeochemical models therefore requires to properly characterize model uncertainties and their consequences. Subsequently, this paper investigates the potential of using random processes to simulate some uncertainties of the 1/4° coupled Physical–Biogeochemical NEMO/PISCES model of the North Atlantic ocean.

Starting from a deterministic simulation performed with the original PISCES formulation, we propose a generic method based on AR(1) random processes to generate perturbations with temporal and spatial correlations. These perturbations are introduced into the model formulations to simulate 2 classes of uncertainties: the uncertainties on biogeochemical parameters and the uncertainties induced by unresolved scales in the presence of non-linear processes. Using these stochastic parameterizations, a probabilistic version of PISCES is designed and a 60-member ensemble simulation is performed.

With respect to the simulation of chlorophyll, the relevance of the probabilistic configuration and the impacts of these stochastic parameterizations are assessed. In particular, it is shown that the ensemble simulation is in good agreement with the SeaWiFS ocean color data. Using these observations, the statistical consistency (reliability) of the ensemble is evaluated with rank histograms. Finally, the benefits expected from the probabilistic description of uncertainties (model error) are discussed in the context of future ocean color data assimilation.

© 2015 Elsevier B.V. All rights reserved.

1. Introduction

Estimating the primary production and its variability in the ocean has been the subject of advanced research over the past 40 years (Marra, 2015), motivated by the need to better understand the oceanic component of the global carbon cycle as well as the interconnections between biogeochemistry and marine ecosystems. In the early nineties, first attempts to use satellite ocean color sensors for estimating phytoplankton carbon fixation were based on productivity models driven by observed concentrations of chlorophyll-a (Behrenfeld and Falkowski, 1997). During the past 25 years, ocean color satellite missions provided a unique means to monitor almost continuously the surface signature of phytoplankton distribution in the upper ocean (Wilson, 2010). The combination of ocean color observations with three-dimensional biogeochemical general circulation models (BGCMs) through integrated approaches is a more recent but promising methodology, which aims at combining model information with data in space and time in order

to estimate global budgets of phytoplankton biomass and monitor its evolution for a variety of applications (Brasseur et al., 2009). However, such integrated approaches cannot be faithfully implemented as long as data sets and model solutions are not consistent statistically one with each other.

In the prospect of model-data comparisons or melding, the model reliability should be properly characterized in terms of uncertainty, or equivalently, “model error”. An appropriate description of uncertainties implies to define a correct separation between the processes that are properly resolved, poorly resolved or unresolved by the model (Brankart et al., 2015). The representation of uncertainties is intimately linked to the particular formulation of the considered biological model. Therefore, adopting an objective methodology to derive a statistical description of these model uncertainties is a major challenge for several types of scientific and operational applications (model assessment, validation, error characterization, data assimilation etc...). This is precisely the purpose of the present paper, which builds on several earlier publications that already addressed different aspects of model uncertainty based on ensemble integrations in the framework of eddy-permitting using the LOBSTER model (Béal et al., 2010; Doron et al., 2013; Fontana et al., 2013).

☆ This document is a collaborative effort from all authors.

* Corresponding author.

E-mail address: fgarnier@lgge.obs.ujf-grenoble.fr (F. Garnier).

In Béal et al. (2010), ensemble experiments performed with perturbed wind forcings enabled to characterize how uncertainties in the physical fields could impact the biological response in a North Atlantic, eddy-permitting model configuration. Using the same configuration, Doron et al. (2013) explored how the spatial dependence between the biological model parameters could explain the misfits between the model and the Globcolour data set, and reduce the uncertainty estimates in the selected parameters. Further, Fontana et al. (2013) implemented a sequential updating scheme to assimilate SeaWiFS (Sea-viewing Wide Field-of-view Scanner) ocean color products every 10 days and reconstruct the biogeochemical state of the North Atlantic during the 1998–2006 period. Finally, these 3 studies show that ocean color data are able to contribute to the reduction of biological uncertainties.

In this study, the more complex PISCES (Pelagic Interaction Scheme for Carbon and Ecosystem Studies) biological model (Aumont et al., 2003) coupled to NEMO is implemented in a $1/4^\circ$ North Atlantic configuration. As explain in Section 2, the PISCES formulation includes two distinct phytoplankton classes and deals with five different micro and macro nutrients to modulate the growth of biomass. In order to enable consistent model–data comparisons, we propose to investigate the potential of a generic approach to simulate appropriate biogeochemical uncertainties originating from several well identified sources and we evaluate their impacts on the model solutions. Introducing stochasticity into ecosystem models was first proposed by (Dowd and Meyer, 2003), by adding white noise processed in the model equations. The concept of stochastic perturbations for model parameter or model state integration was further developed and used in model/data comparisons, estimation and data assimilation studies (e.g.,: Gregg et al., 2009; Jones et al., 2010; Mattern et al., 2010; Dowd, 2011; Jones et al., 2013; Ciavatta et al., 2014).

In this paper, our approach relies on the introduction of stochastic representations of selected model uncertainties associated with some unresolved or poorly-resolved processes in the BGCMS. Random processes are used to generate ensemble members obtained by activating the stochastic processes in a set of independent model runs. Provided the stochasticity of the model is appropriately taken into account by making ensemble simulations, one should expect some skill for uncertainty prediction, though still conditional on model structure. The analysis of the ensemble spread is performed in terms of reliability (i.e., ability in producing pdfs in agreement with the associated observed pdfs) as defined by Candille et al. (2015) using the available ocean color observations. The relevance of this approach is further discussed in the prospect of an ocean color data assimilation system for ocean monitoring and forecasting.

This paper is structured as follows: in Section 2, we present the PISCES model and the North Atlantic configuration of a coupled NEMO/PISCES deterministic simulation; in section 3, are described the stochastic parameterizations proposed to simulate model uncertainties. Section 4 is then dedicated to the impact analysis of the stochastic parameterizations. In particular we explain that the mean effect of the perturbations implies to re-tune the deterministic initial set of parameters. Finally, in Section 5, we analyze the result of a 60-member ensemble simulation. The consistency of this simulation will be assessed from statistical comparisons with ocean color SeaWiFS observations.

2. The coupled physical–biogeochemical deterministic model

In this section, we describe the coupled model and show some general characteristics of a PISCES North Atlantic deterministic simulation (*detIni*) which does not include yet any stochastic process. This simulation is analyzed in terms of chlorophyll structures from comparisons with SeaWiFS ocean color satellite data (for the surface representation) and an in-situ data climatology (for the vertical distribution).

2.1. Model configuration

The deterministic model used in this study is based on a realistic coupled physical–biogeochemical configuration. This system is composed by a dynamical component, which defines the global oceanic configuration, and a biogeochemical part which computes the biological fields.

The physical model is based on the primitive equation ocean circulation model NEMO (Nucleus for European Modeling of the Ocean) covering the North Atlantic region from 20°S to 80°N and from 98°W to 23°E . The numerical grid has a $1/4^\circ$ resolution and includes 46 vertical levels. Such a resolution is considered as “eddy-permitting” as it enables some rough representation of ocean eddies which are key for primary production (Lévy et al., 2011; Oschlies and Garçon, 1998; Smith et al., 1996). The code is issued from the NEMO version 3.4 (www.nemo-ocean.eu) with the the LIM-2 sea ice parameterization (Louvain-la-Neuve sea ice model, Fichet and Maqueda, 1997). A complete description of this model can be find in Madec and the NEMO team (2012). The NEMO/OPA configuration is a duplicate of the NATLO25 North Atlantic configuration developed within the framework of the DRAKKAR project (Barnier et al., 2006). In order to include the years covered by the SeaWiFS data, this NATLO25 configuration is forced by ERA Interim atmospheric forcing fields instead of DFS (Drakkar Forcing sets) atmospheric fields (Brodeau et al., 2009). This configuration is inherited from several earlier studies in the context of ocean color data assimilation (Doron et al. (2011), Fontana et al. (2013)). The physical component was recently validated in Candille et al. (2015) to study the effects of unresolved scales on the model dynamics in the Gulf Stream region. In addition, a realistic dynamical adjustment of the ocean state is obtained after a 13 years physical spin-up (1989–2002) starting from the Levitus climatology for temperature and salinity (Levitus et al., 1998).

The ocean circulation model is coupled with the biogeochemical model PISCES-v2 (Aumont et al., 2015). Originally designed to study the global carbon biogeochemistry, PISCES has a complexity level which is suitable to represent the biological behavior in geographical provinces with various biogeochemical regimes (Aumont and Bopp, 2006; Bopp et al., 2005). In this context, one very important feature is the explicit limitation of phytoplankton growth by iron (Aumont et al., 2003), known as a fundamental micro nutrient element for photosynthesis (Geider and La Roche, 1994; Raven et al., 1999) with for instance its role as a limiting factor for the diazotrophs N_2 fixation (Falkowski, 1997). External inputs of iron in the ocean are specified using explicit forcings (Aumont and Bopp, 2006) which reproduces a supply of iron coming from rivers, continental-shelf sediments and from the atmospheric deposition of mineral dust (Tagliabue et al., 2008). PISCES includes 24 biogeochemical variables which can be grouped into 4 main compartments (Fig. 1). The phytoplankton compartment (PHY) is represented by two different classes, the nanophytoplankton (Nano) and the diatoms (Dia) with their production limited by the availability of five nutrients: iron (Fe), silicate (Si), phosphate (PO_4), ammonium (NH_4) and nitrate (NO_3). In order to simulate various biological behaviors, the phytoplankton compartment contains a formulation to fully predict the ratios $\frac{\text{Chl}}{\text{C}}$ (NCHL and DCHL state variables), $\frac{\text{Si}}{\text{C}}$ (Dsi state variable) and $\frac{\text{Fe}}{\text{C}}$ (NFe and DFe state variables). These ratios are therefore supposed to vary while other Redfield ratios between C, N and P are kept constant in accordance to the values of Takashi et al. (1985). The zooplankton compartment (ZOO) contain 2 species, the mesozooplankton (Meso) and the microzooplankton (Micro). At last, the detritus compartment is divided in 2 pools: the dissolved organic carbon (DOC) and the particulate organic matters including the small particules (sPOM) and the big particules (bPOM) which mostly differ by their sinking velocity. This separation is necessary to study the carbon export into the ocean interior and the associated climatic impacts (Gehlen et al., 2015).

In this study, we only focus on the primary production as estimated by the chlorophyll concentrations which are computed as the sum of

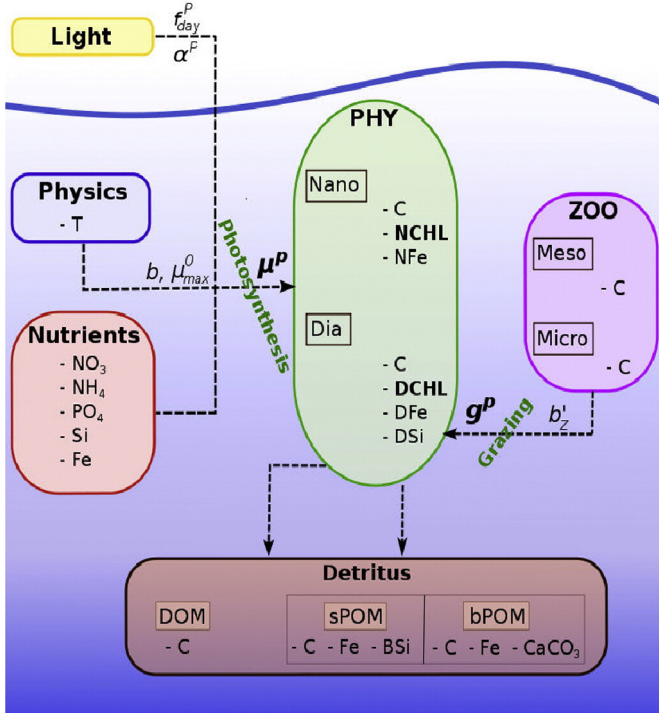


Fig. 1. Schematic representation of the main PISCES biogeochemical compartments. Except for the inorganic carbon (DIC), the alkalinity (Alk) and the oxygen (O_2), all PISCES biogeochemical variables are represented. Only interactions with the phytoplankton compartment (PHY) relevant to our study are shown. This sketch highlights in italic some parameters with an influence on the model estimation of the NCHL (nanophytoplankton chlorophyll) and the DCHL (diatoms chlorophyll) variables. These parameters mainly impact the phytoplankton compartment from Eqs. (1) and (4) of the bold parameters μ^P and g^P . These parameters will be considered as uncertain to set-up a stochastic PISCES version from Section 3.2.

the nanophytoplankton chlorophyll (NCHL) and the diatoms chlorophyll (DCHL) variables. Thus, only some of the interactions between the different biogeochemical compartments and the phytoplankton compartment (PHY) are represented in Fig. 1. These interactions are predominantly regulated by the growth rate μ^P (Eq. (1)) and the grazing rate g^P (Eq. (4)).

$$\mu^P = \mu_p f(L_{day}) g(Z_{mix}) \left[1 - \exp\left(\frac{-\alpha^P \theta^{chl,P} PAR^P}{L_{day} \mu_p L_{Lim}^P}\right) \right] L_{Lim}^P \quad (1)$$

$$\text{with } f(L_{day}) = f_{day}^P \left(\frac{L_{day}}{0.5 + L_{day}} \right) \quad (2)$$

$$\text{and } \mu_p = \mu_{max}^0 \exp(bT) \quad (3)$$

The P or Z (in Eq. (4)) exponents respectively refer to the phytoplankton type (nanophytoplankton or diatoms) and the zooplankton type (mesozooplankton and microzooplankton). Eq. (1) compute the main coefficient μ^P regulating the phytoplankton growth according to the photoadaptive model of (Geider et al., 1996). L_{Lim} is the nutrient limitation term for phytoplankton growth, the P -I slope α is a chlorophyll absorption efficiency coefficient and PAR represents the photosynthetic available radiation. L_{day} is the day length ($\in [0,1]$) used in Eq. (2) to calculate the $f(L_{day})$ dependency of growth rate to the length of the day factor as defined from Gilstad and Sakshaug (1990) and Berges et al. (2002). Note that this dependency only comes from studies on diatoms species but is nevertheless parameterized for the 2 phytoplankton species in PISCES. θ^{chl} is the chlorophyll to carbon ratio and $g(Z_{mix})$ is an attenuation function depending on the relative depths of the mixed layer depth z_{mix} and the euphotic layer (Lévy et al., 1998). The maximum

growth rate μ_p is defined from Eppley (1972) by Eq. (3) in which μ_{max}^0 is the maximum growth at 0 °C and b is a temperature T sensitivity term. Note that these 2 coefficients do not depend on the phytoplankton type. Equivalently, the g^P maximum grazing rate is formulated in Eq. (4) from the $g_{max}^{0,Z}$ maximum grazing rate at 0 °C and the b_z' temperature grazing sensitivity term. In the PISCES standard version used in this study, b_z' is kept identical for the 2 zooplankton.

$$g^P = g_{max}^{0,Z} \exp(b_z' T) F_p^Z \quad (4)$$

The additional factor F_p^Z denotes the preference of the zooplankton for each phytoplankton type. For more details about the PISCES-v2 biogeochemical formulations and the value of the parameters, the reader is invited to consult the latest documentation (Aumont et al., 2015).

In the coupled NATL025/PISCES configuration, the biogeochemical variables are considered as tracers with their evolution determined by the advective–diffusive Eq. (5):

$$\frac{\partial C}{\partial t} = \underbrace{-\nabla \cdot (uC)}_A - \underbrace{K_h \nabla_h^2 C}_{D_h} + \underbrace{\frac{\partial}{\partial z} \left(K_z \frac{\partial C}{\partial z} \right)}_{D_v} + SMS(C) \quad (5)$$

The model state vector C includes the 24 PISCES prognostic variables. The 3 terms A , D_h and D_v represent respectively the advection, the horizontal diffusion and the vertical diffusion computed by the physical model. SMS is the “Source Minus Sink” term of the PISCES model which contains the formulations explicated previously. To ensure a consistent biological initial state, PISCES is initialized in January 2002 from outputs of a global 1/4° PISCES simulation performed by the operational MERCATOR Ocean system. In addition a 3 years biogeochemical spin-up is performed between January 2002 and December 2004. The coupling between the PISCES model and the physics (explicated by Eq. (5)) is performed on-line. The biogeochemical variables are calculated at each time step, which means that approximations coming from temporal averages are avoided. As shown in Fig. 2, the physical NATL025 configuration does not drastically shift the annual mean biogeochemical variable states used for initialization even after 3 years. Small differences between the 2 representations are expected, as the 2 coupled configurations are not exactly similar. For instance, the vertical resolution is 50 levels in the Mercator simulation and the coupling procedure is performed off-line. Finally, it is essential to notice that together with the works of Resplandy et al. (2012) on the Arabian Sea and Dufour et al. (2013) on the Southern Ocean, this simulation is one of the first on-line coupling of PISCES with NEMO at an eddy permitting resolution.

2.2. A deterministic model simulation

Before transforming PISCES into a probabilistic model, a deterministic simulation *detlni* is performed as a first guess using the standard parameter values defined in Aumont et al. (2015). The experiment period is 6 years, extending from January 2005 to December 2010. In the 2 next subsections, the objective is to emphasize some general characteristics of interest about the horizontal and vertical chlorophyll deterministic representations.

2.2.1. Surface chlorophyll representation

We analyze the surface chlorophyll spatial repartition by comparing the deterministic solution with the SeaWiFS satellite data. These SeaWiFS ocean color data have been obtained from the NASA’s Goddard Space Flight Center ([www.http://oceancolor.gsfc.nasa.gov](http://oceancolor.gsfc.nasa.gov)). Near-surface chlorophyll-a concentrations are derived from the OC-4 operational algorithm (O’Reilly et al., 1998). A well admitted error estimation for this sensor is around 30% (Gregg and Casey, 2004). To assess the model consistency right after the biogeochemical spin-up (see previous section), surface chlorophyll maps of the PISCES deterministic simulation *detlni* are compared with SeaWiFS in Fig. 2 for the 31st of

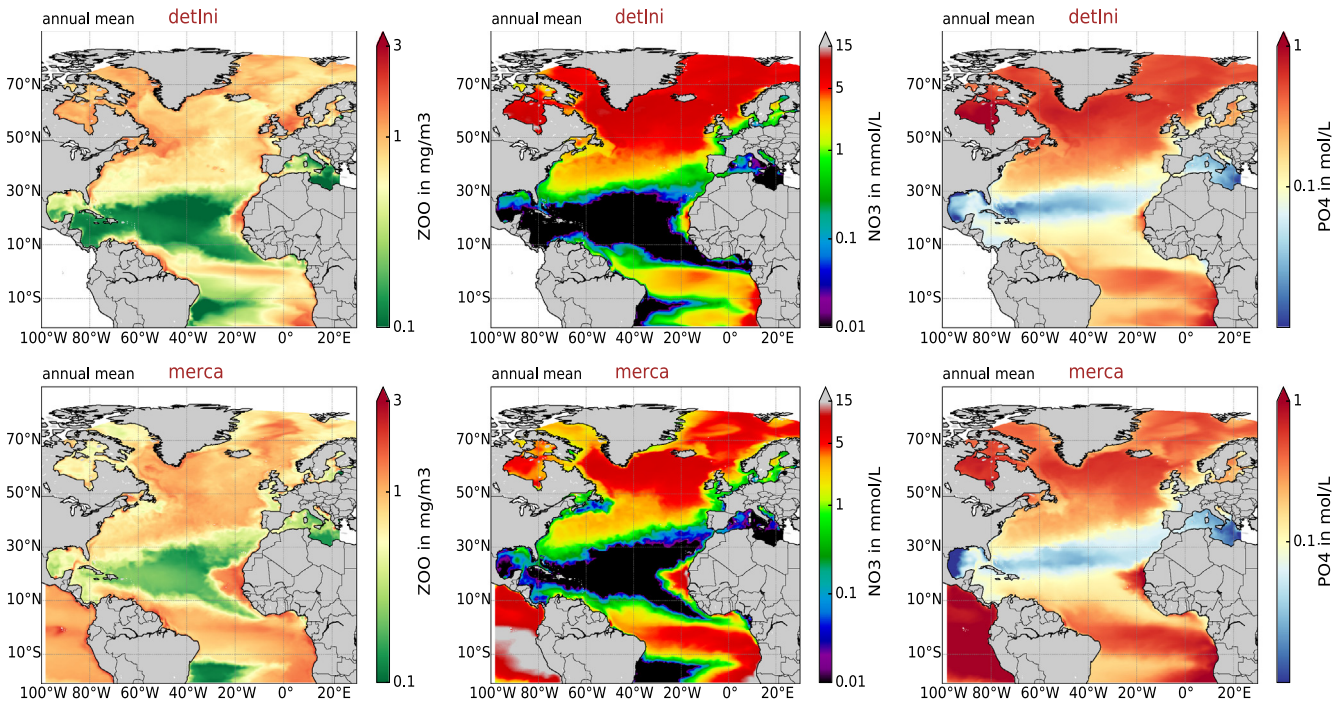


Fig. 2. Surface annual mean concentrations of the deterministic simulation *detlni* analyzed in next sections (top) and the Mercator simulation used to initialize the model in January 2002 (bottom). The panels exhibit the microzooplankton concentrations (left), the nitrate concentrations (middle) and the phosphate concentrations (right). Results are shown for the year 2005.

March 2005 (left panels), the 15th of May 2005 (middle panels) and the 27th of October 2005 (right panels).

In spite of large scale patterns of surface chlorophyll quite well reproduced by the model, one can detect misfits between the satellite data and the deterministic simulation. For instance during the spring bloom period, the surface chlorophyll concentration is always overestimated around the Gulf stream region (mainly between 30°N and 50°N). Nevertheless, PISCES exhibits a relatively coherent northward and shoreline surface chlorophyll propagation in March and October, when the biological activity is less intense. At middle and high latitudes, although the model simulates a realistic persistence during the winter, the model does not represent the “patchy” horizontal chlorophyll structures observed in satellite data. The oligotrophic conditions of the North Atlantic subtropical gyre are too expanded. Also, its transitions with the northern more productive regimes are always too sharp and located southward. More generally, north to south gradients are always overestimated over the domain. In coastal zones, where organic matter concentrations and the supply of iron coming from the sediments and the aeolian dusts are very important, the simulation depicts relevant surface chlorophyll representations. Intense biological productions of major upwellings (e.g., Mauritania) are by consequence quite well simulated.

Finally, as the PISCES model formulations are highly complex, it is quite difficult (and not the purpose here) to identify precisely which process is responsible for each deviation with the observations. Although we have highlighted that the magnitude of chlorophyll concentrations is often imperfectly estimated, PISCES complexity level is able to reproduce the main large scale surface chlorophyll patterns of the North Atlantic basin observed in the SeaWiFS data.

2.2.2. Vertical chlorophyll representation

One of the main difficulty with ocean color data is that the satellite observations are confined to the surface. Extrapolation to the vertical then remains a challenge on which the deterministic model vertical structure is expected to supply a relevant space and time information. In order to compare PISCES vertical structures with global observations,

we use in this section a recent chlorophyll climatology derived from in situ fluorescence data (Sauzède et al., 2014). This climatology is constructed according to the geographical partition of the Longhurst provinces (Longhurst, 1995) presented in Fig. 4.

Fig. 5 compares spatial averages of the in-situ climatological chlorophyll vertical profiles (blue curve) with the *detlni* PISCES simulation (red) in the geographical provinces 4, 6 and 7. To make the comparison significant, the monthly profiles of the simulation *detlni* are calculated from the mean of the 6 years experiment (2005–2010). These chlorophyll vertical profiles can therefore be considered as a “model climatology”. In province 7, the PISCES vertical structure fits quite well with the climatology (except in March but the climatology contains very few data). Due to the Atlantic gyre oligotrophic conditions, surface chlorophyll concentrations are predominantly weak and the model is able to simulate the existence of subsurface deep chlorophyll maxima (DCM). During the spring bloom period (May), it is well reproduced although it is slightly too deep in October. At higher latitudes (province 4), PISCES vertical structures have a shape relatively comparable to the climatology for the months of March and May even if the presence of weak DCM is not represented by the model. In October, far too weak chlorophyll concentrations are simulated in the whole surface layer. In province 6 (around the Gulf Stream region), the model and the climatology are very different during the spring bloom. Within the euphotic layer, PISCES simulates a constant decrease of chlorophyll while the climatology exhibits a strong DCM. Note that the model surface chlorophyll concentrations are strongly overestimated in this region (Fig. 3). A surface chlorophyll agreement with observations is therefore necessary to expect relevant vertical structures. In October, the model surface chlorophyll is in much better agreement with the observations. Even though it is slightly too shallow, the DCM is well represented. The chlorophyll global vertical structure is close to the climatology. In March, the climatological large DCM is not represented by the model.

To conclude, despite important imperfections, PISCES is able to reproduce some of the main large scale biogeochemical structures. This capacity was also highlighted in several earlier studies addressing

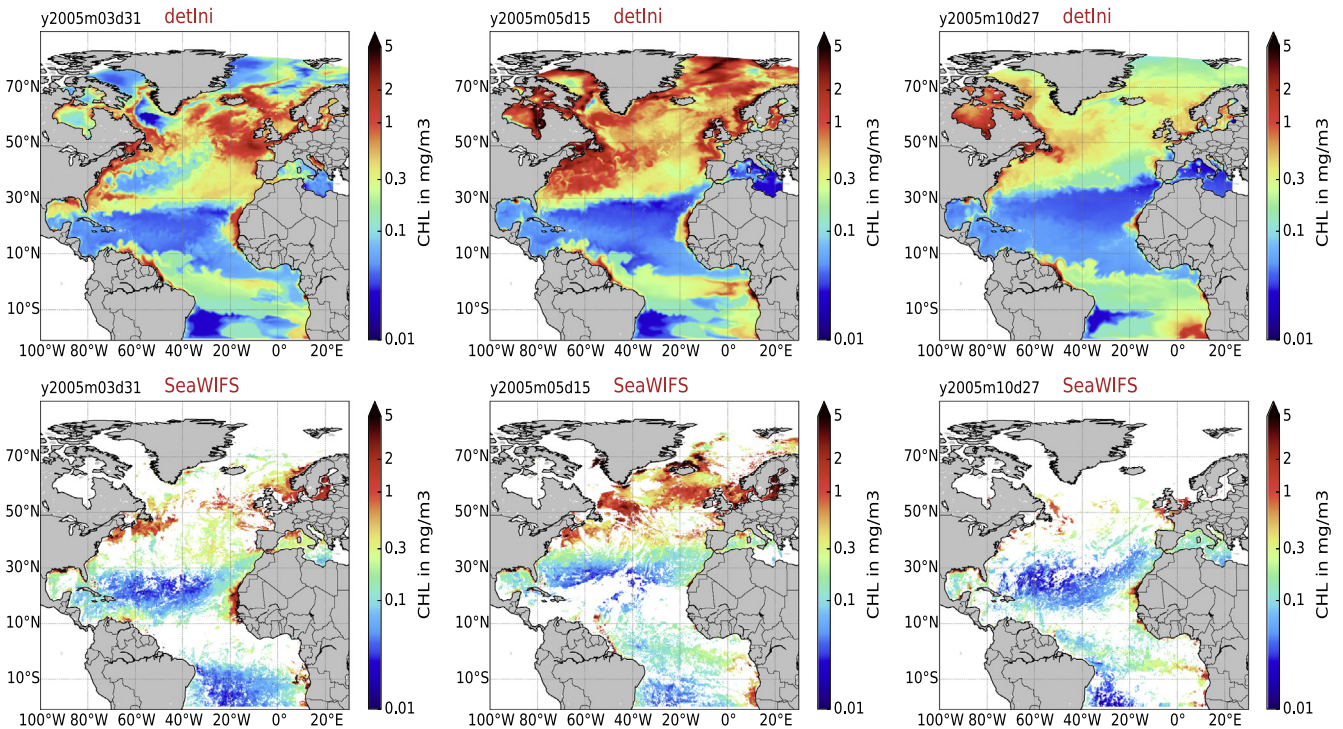


Fig. 3. Surface chlorophyll maps of the PISCES deterministic simulation *detIni* (top) and the SeaWiFS satellite data (bottom) for the 31/03/2005 (left panels), the 15/05/2005 (middle panels) and the 27/10/2005 (right panels). Concentrations are calculated from 5 days averages.

biogeochemical properties for diverse temporal and spatial scales (Aumont and Bopp, 2006; Gehlen et al., 2006; Rodgers et al., 2008; Tagliabue et al., 2009 to cite only a few). In addition, it is consistent with previous studies (for e.g., Uitz et al., 2006) indicating that the representation of surface chlorophyll largely determine the coherence of vertical structures. Finally, the crucial point for our study is that a better agreement with ocean color surface data should thereafter allows to depict better vertical structures within the euphotic layer. Hence, improving surface chlorophyll patterns in agreement with SeaWiFS data is exactly the objective of the next section in which we present a generic stochastic method to simulate some biogeochemical uncertainties.

3. Explicit simulation of biogeochemical uncertainties

In the previous section, it has been suggested that PISCES is able to include many relevant pieces of dynamical information that can be observed by the SeaWiFS ocean color satellite data. However, we also insisted on the fact that the chlorophyll distributions are still imperfect and the model is therefore very uncertain. Biogeochemical uncertainties being ubiquitous, we cannot expect to design a specific parameterization for every single source of uncertainty. We therefore propose to focus on some of the uncertainties with a direct impact on the primary production estimate of the model. As a number of studies focus on the impact of the oceanic component (for instance Anderson and Robinson, 2001; Doney et al., 2004), our goal is to investigate how far some biogeochemical uncertainties alone can explain the deviations with satellite observations. In addition, the originality of the approach comes from the simulation of uncertainties inherent to the biogeochemical model formulations.

3.1. Stochastic parameterizations

The purpose of this section is to propose the generic approach based on direct stochastic parameterizations of some biogeochemical model uncertainties. We do not intend here to encompass every source of biogeochemical uncertainties or design the best refined stochastic parameterizations. The objective is mainly to highlight the potential of such a probabilistic approach by a focus on 2 main classes of biogeochemical uncertainties. The first class is related to approximations on the biogeochemical parameters (see Section 3.2). This can be analyzed as mainly resulting from the reduced description of the system by a limited number of state variables. The second class of uncertainties is related to the limitation of the spatial scales resolved by the model (Section 3.3). Because of the non-linearity of the bio-model formulation, unresolved scales can indeed produce an averaged effect on the ecosystem behavior.

To parameterize these 2 classes of uncertainties, the approach adopted here is to modify the model equations by including stochastic

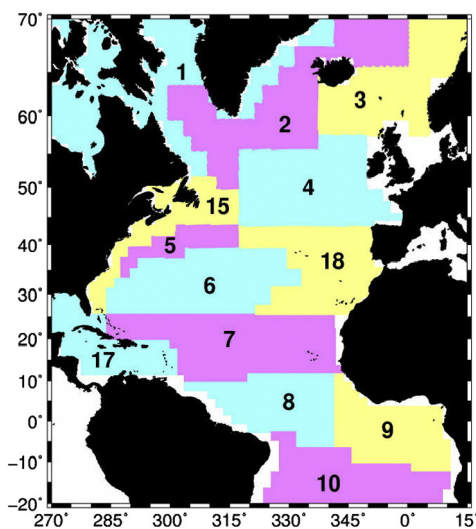


Fig. 4. Longhurst (1995) biogeochemical provinces as presented in Doron et al. (2011). The coastal areas are excluded.

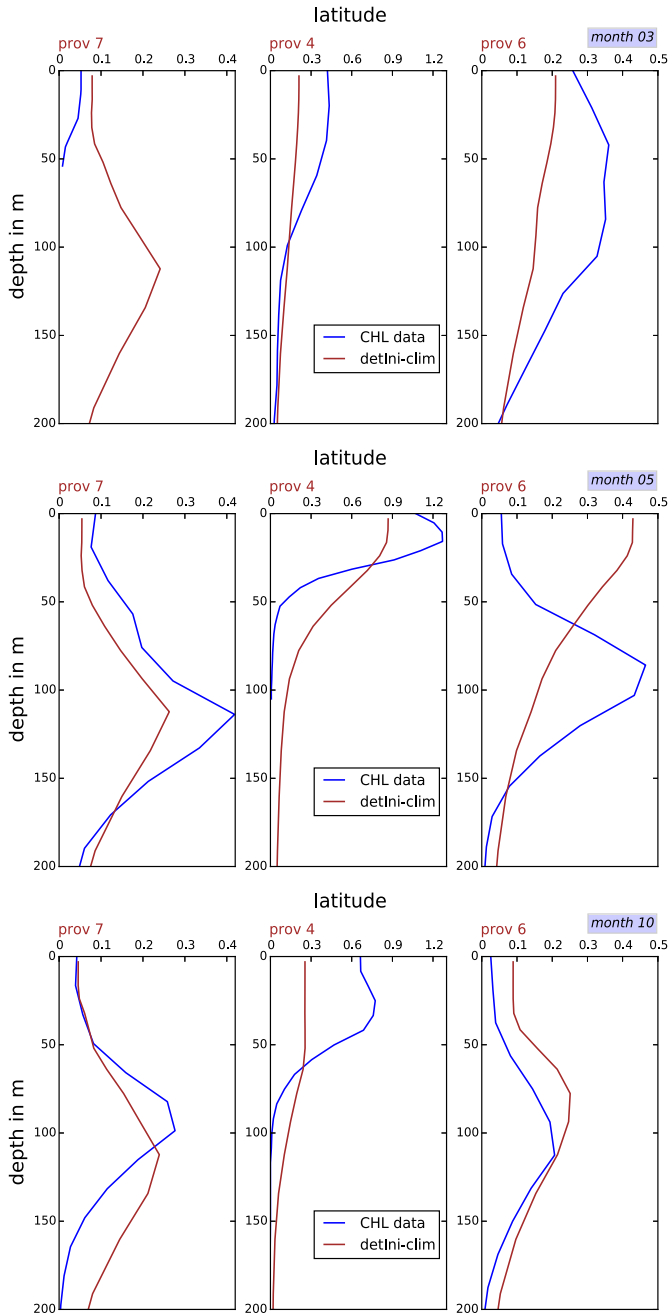


Fig. 5. Chlorophyll vertical profiles (in mg/m^3) spatially averaged in Longhurst provinces 4, 6 and 7 (Longhurst, 1995). The blue curve is the climatology (Sauzède et al., 2014) and the brown curve is a *detIni* PISCES simulation climatology calculated from an average between 2005 and 2010 (6 years). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

processes to explicitly simulate uncertainties in the model formulations. These stochastic processes must be tuned to fit reasonable prior assumptions about model uncertainties in terms of magnitude and space/time correlation structure. For all perturbations, the starting point is a first-order autoregressive (AR(1)) process. For each stochastic perturbation (the 7 biological parameters and the unresolved scales), at each grid point, and at each time step k , a random noise $\xi(t)$ is drawn from an AR(1) process:

$$\xi(t_k) = \varphi \xi(t_{k-1}) + \sqrt{1 - \varphi^2} w \quad (6)$$

where w is a Gaussian white noise and φ is related to a decorrelation time scale τ : $\varphi = e^{-1/\tau}$. To introduce spatial correlations in the stochastic perturbations, these maps of random noises are convoluted with a two-dimensional 11×11 smoothing kernel. The statistical parameters of the stochastic processes are recapitulated in Table 1 (Section 4). This implementation generalizes the work of Brankart (2013) to simulate the effect of unresolved scales in the computation of the large scale horizontal density gradients. These maps of smoothed, time-correlated random noises are then used to perturb biogeochemical parameters and to simulate some effects of unresolved scales as described below. In the next 2 sections, we explain how these stochastic processes are used to simulate some uncertainties inherent to the biogeochemical model.

3.2. Stochastic parameterization of uncertain biogeochemical parameters

As already mentioned, one of the main difficulties to simulate marine ecosystem behaviors comes from the empirical formulation of biogeochemical processes in the model. The mathematical formulations are mainly based on laboratory experiments and only a few parameters are supposed to figure out some complex phenomena as for instance the diversity of species or the biological adaptation. In addition, the uncertainties of the biological parameters also result from the spatial and the temporal variation of the biological behaviors like the seasonality or the oceanic diversity (Doron et al., 2011). For the perturbations of biological parameters, the AR(1) processes are set up with a standard deviation $\sigma_{up} = 0.3$ and a decorrelation time scale $\tau = 1$ month (Table 1). After spatial filtering, the Gaussian noises are transformed into Lognormal noises to guarantee positivity. The stochastic perturbations are introduced by the multiplication of the parameters with these Lognormal noises built from the exponential of the previously spatially filtered AR(1) Gaussian variables. A biological parameter p is then transformed into $p' = \exp(\xi(t))p$. Then, the SMS term of Eq. (5) is computed using p' instead of p . To preserve vertical consistency, all multiplicative perturbations are set identical for the whole water column. The perturbations on the different parameters are uncorrelated with each other. Note that the mean of a $\text{LogN}(\mu = 0, \sigma_{up} = 0.3)$ is $\exp(0.045) \approx 1.04$. By consequence, the mean of a perturbed parameter p' is really close to the original parameter p . This will be discussed in details in Section 4. To be consistent with Doron et al. (2013) and to produce a quite simple configuration able to take into account some of the main first order biological uncertainties, 7 influent biogeochemical parameters are perturbed:

- the growth rate at 0°C μ_{max}^0 and the temperature sensitivity of growth b of the 2 phytoplankton classes.
- the temperature sensitivity of grazing b_z' of the 2 phytoplankton classes.
- the initial P-I slope (α^p) for the 2 phytoplankton classes.
- the f_{day}^p dependency of growth rate to the length of the day factor for the 2 phytoplankton classes.

This makes a total of 7 independent stochastic processes introduced in the model equations. Note that this formulation is similar to the SPPT scheme implemented by Buizza et al. (1999) to simulate uncertainties in the ECMWF ensemble prediction system.

3.3. Stochastic parameterization of unresolved scales

Stochastic processes have also been shown to provide a relevant solution to represent the effect of unresolved scales in ocean circulation model, as a result of the non linearity of the advection term (Frederiksen et al., 2012) or of the seawater equation of state (Brankart, 2013). Biogeochemical model equations being particularly nonlinear, unresolved fluctuations do not cancel by averaging the model equations [$\overline{SMS(C_i)} \neq SMS(\overline{C_i})$] and can thus represent a major source of uncertainties

(Lévy and Martin, 2013). As the model equations are solved at each grid point for a mean value of C_i , the effects of unresolved scales will have an impact on the large scale biogeochemical representation and therefore cannot be neglected. To simulate this effect with stochastic parameterizations, we create a perturbation δC which simulates the unresolved fluctuations of concentration C within every model grid box. The perturbation is then used in the equation of evolution of C :

$$\frac{\partial C}{\partial t} \Big|_{\text{bio}} = \frac{1}{2} [SMS(C + \delta C, u, p, t) + SMS(C - \delta C, u, p, t)] \quad (7)$$

The perturbations δC are dimensioned from the comparison of SeaWiFS high resolution ($1/12^\circ$) maps and their upscaled, filtered counterparts at a $1/4^\circ$ resolution. At first approximation, it shows that the differences in chlorophyll concentration due to unresolved scales are approximately proportional to the chlorophyll concentration itself, in a proportion of 20–30%. Therefore, perturbations are constructed from the expression $\delta C = \xi(t)C$, where the parameters of the AR(1) process that computes $\xi(t)$ are adjusted so that the standard deviation is 0.2 (20%). Note that as for the parameterization of uncertain parameters, the same $\xi(t)$ are used along the z axis. In this procedure, $\xi(t) \in \mathcal{N}(\mu = 0, \sigma_{up} = 0.2)$. Adding δC to the tracer C is therefore equivalent to the multiplication of C by a normal perturbation $\xi'(t) \in \mathcal{N}(\mu = 1, \sigma_{up} = 0.2)$. With this construction, the parameterization has no effect on the solution when the SMS equation is linear. It therefore accounts for the non linearity degree of the SMS Equations.

4. Joint tuning of deterministic and stochastic parameters

As already mentioned, a key objective of this study is to perform a probabilistic simulation with a sufficient level of uncertainty to be compared with ocean color images. The consistency of the simulation is therefore examined from comparisons with ocean color data (SeaWiFS), considering that each member of the simulation need to be a possible representation of the observed large scale chlorophyll distribution. In this section, we explain that the introduction of stochastic parameterizations in the model produces a mean effect due to the

Table 1

List of the parameter values used in the different simulations performed to set up a consistent PISCES probabilistic model version. All PISCES parameters presented here are perturbed according to the same statistical parameters corresponding to the uncertain parameter perturbations. The PISCES model Formulations involving these parameters are described in Section 3.2.

Tuning of parameters				
PISCES parameters	Simulations			
	detIni	stoIni	detUpd	stoUpd
Nanophytoplankton initial P–I slope: α^N	2	2	4	4
Diatoms initial P–I slope: α^D	2	2	4	4
Phytoplankton growth rate at 0 °C: μ_{max}^0	0.6	0.6	0.5	0.5
Phytoplankton temperature sensitivity of growth: b	0.064	0.064	0.045	0.045
Zooplankton temperature sensitivity of grazing: b_z'	0.076	0.076	0.084	0.084
Nanophytoplankton growth dependency to the day length: f_{day}^N	1.5	1.5	1.5	1.5
Diatoms' growth dependency to the day length: f_{day}^D	1.5	1.5	1.5	1.5
Statistical parameters				
	Simulations			
	detIni	stoIni	detUpd	stoUpd
Decorrelation time step (in month): τ	0	1	0	1
Gaussian distribution mean μ	0	0	0	0
Standard deviation for unresolved scale perturbations: σ_{us}	0	0.2	0	0.2
Standard deviation for uncertain parameter perturbations: σ_{up}	0	0.3	0	0.3
Laplacian spatial correlation filter (in grid points):	0	5	0	5

combined effects of (i) small scale variations of chlorophyll and (ii) the non linearity of the SMS function. To deal with this systematic bias, we propose to redefine the initial set of biological parameters used in Section 2 to perform the first guess deterministic simulation *detIni*.

4.1. Impacts of uncertainties using the original deterministic parameters

Using the stochastic perturbations presented in the previous section, we can setup a probabilistic version of the PISCES model. This probabilistic model includes a stochastic component (presented in Section 3) as a complement to the deterministic model dynamics (presented in Section 2). This probabilistic model is therefore characterized by an extended set of parameters including the stochastic parameters in addition to the deterministic parameters. As a first attempt, a probabilistic model simulation *stoIni* is performed using the same deterministic parameters as the *detIni* simulation described in Section 2. The stochastic parameters characterize the random perturbations $\xi(t) \in \text{LogN}(\mu = 0, \sigma_{up} = 0.3)$ used to simulate parameter uncertainties and the random perturbations $\xi(t) \in \mathcal{N}(\mu = 0, \sigma_{us} = 0.2)$ used to simulate the effect of unresolved scales (see previous section). In addition, a $\tau = 1$ month decorrelation time scale and a 5 grid points spatial filter are also applied for all perturbations. These parameters values are summarized in Table 1. The impacts of the perturbations on the model solution are investigated by performing a 3 member ensemble simulation initialized on the 01/01/2005 from the *detIni* deterministic model simulation outputs. In Fig. 6, we present these 3 members together with the deterministic simulation *detIni* during the spring bloom.

The most relevant thing highlighted in Fig. 6 is that these stochastic parameterizations produce a strong effect on the model representation. Moreover, as compared to the deterministic simulation, the 3 members of the stochastic simulation clearly show much stronger surface chlorophyll concentrations nearly everywhere. Consequently, these simulations systematically overestimate the surface chlorophyll concentrations compared to the SeaWiFS map presented in Fig. 3. This effect is observed in any member as soon as these uncertainties are explicitly simulated. Many experiments investigating the impacts of various stochastic parameterizations have always shown this tendency. Note that this surface chlorophyll increase is not the result of the Log-normal shape of the distribution imposed to simulate biogeochemical parameter uncertainties (to avoid negative value), since the mean of the perturbed parameter is really close to the original parameter value (see Section 3.2). Actually, the surface chlorophyll increase observed in Fig. 6 mainly comes from the unresolved scale parameterization, which has been tuned in our experiment using the root mean square of the small scales in high resolution SeaWiFS images (see Section 3.3). With non linear equations, these perfectly symmetric Gaussian perturbations lead to a mean effect (bias) depending on the concavity/convexity shape of the equations (non linearity degree of the SMS equations).

4.2. Re-tuning of the PISCES deterministic parameters

From the previous considerations, the initial deterministic simulation *detIni* can be seen as a single simulation of the probabilistic model with all stochastic parameters set to 0 (see Table 1). In this simulation, the initial deterministic parameters were calibrated to obtain in average, a best fit with observations. This calibration explains the consistency between the yellow curve and the blue curve in Fig. 7. To reach this level of consistency, model uncertainties (which should have lead to strong misfits between the deterministic simulation and the observations) were indirectly compensated by a specific tuning of the deterministic parameters (e.g., an excessively high photosynthesis efficiency to simulate unresolved biological light adaptation). The initial parameter values should therefore not be considered as written in stone but as an optimal choice depending on the level of simulated uncertainties. When

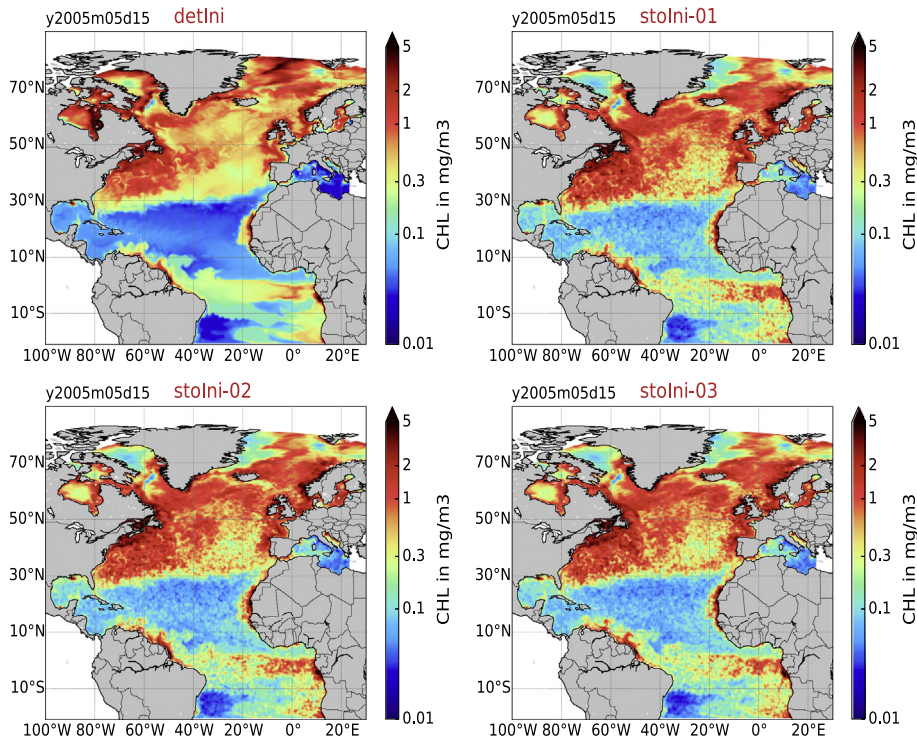


Fig. 6. Surface chlorophyll maps of the 15/05/2005 (5 days average) for the PISCES *detIni* deterministic simulation and the 3 members of the stochastic *stolNi* ensemble simulation.

stochastic parameters are thereafter different from 0 (simulation *stolNi*), the simulated level of uncertainty is increased which produces a surface chlorophyll increase. To counterbalance this mean effect, we propose to re-tune the value of some of the deterministic parameters as presented in Table 1. As a first step mainly due to model computation complexity, these parameters are re-tuned empirically (based on sensitivity experiments). Thus, it must not be considered as optimal and should be improved in the future. The parameter values are fixed with the specifications to 1) provide ensemble members coherent with SeaWiFS data, 2) keep consistency with the initial deterministic values and 3) keep a sufficient level of uncertainty to ensure dispersion.

To investigate the impacts of this updated deterministic set of parameters, we perform a 1-year deterministic simulation *detUpd* (with zero standard deviation perturbations) and a 3 member-ensemble simulation *stoUpd* with the deterministic and stochastic updated parameter values (see Table 1). Surface chlorophyll concentrations are shown in Fig. 8 for the 15/05/2005. As we can observe in Figs. 7 and 8, the main effect of this re-parameterization is a strong chlorophyll decrease in all Atlantic zones. If this effect seems to well improve the surface chlorophyll in the Gulf Stream area (simulation *detUpd*), the solution is

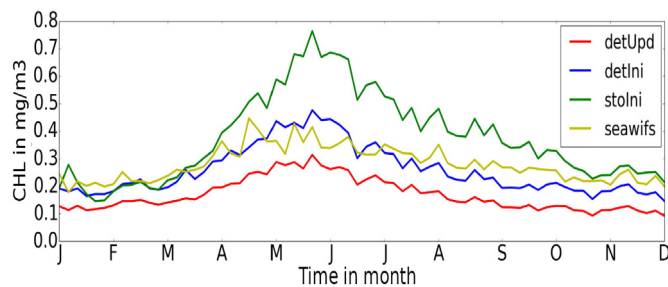


Fig. 7. Time series of surface chlorophyll spatially averaged over the whole domain for the year 2005. The PISCES deterministic simulation *detIni* is represented in blue, the SeaWiFS satellite data in yellow, the perturbed simulation *stolNi* in green and the deterministic simulation *detUpd* in red. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

degraded nearly everywhere else and most features of the PISCES deterministic simulation *detIni* (see Section 2) are worsen. Nevertheless, each realization of the ensemble *stoUpd* is qualitatively similar to the SeaWiFS image presented in Fig. 3 in terms of surface chlorophyll spatial distribution and intensity. Chlorophyll concentrations are quite well represented in key areas like upwelling zones, high latitudes, the south equatorial current, and coastal regions. It also relevant to record strong small scale effects, very different between one member and another. These small scales are directly associated with the ensemble dispersion, required to compare the ensemble with observations (see Section 5.1). Everywhere and for each member, surface chlorophyll patterns are also more patchy than the smooth deterministic simulation *detIni* which is consistent with earlier studies as Maranon et al. (2000) or Gist et al. (2008).

To conclude, one of the main feature to point out is that the updated set of parameters, without stochastic processes, cannot be used to simulate a biological state comparable to the SeaWiFS observations. Because the stochastic parameterizations are inherent to the model formulation, this updated configuration can only be appropriate for a simulation which includes the stochastic perturbations. We therefore cannot separate stochastic parameterizations and model configuration but they must be considered together to perform a probabilistic simulation. Once this re-tuning of the whole set of parameters (deterministic and stochastic) is accomplished, we can expect to perform a reliable probabilistic simulation (with many more members), which we can analyze in more details and compared to observations using an appropriate statistical approach. This is the purpose of the next sections.

5. A 60-member probabilistic simulation

To explore the biogeochemical model response to some of the major biological uncertainties, we present here the results of a 60-member ensemble simulation. This probabilistic simulation is performed from January 2005 to December 2005 with the parameters of the simulation *stoUpd* described in the previous section. In this section, the objectives are to describe the effects of the stochastic parameterizations on the

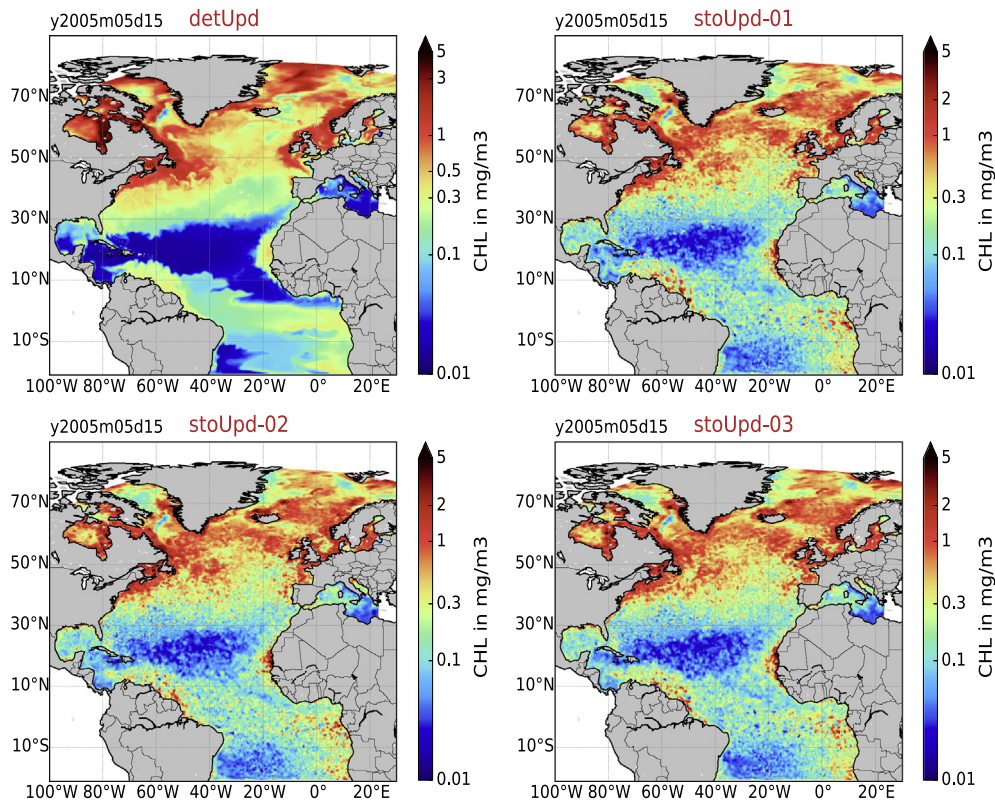


Fig. 8. Surface chlorophyll maps of the 15/05/2005 for the PISCES (*detUpd*) deterministic simulation and the 3 members of the stochastic *stoUpd* ensemble simulation.

chlorophyll representation. Before we analyze these effects, it is worth noting that these results depend on the stochastic parameterizations presented before. Other assumptions about the noise structure or the consideration of other biological parameters will presumably lead to different results. In the first part, we show some general characteristic about the chlorophyll distribution in terms of ensemble dispersion. Then, we show that the probabilistic approach allows to objectively compare the model with SeaWiFS observations. In particular, we prove that the ensemble spread is able to capture ocean color data with a consistent statistical representation of their probability of occurrence (reliability). Note that a very significant CPU time (about 50 000 h) has been necessary to set-up the stochastic configuration (previous section) and perform this ensemble simulation. The choice of 60 members to define the probability distribution is therefore mainly driven by computer costs.

5.1. Biological response to the stochastic parameterizations

5.1.1. Surface chlorophyll response

In the previous section, we have shown that each member of the simulation *stoUpd* is able to represent a biological state in agreement with the SeaWiFS observations. To extend the analysis of the impacts of the stochastic perturbations, Fig. 9 shows surface chlorophyll maps of the ensemble mean and the relative standard deviation (absolute standard deviation normalized by the ensemble mean).

The first relevant feature is that the ensemble mean exhibits several improvements compared to the deterministic simulation (see Fig. 3). For instance, improvements in the Gulf Stream region coming from the re-tuning of parameters (Fig. 8) are conserved. PISCES chlorophyll overestimations in the south of the Gulf Stream pathway had indeed been identified as a main problem of the pre-operational MERCATOR PISCES simulation used to initialize the simulations (see Section 2.1). Note that this improvement could be observed in Fig. 8, as the mean effect

exists in each member of the ensemble simulation. The direct consequence is a smoothing of the too strong north to south gradient simulated around 30°N by the deterministic simulation. This effect also exists in the equatorial region: the surface chlorophyll mean is decreased and so the gradients are lowered which tends to smooth the large scale surface chlorophyll distributions. In addition, the relevant features observed for March and October in the deterministic simulation are conserved by the ensemble mean surface chlorophyll representation. In fact, Fig. 9 highlights improvements coming for the stochastic parameterizations as a consequence of the mean effect of the stochastic parameterizations.

As we could expect from previous results, the ensemble dispersion is strongly dependent to the geographical area. Except for the Mediterranean sea and several locations in high latitudes, the ensemble still exhibit a quite strong dispersion over the domain. The relative standard deviation maps in Fig. 9 show maximal dispersion in the Gulf Stream region, the equatorial current and around the coastal zones, where small scale effects are predominant (Fig. 8). Also, it is relevant to observe that in spite of oligotrophic conditions (i.e., low chlorophyll concentrations) the dispersion is nevertheless significant in the subtropical gyre. On the contrary, a relatively small relative standard deviation is observed in high latitudes where chlorophyll concentrations are quite high. Besides, the small dispersion diagnosed with the relative standard deviation is partly compensated by strong chlorophyll concentrations and a significant interquartile range (see Section 5.2). As we could expect, the dispersion is smaller in March and October, when the biological activity is less intense.

5.1.2. Chlorophyll vertical distribution

Regarding the chlorophyll vertical distributions, the validation of the ensemble simulation is more delicate since there are no observation network comparable to the satellite data. Comparisons with the climatology constructed from space and time averages were possible in the deterministic case (Section 2.2.2) because we needed to counterbalance

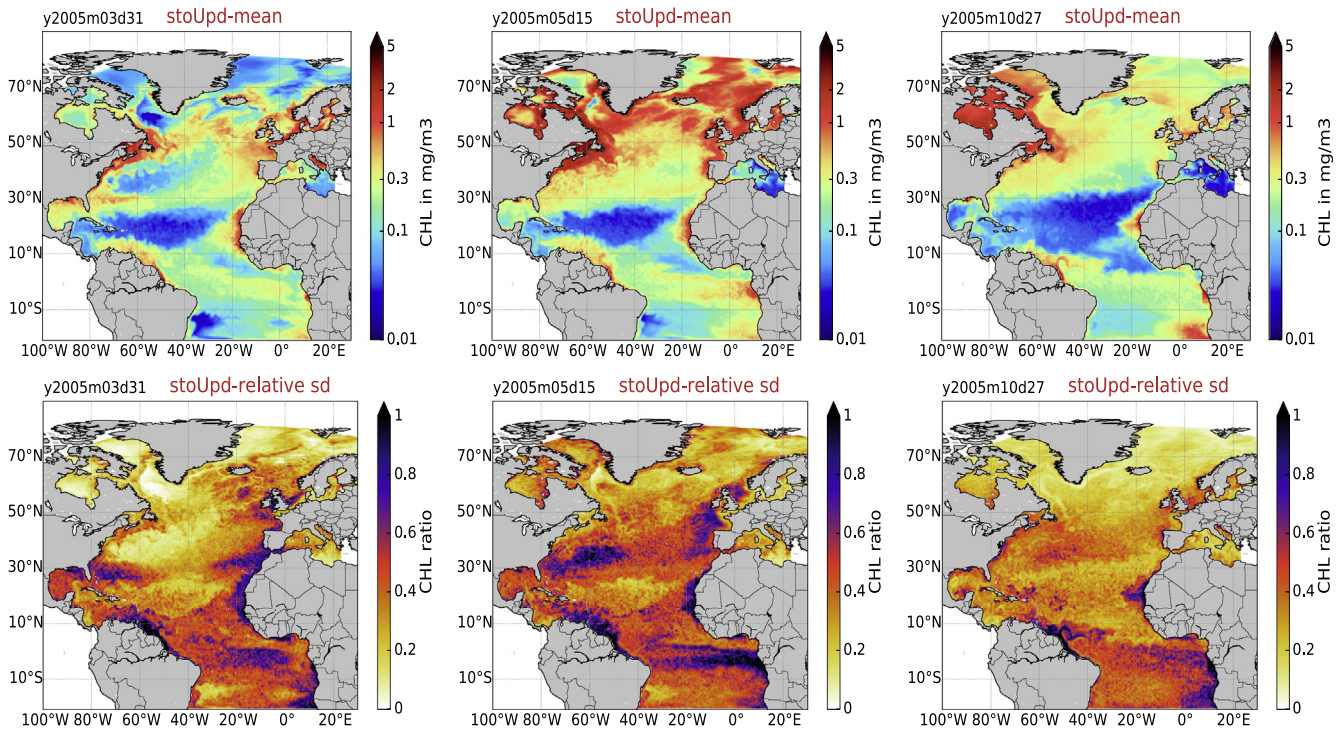


Fig. 9. Surface chlorophyll ensemble mean (top) and relative standard deviation (bottom) for the 31/03/2005 (left panels), the 15/05/2005 (middle panels) and the 27/10/2005 (right panels). Relative standard deviations are calculated at each grid point from the ratio between the absolute standard deviation and the ensemble mean.

the spatial heterogeneity in order to analyze the large scale vertical distributions. Since we consider now a probabilistic simulation, the expected objective is exactly to simulate this strong biological heterogeneity, considered as a representative sample of the model uncertainty. In Section 2.2.2, we highlighted that the *detIni* first guess deterministic simulation performs quite well on the vertical. We also showed that the vertical structure consistency depends on the surface chlorophyll representation. Thus, the objectives are here to verify that 1) the stochastic perturbations do not deteriorate the deterministic

representation (considered as the best space and time information) and 2) that a vertical dispersion is generated.

In Fig. 10, we compare the chlorophyll vertical structures of 4 members with the ensemble mean and the deterministic simulation along a 20°W longitudinal section during the spring bloom (15/05/2005). This points out that the large scale vertical structure of the deterministic simulation is conserved by the ensemble simulation. The mean structure correctly represent the main vertical patterns such as the geographical chlorophyll range of magnitude, the depth of propagation and maximum

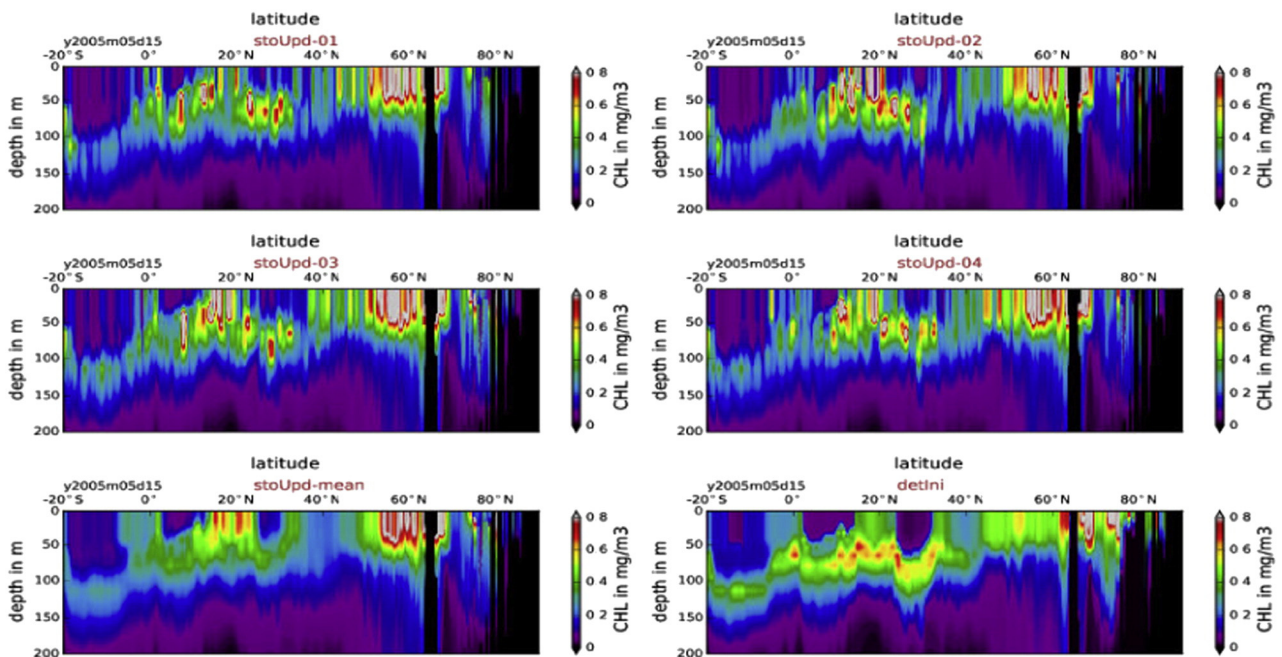


Fig. 10. Chlorophyll vertical sections at 20°W for the 15/05/2005. We represent 4 members of the 60-member ensemble simulation (stoUpd-01 to stoUpd-04), the ensemble mean (stoUpd-mean) and the PISCES deterministic simulation *detIni*. Note that the members stoUpd-01, stoUpd-02 and stoUpd-03 were presented in Fig. 8 for the surface chlorophyll.

chlorophyll depth (DCM) localizations. As we expected by setting the stochastic processes identical along the vertical axis (see section 3), each member is also comparable with each other's. The main vertical patterns (explicited above) are also simulated by each member, which highlights the coherence of the ensemble simulation. In addition, a dispersion is also generated within the euphotic layer. It means that the stochastic parameterizations reach to simulate different vertical distributions. Each realization is characterized by discontinuities induced by small scale effects, probably more realistic than the smooth solution of the deterministic simulation. Finally, at a given grid point, the stochastic model is able to display a variety of vertical behaviors, which is a key feature in order to simulate vertical profiles in agreement with observations. In addition, it would allow to get relevant correlations between surface chlorophyll and vertical distributions, which is fundamental for data assimilation.

To conclude, we showed that the stochastic parameterizations presented in Section 3 provide a means to perform a relevant probabilistic simulation. Compared to the deterministic simulation *detIni*, several improvements on the surface chlorophyll patterns have been highlighted. In addition, the vertical deterministic distributions are not deteriorated by the stochastic parameterizations. Each member is able to simulate different vertical profiles, which can be seen as the signature of the existence of different phytoplankton species. In the next section, we show that a probabilistic simulation contains a very large quantity of information that must be considered statistically, using specific diagnostics.

5.2. Probabilistic response to biological uncertainties: comparison with SeaWiFS ocean color data

In Section 4.2, we compared 3 members of the stochastic simulation with both the deterministic simulation and the SeaWiFS satellite data to check that each realization of the ensemble provide a possible representation of the biological state. Nevertheless, if this overview was sufficient to validate the relevance of the stochastic parameterizations, these comparisons do not effectively take into account the statistical information given by the ensemble. For this purpose, we investigate in the

following sections how the observations are included (or not included) within the ensemble.

5.2.1. Ensemble quantiles

The statistical surface chlorophyll response to uncertainties for the 15/05/2005 is presented in Fig. 11. In this figure, surface chlorophyll concentrations are examined from quantile maps which consider simultaneously the information coming from all ensemble members. Thus, these maps do not show model states but are composites representing at each grid point the minimum, the 25% percentile, the median, the 75% percentile and the maximum of the ensemble. To make the comparison easier, the model data have been interpolated on the grid of SeaWiFS. These observations are also presented in Fig. 11.

Here, the dispersion is diagnosed by the interquartile range, which does not take into account the magnitude range of the chlorophyll concentrations (unlike the relative standard deviation maps presented in Fig. 9). It shows that a significant dispersion is also generated at high latitudes. When we compare the ensemble spread with the SeaWiFS data, the most striking feature is that the envelope (distance between the minimum and the maximum of the ensemble) captures most of the observations. More precisely, without taking into account any observation error, 70.5% of the 383 827 observation points are inside the ensemble. Note that this ratio is significantly higher if we remove the observations of some regions like the Mediterranean and the Baltic seas where the ensemble is clearly not dispersive enough. These regions have indeed their own particular biogeochemical behavior, very different from the rest of the Atlantic Ocean, so that the generic stochastic parameterizations implemented here might not be completely appropriate. Another major point is that the difference between the maximum and the third quartile is higher than the difference between the first quartile and the minimum. This is mainly the result from the zero lower bound. Where concentrations are small, the perturbations can hardly make it decrease, while extreme values are always high chlorophyll concentrations. In addition, it means that the shape of the stochastic perturbations does not modify the prior lognormal chlorophyll distribution. Fig. 11 also shows that the ensemble is able to simulate a large range of chlorophyll

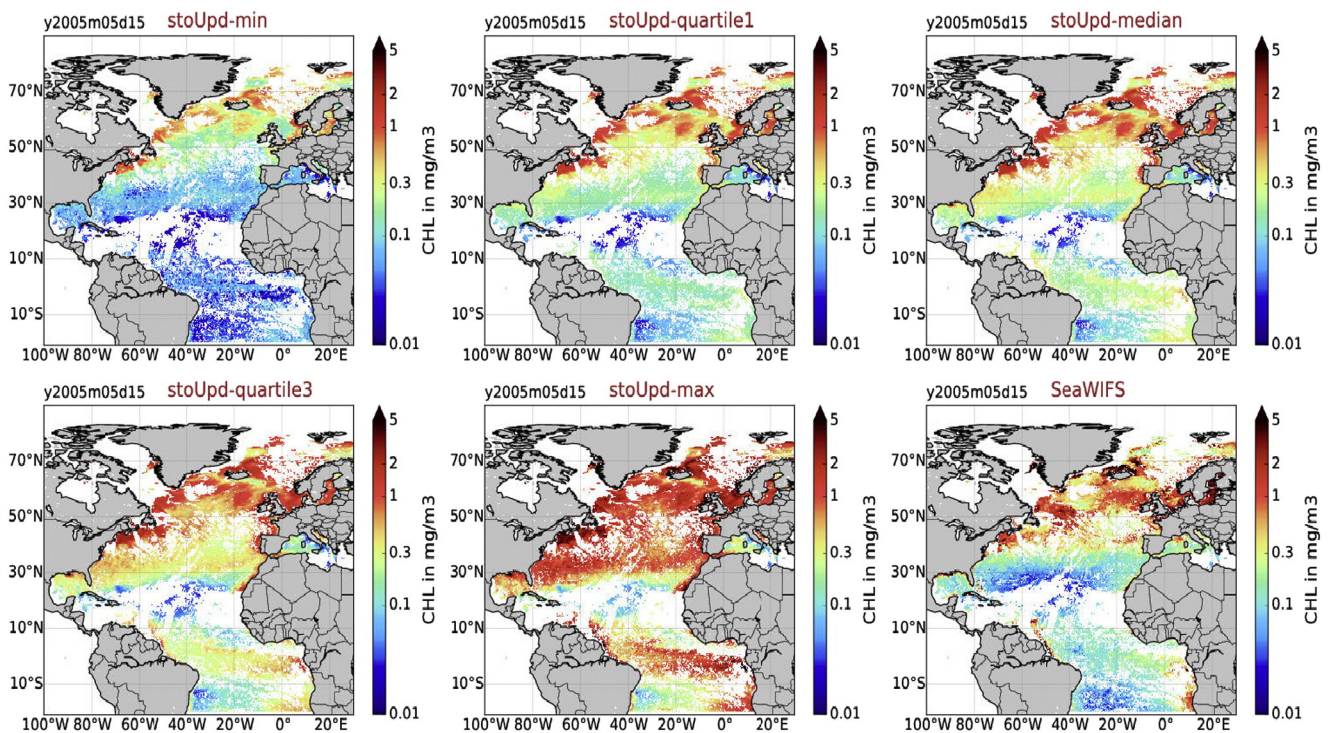


Fig. 11. Surface chlorophyll response to uncertainty as described by the 60-member ensemble simulation for the 15/05/2005. The top panels present the minimum of the ensemble, the 25% percentile, the median and the bottom panels present the 75% percentile and the maximum of the ensemble.

concentrations, which is necessary to include the observations. This confirms the importance of considering a probability density to characterize the ensemble which cannot be only summarized by an ensemble mean and covariances.

To conclude, we have shown in this section that the stochastic parameterizations generate a strong dispersion consistent with the different spatial level of uncertainties. In most places, the envelope defined by the probabilistic representation captures the observations. Thus, this suggests that the perturbation method and the choice of the simulated uncertainties are both quite pertinent.

5.2.2. Rank of observations in the ensemble

To go further, the goal of this section is to perform a more objective comparison of the ensemble with the SeaWiFS satellite data. The term “objective” is employed here in the sense that we evaluate the probabilistic system using standard statistical series based on a very large quantity of observations to assess the quality of the ensemble simulation. The statistical consistency is evaluated with the metric of reliability (Murphy, 1973). It measures the consistency between the probability distribution given by the ensemble and the statistical distribution of the observations. This diagnostic is very often used to evaluate the performance of meteorological forecast systems (Buizza et al., 2004; Palmer et al., 2005) and will be deduced from the shape of rank histograms (Candille and Talagrand, 2005). A flat histogram indicates a perfect reliability.

The rank histograms presented in Fig. 12 are computed as follows: for each grid point and at a given date, all members are sorted in ascending order (relatively to their chlorophyll concentration). Each observation is then ranked relatively to its location within this sorted ensemble. All observations smaller than the minimum of the ensemble take the rank 0 while all observations higher than the maximum of the ensemble take the rank n . Rank histograms therefore represent the cumulative ranks of all available observations at a given date over the domain and contain $n + 1$ bins. The blue line indicates the perfect reliability which means the observation frequency of occurrence required

for each member to have a flat rank histogram. In Fig. 12, together with rank histograms at 3 different dates (31/03/2005, 15/05/2005, 04/07/2005), we also present the rank maps which exhibits the rank of the observation at each grid point. In order to take into account the 30% SeaWiFS observation error (see Section 2.2.1), one solution is to increase the ensemble spread. Thus, ranks are computed as follows.

$$x'_i = x_i + y \quad \text{with} \quad y \sim \mathcal{N}(0, \sigma^2 I) \quad (8)$$

For each realization (observation) a Gaussian white noise with a standard deviation $\sigma = 30\%$ of the SeaWiFS chlorophyll concentration is added to the state x_i of the ensemble member i to obtain a new state x'_i . The observation error is therefore indirectly simulated by computing a new ensemble with an expected larger dispersion. Ranks are then computed using x'_i .

The most important thing to point out is that over the entire domain, with only a 30% SeaWiFS observation error, the distribution of the ranks is not very far from being uniform. If these histograms exhibits a slight underdispersion (too many observations are still out of the ensemble), only 20% of the observations are outside the ensemble during the spring bloom period (15/05/2005) and less than 15% are outside for the 31/03/2005 and the 27/10/2005. Because of a less intense biological dynamics, the dispersion is weaker during autumn and winter periods than during the spring bloom. Once again, we observe that the ensemble dispersion is very dependent to the geographical region. Note that we decided to show global rank histograms, with all available observations, instead of dividing the domain into smaller divisions (like Longhurst provinces for example). These histograms could therefore be flattened and the amount of observations included within the ensemble increased if we remove observations of some areas with singular biogeochemical behaviors (see Section 5.2.1). Finally, the ensemble simulation exhibits high statistical consistencies with observations and can therefore be considered as a well reliable ensemble.

Nevertheless, the shape of these rank histograms also highlights a positive bias emerging during the spring bloom period. Around the 30°N

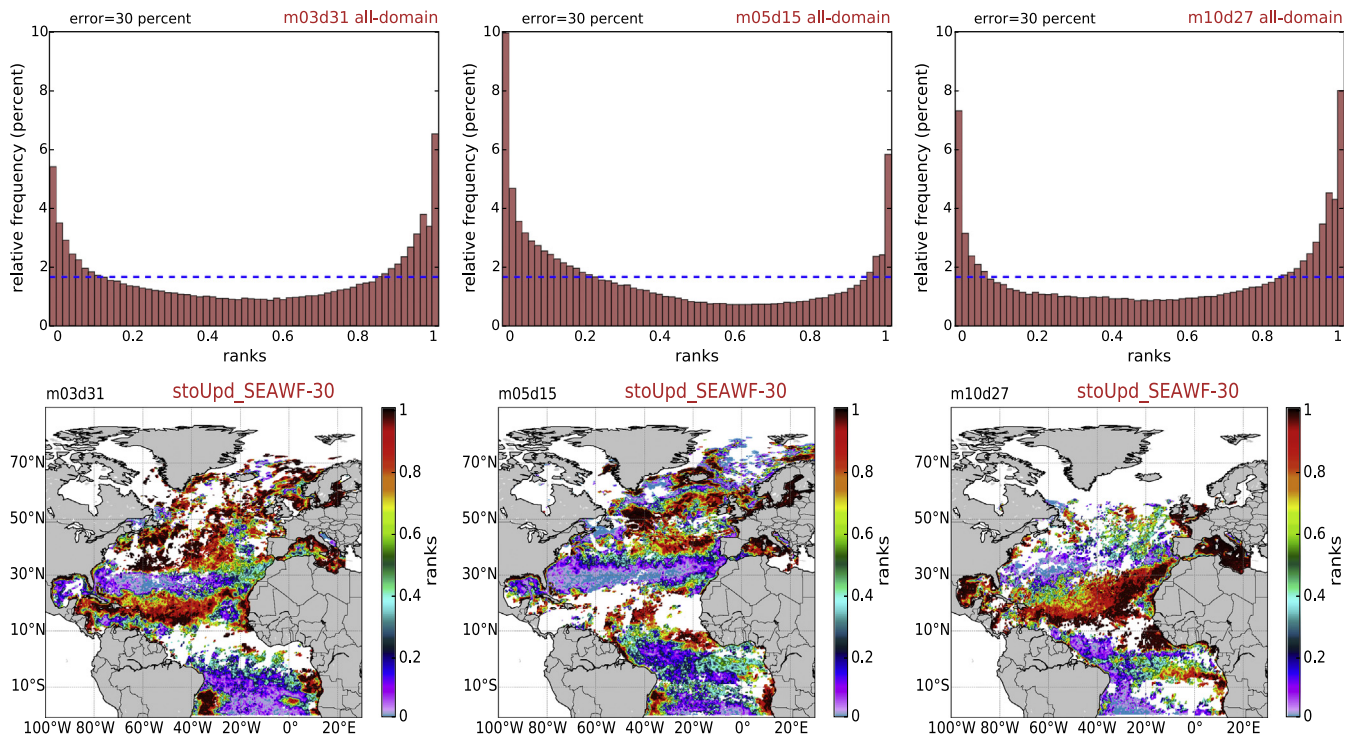


Fig. 12. Top: Surface chlorophyll rank histograms (observations given by SeaWiFS data) over the whole domain for the 31/03/2005 (left), the 15/05/2005 (middle) and the 27/10/2005 (right). A 30% SeaWiFS observation error is taken into account. The blue line refers to the rank of a flat histogram. Bottom: Surface distribution of ranks taking into account a 30% observation error for the corresponding rank histogram dates. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

section or along the Canadian coast, the rank maps show that the model tends to overestimate the chlorophyll concentrations. Fig. 12 presents very heterogeneous rank distributions over the domain with a shape of histograms therefore strongly dependent to the considered geographical area. Note that it was expected since the stochastic parameterizations presented in this paper are now one first original step in the context of biogeochemistry and are still perfectible. For instance, space and time dependent time correlations and standard deviations could be considered. To conclude, despite some imperfections, the ability to obtain a reliable ensemble is an essential property. It means that the stochastic method used to simulate uncertainties is well adapted and that this ensemble simulation should perform consistent predictions.

6. Discussion and conclusions

In this study, we assessed the statistical consistency (reliability) between the ocean color SeaWiFS data and the surface chlorophyll representations of a coupled NEMO/PISCES 60-member ensemble simulation performed with stochastic parameterizations of biogeochemical uncertainties. Starting from a first guess PISCES deterministic simulation, we showed that despite many imperfections, this deterministic simulation was able to describe the basic observed features of the large scale North Atlantic basin in terms of horizontal patterns, vertical structures and temporal variability. To account for the uncertainty of the biogeochemical model and generate an ensemble of model realizations, we investigated a new paradigm for modeling the biogeochemistry by changing PISCES into a probabilistic model. For this purpose, we proposed a generic “built-in” method based on stochastic parameterizations to simulate the joint effects of the unresolved scales and a selection of uncertain parameters. In complement, a re-tuning of some PISCES parameters was necessary to counterbalance the side effects of stochastic perturbations into the original model formulation. Thereafter, a 60-member ensemble simulation performed with these stochastic parameterizations was shown able to produce surface chlorophyll distributions statistically consistent with the SeaWiFS data. The reliability measure of surface chlorophyll distributions indicates that the information of each member is necessary to explain the spatial and temporal variations observed in ocean color satellite fields. Also, it highlights the fact that an ensemble simulation has to be analyzed using diagnostics in agreement with the probabilistic approach.

The statistical consistency with observations is also a prerequisite to build next generation ocean color data assimilation systems. In oceanography, data assimilation is usually performed with sequential techniques derived from Kalman Filter theory like the Sequential Evolutive Extended Kalman (SEEK) filter (Brasseur and Verron, 2006; Pham et al., 1998), thereafter adapted to ocean color assimilation problems (Carmillet et al., 2001; Natvik and Evensen, 2003; Triantafyllou et al., 2007; Butenschön and Zavatarelli, 2012; Fontana et al., 2013, to cite only a few). In these methods, the prior error covariances rely on prior approximations which are not adapted to the biogeochemical context. Moreover, a good specification of error statistics is essential for analysis performances (Nerger et al., 2005). Using a probabilistic approach, data assimilation can be performed with analysis schemes (e.g EnKF, Evensen, 2003) in which the prior error covariances are calculated at each time step using the probability distributions computed by the ensemble simulation. Using the stochastic simulation approach presented in this article, the structure of the error statistics also takes into account uncertainties inherent to the biogeochemical formulations. The probability distribution of the ensemble being coherent with the distribution of the SeaWiFS observations, an improved definition of the prior covariance matrix can be expected. Besides the importance of reliability, the impacts of ocean color data assimilation on the vertical structures are another major difficulty. Indeed, the vertical structure of biogeochemical fields after analysis is strongly related to the prior statistical correlations predicted by the ensemble. For that purpose, the ensemble simulation was calibrated to generate a probabilistic distribution which preserves the main vertical

features expected in the real ocean. In addition, our results indicate that the deterministic vertical structures are strongly dependent to the surface chlorophyll patterns. Thus, we would expect that improvements on the surface chlorophyll representation induced by data assimilation could cascade onto the vertical.

Finally, we emphasize that the stochastic parameterizations presented in this article must be considered as a first step toward more specific probabilistic approaches in ocean biogeochemistry. Several improvements should be investigated further. For instance, standard deviations and time correlations of random processes could be refined according to Longhurst province spatial biogeochemical dependencies. Simulating some physical model uncertainties, known to play an essential role on the biology (Doney et al., 2004; Follows and Dutkiewicz, 2001; Williams et al., 2006), should also allow a better capture of the ocean satellite data in some specific areas like the Gulf Stream region. To conclude, this study shows that considering biogeochemical uncertainties using a probabilistic approach is an emerging field of investigations that should in the long term improve our ability to quantify primary production using models and data in a more consistent way.

Acknowledgments

This work has received funding from the European Community's Seventh Framework Programme FP7/2007–2013 under grant agreements 283367 (MyOcean2), H2020 633085 (MyOcean-FO) and 283580 (SANGOMA), with additional support from CNES. It is also a contribution to the CNRS/INSU/LEFE program. The calculations were performed using HPC resources from GENCI-IDRIS (grant 2015-011279). The authors are grateful to the two anonymous reviewers for their relevant and constructive comments.

References

- Anderson, L., Robinson, A., 2001. Physical and biological modeling in the gulf stream region part II. physical and biological processes. *Deep-Sea Res.* 1 48, 1139–1168. [http://dx.doi.org/10.1016/S0967-0637\(00\)00092-3](http://dx.doi.org/10.1016/S0967-0637(00)00092-3).
- Aumont, O., Bopp, L., 2006. Globalizing results from ocean in situ iron fertilization studies. *Glob. Biogeochem. Cycles* 20. <http://dx.doi.org/10.1029/2005GB002591>.
- Aumont, O., Maier-Reimer, E., Blain, S., Monfray, P., 2003. An ecosystem model of the global ocean including Fe, Si, P colimitations. *Glob. Biogeochem. Cycles* 17. <http://dx.doi.org/10.1029/2001GB001745>.
- Aumont, O., Ethe, C., Tagliabue, A., Bopp, L., Gehlen, M., 2015. PISCES-v2: an ocean biogeochemical model for carbon and ecosystem studies. *Geosci. Model Dev.* 8. <http://dx.doi.org/10.5194/gmdd-8-1-2015>.
- Barnier, B., Madec, G., Penduff, T., Molines, J.-M., Treguier, A.-M., Le Sommer, J., Beckmann, A., Biastoch, A., Böning, C., Dengg, J., Derval, C., Durand, E., Gulev, S., Remy, E., Talandier, C., Theetten, S., Maltrud, M., Mc Clean, J., De Cuevas, B., 2006. Impact of partial steps and momentum advection schemes in a global ocean circulation model at eddy permitting resolution. *Ocean Dyn.* 56. <http://dx.doi.org/10.1007/s10236-006-0082-1>.
- Béal, D., Brasseur, P., Brankart, J.M., Ourmières, Y., Verron, J., 2010. Characterization of mixing errors in a coupled physical biogeochemical model of the north Atlantic: implications for nonlinear estimation using Gaussian anamorphosis. *Ocean Sci.* 6, 247–262.
- Behrenfeld, M., Falkowski, P., 1997. Photosynthetic rates derived from satellite-based chlorophyll concentration. *Limnol. Oceanogr.* 42, 1–20.
- Berges, J., Varela, D., Harrison, P., 2002. Effects of temperature on growth rate, cell composition and nitrogen metabolism in the marine diatom *Thalassiosira pseudonana* (Bacillariophyceae). *Mar. Ecol. Prog. Ser.* 225, 139–146.
- Bopp, L., Aumont, O., Cadule, P., Alvain, S., Gehlen, M., 2005. Response of diatoms distribution to global warming and potential implications: a global model study. *Geophys. Res. Lett.* 32. <http://dx.doi.org/10.1029/2005GL023653>.
- Brankart, J.M., 2013. Impact of uncertainties in the horizontal density gradient upon low resolution global ocean modelling. *Ocean Model.* 66, 64–76. <http://dx.doi.org/10.1016/j.oceomod.2013.02.004>.
- Brankart, J., Candille, G., Garnier, F., Calone, C., Melet, A., Bouttier, P., Brasseur, P., Verron, J., 2015. A generic approach to explicit simulation of uncertainty in the nemo ocean model. *Geosci. Model Dev.* 8, 615–643. <http://dx.doi.org/10.5194/gmdd-8-615-2015> (Discuss).
- Brasseur, P., Verron, J., 2006. The seek filter method for data assimilation in oceanography: a synthesis. *Ocean Dyn.* 56. <http://dx.doi.org/10.1007/s10236-006-0080-3>.
- Brasseur, P., Gruber, N., Barciela, K., Brander, M., Doron, M., El Moussaoui, A., Hobday, A., Huret, M., Krémeur, A.-S., Lehodey, P., Matear, R., Moulin, C., Murtugudde, R., Senina, I., Svendsen, E., 2009. Integrating biogeochemistry and ecology into ocean data assimilation systems. *Oceanography* 22 (3), 206–215. <http://dx.doi.org/10.5670/oceanog.2009.80>.

- Brodeau, L., Barnier, B., Treguier, A.M., Penduff, T., Gulev, S., 2009. An era40-based atmospheric forcing for global ocean circulation models. *Ocean Model.* 31. <http://dx.doi.org/10.1016/j.ocemod.2009.10.005>.
- Buizza, R., Miller, M., Palmer, T., 1999. Stochastic representation of model uncertainties in the ECMWF ensemble prediction system. *Q. J. R. Meteorol. Soc.* 125, 2887–2908.
- Buizza, R., Houtekamer, P., Toth, Z., Pellerin, G., Mzhenh, W., Zhu, Y., 2004. A comparison of the ECMWF, MSC, and NCEP global ensemble prediction systems. *Mon. Weather Rev.* 133, 1076–1097.
- Butenschön, M., Zavatarelli, M., 2012. A comparison of different versions of the seek filter for assimilation of biogeochemical data in numerical models of marine ecosystem dynamics. *Ocean Model.* 54–55, 37–54. <http://dx.doi.org/10.1016/j.ocemod.2012.06.003>.
- Candille, G., Talagrand, O., 2005. Evaluation of probabilistic prediction systems for a scalar variable. *Q. J. R. Meteorol. Soc.* 131, 2131–2150. <http://dx.doi.org/10.1256/qj.04.71>.
- Candille, G., Brankart, J., Brasseur, P., 2015. Assessment of an ensemble system that assimilates Jason-1/Envisat altimeter data in a probabilistic model of the north Atlantic ocean circulation. *Ocean Sci.* 11, 425–438. <http://dx.doi.org/10.5194/os-11-425-2015>.
- Carmillet, V., Brankart, J.M., Brasseur, P., Drange, H., Evensen, G., Verron, J., 2001. A singular evolute extensive kalman filter to assimilate ocean color data in a coupled physical–biogeochemical model of the north Atlantic ocean. *Ocean Model.* 3, 167–192.
- Ciavatta, S., Torres, R., Martinez-Vicente, V., Smyth, T., Dall'Olmo, G., Polimene, L., Allen, J., 2014. Assimilation of remotely-sensed optical properties to improve marine biogeochemistry modelling. *Prog. Oceanogr.* 127, 74–95. <http://dx.doi.org/10.1016/j.pocean.2014.06.002>.
- Doney, S., Lindsay, K., Caldeira, K., Campin, J., Drange, H., Dutay, J., Follows, M., Gao, Y., Gnanadesikan, A., Gruber, N., Ishida, A., Joos, F., Madec, G., Maier-Reimer, E., Marshall, J., Matear, R., Monfray, P., Mouchet, A., Najjar, R., Orr, J., Plattner, G., Sarmiento, J., Schlitzer, R., Slater, R., Totterdell, I., Weiring, M., Yamanaka, Y., Yool, A., 2004. Evaluating global ocean carbon models: the importance of realistic physics. *Glob. Biogeochem. Cycles* 18. <http://dx.doi.org/10.1029/2003GB002150>.
- Doron, M., Brasseur, P., Brankart, J.M., 2011. Stochastic estimation of biogeochemical parameters of a 3d ocean coupled physical–biogeochemical model: twin experiments. *J. Mar. Syst.* 87, 194–207. <http://dx.doi.org/10.1016/j.jmarsys.2013.02.007>.
- Doron, M., Brasseur, P., Brankart, J.M., Losa, S.N., Melet, A., 2013. Stochastic estimation of biogeochemical parameters from global colour ocean colour satellite data in a north Atlantic 3d ocean coupled physical–biogeochemical model. *J. Mar. Syst.* 117–118, 81–85. <http://dx.doi.org/10.1016/j.jmarsys.2013.02.007>.
- Dowd, M., 2011. Estimating parameters for a stochastic dynamic marine ecological system. *Environmetrics* 22, 501–515. <http://dx.doi.org/10.1002/env.1083>.
- Dowd, M., Meyer, R., 2003. A bayesian approach to the ecosystem inverse problem. *Ecol. Model.* 168, 39–55. [http://dx.doi.org/10.1016/S0304-3800\(03\)00186-8](http://dx.doi.org/10.1016/S0304-3800(03)00186-8).
- Dufour, C.O., Le Sommer, J., Gehlen, M., Orr, J.C., Molines, J.M., Simeon, J., Barnier, B., 2013. Eddy compensation and controls of the enhanced sea-to-air CO₂ flux during positive phases of the southern annular mode. *Glob. Biogeochem. Cycles* 27, 950–961. <http://dx.doi.org/10.1002/gbc.20090>.
- Eppley, R., 1972. Temperature and phytoplankton growth in the sea. *Fish. Bull.* 70, 1063–1085.
- Evensen, G., 2003. The ensemble kalman filter: theoretical formulation and practical implementation. *Ocean Dyn.* 53, 343–367. <http://dx.doi.org/10.1007/S10236-003-0036-9>.
- Falkowski, P., 1997. Evolution of the nitrogen cycle and its influence on the biological sequestration of CO₂ in the ocean. *Nature* 387.
- Fichefet, T., Maqueda, A., 1997. Sensitivity of a global sea ice model to the treatment of ice thermodynamics and dynamics. *J. Geophys. Res.* 102, 12,609–12,646 (Doi:10.1481/97JG00480\$09.00).
- Follows, M., Dutkiewicz, S., 2001. Meteorological modulation of the north Atlantic spring bloom. *Deep-Sea Res.* 49, 321–344.
- Fontana, C., Brasseur, P., Brankart, J.M., 2013. Toward a multivariate reanalysis of the north Atlantic biogeochemistry during 1998–2006 based on the assimilation of seaWIFS chlorophyll data. *Ocean Sci.* 9, 37–56. <http://dx.doi.org/10.5194/os-9-1-2013>.
- Frederiksen, J.S., O'Kane, T.J., Zidikheri, M.J., 2012. Stochastic subgrid parametrizations for atmospheric and oceanic flows. *Phys. Scr.* 85.
- Gehlen, M., Bopp, L., Emprin, N., Heinze, C., Ragueneau, O., 2006. Reconciling surface ocean productivity, export fluxes and sediment composition in a global biogeochemical ocean model. *Biogeosciences* 3, 521–537 (www.biogeosciences.net/3/521/2006/).
- Gehlen, M., Barciela, R., Bertino, L., Brasseur, P., Butenschön, M., Chai, F., Crise, A., Drillet, Y., Ford, D., Lavoie, D., Lehodey, P., Perruche, C., Samuelsen, A., Simon, E., 2015. Building the capacity for forecasting marine biogeochemistry and ecosystems: recent advances and future developments. *J. Oper. Oceanogr.* 8 (S1), 168–187. <http://dx.doi.org/10.1080/1755876X.2015.1022350>.
- Geider, R., La Roche, J., 1994. The role of iron in phytoplankton photosynthesis, and the potential for iron-limitation of primary productivity of the sea. *Photosynth. Res.* 39, 275–301.
- Geider, R.J., MacIntyre, H., Kana, T., 1996. A dynamic model of photoadaptation in phytoplankton. *Limnol. Oceanogr.* 41, 1–15.
- Gilstad, M., Sakshaug, E., 1990. Growth rates of ten diatom species from the Barents sea at different irradiances and day lengths. *Mar. Ecol. Prog. Ser.* 64, 169–173.
- Gist, N., Serret, P., Woodward, E., Chamberlain, K., Robinson, C., 2008. Seasonal and spatial variability in plankton production and respiration in the subtropical gyres of the Atlantic ocean. *Atl. Meridional Transect Programm.* 56, 931–940. <http://dx.doi.org/10.1016/j.dsr2.2008.10.035>.
- Gregg, W., Casey, N., 2004. Global and regional evaluation of the seaWIFS chlorophyll data set. *Remote Sens. Environ.* 93, 463–479. <http://dx.doi.org/10.1016/j.rse.2003.12.012>.
- Gregg, W., Friedrichs, M.A.-M., Rose, K., Schlitzer, R., Thompson, K., Doney, S., 2009. Skill assessment in ocean biological data assimilation. *J. Mar. Syst.* 76, 16–33. <http://dx.doi.org/10.1016/j.jmarsys.2008.05.006>.
- Jones, E., Parslow, J., Murray, L., 2010. A bayesian approach to state and parameter estimation in a phytoplankton–zooplankton model. *Aust. Meteorol. Oceanogr. J.* 59, 7–16.
- Jones, E., Parslow, J., Dowd, M., 2013. Summary of discussions: Marine Biogeochemical Data assimilation symposium. CSIRO Marine and Atmospheric Research. Technical Rep (14 pp.).
- Levitus, S., Boyer, T.P., Conkright, M.E., O'Brien, T., Antonov, J., Stephen, C., Stathoplos, L., Johnson, D., Gelfeld, R., 1998. NOAA atlas NESDIS 18.
- Lévy, M., Martin, A., 2013. The influence of mesoscale and submesoscale heterogeneity on ocean biogeochemical reactions. *Glob. Biogeochem. Cycles* 27, 1–12. <http://dx.doi.org/10.1002/2012GB004518>.
- Lévy, M., Mémerly, L., Madec, G., 1998. The onset of a bloom after deep winter convection in the northwestern mediterranean sea: mesoscale process study with a primitive equation model. *J. Mar. Syst.* 16, 7–21.
- Lévy, M., Iovino, D., Resplandy, L., Klein, P., Madec, G., Tréguier, A.M., Masson, S., Takahashi, K., 2011. Large scale impacts of mesoscale dynamics on phytoplankton: local and remote effects. *Ocean Model.* 43, 77–93. <http://dx.doi.org/10.1016/j.ocemod.2011.12.003>.
- Longhurst, A., 1995. Seasonal cycles of pelagic production and consumption. *Prog. Oceanogr.* 36. [http://dx.doi.org/10.1016/0079-6611\(95\)00015-1](http://dx.doi.org/10.1016/0079-6611(95)00015-1).
- Madec, G., the NEMO team, 2012. NEMO Ocean Engine. version 3.4 ed. (URL: <http://www.nemo-ocean.eu>).
- Maranon, E., Holligan, P., Varela, M., Mourino, B., Bale, A., 2000. Basin scale variability of phytoplankton biomass, production and growth in the Atlantic ocean. *Deep-Sea Res.* 47, 825–857.
- Marra, J., 2015. Ocean productivity: a personal perspective since the first liege colloquium. *J. Mar. Syst.* 147, 3–8. <http://dx.doi.org/10.1016/j.jmarsys.2014.01.012>.
- Mattern, J.-P., Fennel, K., Dowd, M., 2010. Introduction and assessment of measures for quantitative model-data comparison using satellite images. *Remote Sens.* 794–818. <http://dx.doi.org/10.3390/rs2030794>.
- Murphy, A., 1973. A new vector partition of the probability score. *J. Appl. Meteorol.* 12, 595–600.
- Natvik, L.J., Evensen, G., 2003. Assimilation of ocean colour data into a biogeochemical model of the north Atlantic part 1. Data assimilation experiments. *J. Mar. Syst.* 40, 127–153. [http://dx.doi.org/10.1016/S0924-7963\(03\)00016-2](http://dx.doi.org/10.1016/S0924-7963(03)00016-2).
- Nerger, L., Hiller, W., Jens, S., 2005. A comparison of error subspace kalman filter. *Tellus* 57A, 715–735.
- O'Reilly, J.E., Maritorena, S., Greg, M.B., Siegel, D.A., Carder, K.L., Garver, S.A., Kahru, M., McClain, C., 1998. Ocean color chlorophyll algorithms for seaWIFS. *J. Geophys. Res.* 103, 24,937–24,953.
- Oschlies, A., Garçon, V., 1998. Eddy-induced enhancement of primary production in a model of the north Atlantic ocean. *Nature* 394, 266–269. <http://dx.doi.org/10.1038/28373>.
- Palmer, T., Shutts, G., Hagedorn, R., Doblas-Reyes, F., Jung, T., Leutbecher, M., 2005. Representing model uncertainty in weather and climate prediction. *Annu. Rev. Earth Planet. Sci.* 33, 163–193. <http://dx.doi.org/10.1146/annurev.earth.33.092203.122552>.
- Pham, D., Verron, J., Roubaud, M., 1998. A singular evolute extended kalman filter for data assimilation in oceanography. *J. Mar. Syst.* 16, 323–340.
- Raven, J., Evans, M., Korb, R., 1999. The role of trace metals in photosynthetic electron transport in CO₂-evolving organisms. *Photosynth. Res.* 60, 111–150.
- Resplandy, L., Lévy, M., Bopp, L., Echevin, V., Pous, S., V.V.S.S., S., Kumar, D., 2012. Controlling factors of the oxygen balance in the Arabian sea's OMZ. *Biogeosciences* 9, 5095–5109. <http://dx.doi.org/10.5194/bg-9-5095-2012>.
- Rodgers, K., Aumont, O., Menkes, C., Gorgues, T., 2008. Decadal variations in equatorial Pacific ecosystems and ferrocline/pycnocline decoupling. *Glob. Biogeochem. Cycles* 22. <http://dx.doi.org/10.1029/2006GB002919>.
- Sauzède, R., Claustre, H., Jamet, C., Uitz, J., Ras, J., Mignot, A., D'Ortenzio, F., 2014. Retrieving the vertical distribution of chlorophyll a concentration and phytoplankton community composition from in situ fluorescence profiles. A Method Based on a Neural Network with Potential for Global-Scale Applications vol. 120. <http://dx.doi.org/10.1002/2014JC010355>.
- Smith, C., Richards, K., Fasham, M., 1996. The impact of mesoscale eddies on plankton dynamics in the upper ocean. *Deep-Sea Res. I Oceanogr. Res. Pap.* 43, 1807–1832. [http://dx.doi.org/10.1016/S0967-0637\(96\)00035-0](http://dx.doi.org/10.1016/S0967-0637(96)00035-0).
- Tagliabue, A., Bopp, L., Aumont, O., 2008. Ocean biogeochemistry exhibits contrasting responses to a large scale reduction in dust deposition. *Biogeosciences* 5, 11–24.
- Tagliabue, A., Bopp, L., Roche, D., Bouttes, M., Dutay, J.C., Alkama, R., Kageyama, M., Michel, E., Paillard, D., 2009. Quantifying the roles of ocean circulation and biogeochemistry in governing ocean carbon-13 and atmospheric carbon dioxide at the last glacial maximum. *Clim. Past* 5, 695–706 (www.clim-past.net/5/695/2009/).
- Takashi, T., Broecker, W., Langer, S., 1985. Redfield ratio based on chemical data from isopycnal surfaces. *J. Geophys. Res.* 90, 6907–6924. <http://dx.doi.org/10.1029/JC090iC04p06907>.
- Triantafyllou, G., Hoteit, L., Petihakis, G., Banks, A., 2007. Assimilation of ocean colour data into a biogeochemical flux model of the eastern mediterranean sea. *Ocean Sci.* 3, 397–410.
- Uitz, J., Claustre, H., Morel, A., S.B., 2006. Vertical distribution of phytoplankton communities in open ocean: an assessment based on surface chlorophyll. *J. Geophys. Res.* 111 (Doi:10.1029).
- Williams, R., Rousseinov, V., Follows, M., 2006. Nutrients streams and their induction into the mixed layer. *Glob. Biogeochem. Cycles* 20. <http://dx.doi.org/10.1029/2005GB002586>.
- Wilson, C., 2010. The rocky road from research to operations for satellite ocean-colour data in fishery management. *J. Mar. Sci.* 71. <http://dx.doi.org/10.1093/icesjms/fsq168>.

