Apports de données Argo pour caractériser les erreurs modèles et contraindre les systèmes d'assimilation

Floriane Ninove

▶ **To cite this version:**

Floriane Ninove. Apports de données Argo pour caractériser les erreurs modèles et contraindre les systèmes d'assimilation. Océan, Atmosphère. Université Paul Sabatier - Toulouse III, 2015. Français. NNT : 2015TOU30230 . tel-01363068

HAL Id: tel-01363068

https://theses.hal.science/tel-01363068

Submitted on 9 Sep 2016

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

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

Université de Toulouse

THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par : *l'Université Toulouse 3 Paul Sabatier (UT3 Paul Sabatier)*

Présentée et soutenue le *17 Novembre 2015* **par :**

Floriane NINOVE

Apports des données Argo pour caractériser les erreurs modèles et contraindre les systèmes d'assimilation

---

JURY

Pierre Brasseur	Directeur de Recherche, LGGE	Rapporteur
Gilles Reverdin	Directeur de Recherche, LOCEAN	Rapporteur
Bruno Zakardjian	Professeur des universités, Toulon	Rapporteur
Frank Roux	Professeur des universités, Toulouse	Examinateur
Pierre-Yves Le Traon	Directeur Scientifique, Ifremer-Mercator Océan	Directeur de thèse
Elisabeth Remy	Chercheur, Mercator Océan	Co-Directrice de thèse

---

École doctorale et spécialité :
SDU2E : Astrophysique, Sciences de l'Espace, Planétologie

Unité de Recherche :
Mercator Ocean/LEGOS

Directeur(s) de Thèse :
Pierre Yves Le Traon et *Elisabeth Remy*

Rapporteur :
Noms des rapporteurs (s'ils ne font pas partie des membres du jury)

Thèse

pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ DE TOULOUSE

par

Floriane Ninove

Apports des données Argo pour caractériser les erreurs modèles et contraindre les systèmes d'assimilation.

Sous la direction de

Pierre-Yves Le Traon (Ifremer/Mercator Océan)

Elisabeth Remy (Mercator Océan)

soutenue le 17 Novembre 2015 devant le jury composé de

Pierre Basseur	LGGE	Rapporteur
Gilles Reverdin	LOCEAN	Rapporteur
Bruno Zakardjian	Université Toulon	Rapporteur
Frank Roux	Université toulouse	Examinateur
Pierre-Yves Le Traon	Ifremer-Mercator Océan	Directeur de Thèse
Elisabeth Remy	Mercator Océan	Co-directrice de Thèse

Remerciements

La reconnaissance est la mémoire du cœur

Parce que chaque expérience dans la vie s'accompagne bien souvent de reconnaissance. Je tiens à remercier toutes celles et ceux qui ont soutenu de près ou de loin ce travail doctoral.

En premier lieu, je remercie Pierre-Yves Le Traon, directeur de thèse et directeur scientifique de Mercator Océan, pour m'avoir accordé sa confiance et son aide tout au long de ma thèse. Je remercie Elisabeth Remy qui a co-encadré ce travail de thèse, pour son soutien et sa patience ! Je suis grandement reconnaissante à messieurs Pierre Brasseur, Gilles Reverdin et Bruno Zakardjian d'avoir accepté de prendre connaissance de mes travaux et de les avoir évaluer sous l'angle de leur expertise aiguisée, didactique et appliquée. Je tiens à remercier également Frank Roux pour avoir présidé ce jury de thèse et avoir su y insuffler une atmosphère aussi détendue qu'académique.

Je remercie l'Ifremer qui a financé cette thèse en m'accordant un poste d'allocataire de recherche ainsi que Mercator Océan qui m'a accueilli et accompagné pendant ces années de thèse.

Un mot pour cette grande famille évoluant à Mercator Océan, j'ai eu plaisir à apprendre et travailler parmi eux. Un grand merci à l'équipe R&D de Mercator Océan, pour leur bienveillance et leur bonne humeur. De chaleureux remerciements à Charly Reigner, Marie Drevillon, Jean-Michel Lellouche, Gilles Garric. Merci à la très sympathique équipe Info toujours présente et efficace : Mathieu Peltier, Carine Castillon, Mondher Chekki et Bertand Ferret. Charles mon cher co-bureau, merci pour nos interminables discussions, ton écoute et tes conseils. Merci à Julien pour tes multiples conseils, tes blagues et ta patience lorsque la mienne semblait épuisée ! Merci à Geneviève Chabot, Véroniques Landes, Alice Vu pour nos discussions de filles et vos attentions. Merci aux sportif(ve)s Mercatoriens, nos footings le long du canal du Midi me manqueront : Lucas Nouel, Guillaume Reffray, Elodie Gutknecht.

Un merci particulier à Cedric Giordian pour ta force de vie et ta bonne humeur quotidienne. Puis une pensée à tous les non-permanents, stagiaires, thésards, post-docs, voisins de bureaux avec qui j'ai pu travailler, discuter, rire, déguster des spécialités gastronomiques diverses, prendre le temps de ne rien faire... merci à Jonathan, Cyrille, LZ, Anne D., Stephane, Guillaume S., Giovanni, Angélique, Simon, Matthieu, Victor. Merci à Vanessa D., soeur de coeur, compagne de galère merci pour nos fous rires, notre premier colloque à Venise. Merci aux amis de promo, Vanessa M., Anne F., à la Pic Team rencontrée aux Doctoriales Albigeoises.

L'équilibre entre vie personnelle et professionnelle est semble t-il indistinct dans la vie d'un thésard mais je tiens à remercier Emilie et Charlotte Jacq pour leur bulle d'Art dans cet épisode studieux : merci pour votre énergie infinie, votre vision de la vie et de la danse ainsi que notre aventure in New York City.

Enfin, merci à ma famille, à mes parents, grands-parents et mon grand frère pour leur soutien inconditionnel. Merci à Caroline et mon amour de neveu Simon qui ont su me rappeler l'essentiel. Je souhaiterais terminer ces remerciements par la personne qui a partagé au fil des années cette aventure : merci Antoine pour ton soutien, ton aide, et ton réconfort.

à Guy Ninove,

Résumé

Le programme international Argo a révolutionné l'observation globale des océans. Une flotte de plus de 3000 instruments sous-marins autonomes est en place, programmés pour fournir des mesures globales de profils de température et de salinité sur les 2000 premiers mètres de l'océan. Ces mesures sont assimilées dans des modèles océaniques conjointement aux observations satellitaires afin de décrire et prévoir l'océan. Dans cette thèse nous proposons une analyse permettant de caractériser les erreurs d'un modèle global océanique par comparaison aux données du réseau Argo. Ces erreurs sont décrites via leurs amplitudes, leurs variations régionales et temporelles ainsi que via les échelles spatiales associées. Une caractérisation des échelles spatiales à la fois pour la variabilité océanique et pour les erreurs modèles est, en particulier, menée permettant de relier la structure des erreurs modèles à celle du signal. Enfin, des techniques basées sur le contenu en information sont testées afin de pouvoir à terme mieux quantifier l'impact des observations Argo sur les systèmes d'assimilation de Mercator Océan.

Abstract

The international Argo program has revolutionized the observation of the global ocean. An array of more than 3000 profiling floats is in place and provides global measurements of temperature and salinity on the first 2000 meters of the ocean. These measurements are assimilated into ocean models together with satellite observations to describe and forecast the ocean state. We propose here to characterize model errors using Argo observations. Model errors are described through their amplitude, geographical and temporal variations as well as their spatial scales. Spatial scales of both model errors and ocean signals are, in particular, estimated. This allows a comparison of model errors and ocean variability structure. Finally, techniques based on information content are tested in the longer run quantifying the impact of Argo observations in the Mercator Ocean data assimilation systems.

Table des matières

Remerciements

Résumé

Abstract

1 Introduction générale **2**
- 1.1 Le Système Global d'Observation des Océans . . . 3
- 1.2 Apport des données Argo . . . 4
- 1.3 Objectifs de la thèse . . . 6
- 1.4 Structure du manuscrit . . . 8

2 Le programme international Argo et le centre d'océanographie opérationnelle Mercator Ocean **10**
- 2.1 Le programme international Argo . . . 11
  - 2.1.1 Les objectifs . . . 11
  - 2.1.2 Cycle de plongée et flux de données Argo . . . 11
  - 2.1.3 Résultats prometteurs du réseau Argo . . . 13
  - 2.1.4 Participation très active de la France . . . 14
  - 2.1.5 Les nouveaux défis pour Argo . . . 14
- 2.2 Mercator Ocean, Centre francais d'analyse et de prévision océanique 15
  - 2.2.1 Histoire et projets européens . . . 15
  - 2.2.2 Modèle océanique et systèmes d'assimilation . . . 16

3 Caractérisation des écarts modèle observations pour une simulation sans assimilation **19**
- 3.1 Méthodes statistiques . . . 20
- 3.2 Caractéristiques propres du modèle . . . 24
- 3.3 Présentation des observations in-situ Argo . . . 26
- 3.4 Evaluation des écarts entre la simulation et les observations in-situ Argo . . . 29

3.4.1 Séries temporelles des écarts moyens . . . 29
3.4.2 Moyennes et variances en température et salinité sur la période 2006-2009 . . . 31
3.4.3 Moyennes saisonnières en température et salinité sur la période 2006-2009 . . . 37

4 Caractérisation des échelles spatiales des erreurs modèles et de la variabilité océanique pour la température et la salinité 41
4.1 Introduction . . . 43
4.2 Méthodologie pour l'estimation des échelles spatiales . . . 44
4.2.1 Discrétisation des covariances . . . 44
4.2.2 Ajustement d'une fonction analytique de corrélation . . . 46
4.2.3 Validation de la méthodologie à partir de données simulées . 47
4.2.4 Conclusion . . . 55
4.3 Caractérisation des échelles spatiales des erreurs modèles et du signal à partir des données Argo . . . 56
4.3.1 Structures et caractéristiques des régions examinées . . . 56
4.3.2 Utilisation d'une nouvelle formulation analytique du modèle de corrélation . . . 59
4.3.3 Analyse globale des échelles spatiales des erreurs modèles en température et salinité . . . 59
4.3.4 Comparaison des échelles spatiales des erreurs modèles par rapport aux échelles de la variabilité océanique . . . 63
4.4 Estimation des échelles spatiales des variations de température et salinité à partir des données Argo . . . 71
4.4.1 Article paru dans le journal *Ocean Science* . . . 71
4.4.2 Erreur formelle sur l'estimation de la covariance . . . 87
4.5 Conclusion . . . 89

5 Analyse des méthodes basées sur le contenu en information et étude de sensibilité 90
5.1 Introduction . . . 91
5.2 Mesure d'impact des observations sur l'analyse . . . 92
5.2.1 Définition analytique . . . 92
5.2.2 Définition dans le cadre du BLUE . . . 92
5.2.3 Definition dans le cadre de l'analyse objective . . . 93
5.2.4 Méthodes alternatives pour l'approximation du DFS . . . 94
5.3 Etude de sensibilité du contenu en information . . . 96
5.3.1 Simulation de distributions gaussiennes et estimateur de Gauss-Markov . . . 97
5.3.2 Tests d'influence sur le DFS . . . 98

5.4 Comparaison des estimations du contenu en information 105
5.4.1 Illustration à partir d'une application au système d'assimilation de Mercator Océan 105
5.4.2 Mise en pratique dans le cadre d'analyse objective 106
5.5 Conclusion 109

Conclusion et perspectives **110**

Annexes **113**

A Catalogue des biais annuels et saisonniers modèle-observations sur la période 2006-2009 **114**

B Catalogue des échelles spatiales de corrélation des erreurs modèles et de la circulation océanique **126**

C Méthode d'assimilation BLUE **137**

Chapitre 1

Introduction générale

Sommaire

1.1 Le Système Global d'Observation des Océans 3
1.2 Apport des données Argo 4
1.3 Objectifs de la thèse 6
1.4 Structure du manuscrit 8

1.1 Le Système Global d'Observation des Océans

L'océan, difficilement observable par l'homme, est resté longtemps méconnu. Ce n'est qu'à partir du XIX^{me} siècle que les premières campagnes d'explorations maritimes autant européennes qu'américaines débutent notamment avec la campagne océanographique anglaise du Challenger (1872 – 1876). D'ambitieux programmes internationaux d'observation de l'océan ont été menés dans les années 80 et 90 (TOGA, WOCE) et ont conduit à des avancées majeures dans la compréhension de l'océan et de son rôle sur le climat. La mise en place depuis une dizaine d'années des moyens nouveaux d'observations temps réel (satellites, instruments autonomes in-situ dont le réseau global de flotteurs profilants Argo) ainsi que les progrès sur la modélisation numérique ont permis plus récemment le développement de l'océanographie opérationnelle et des services océaniques associés.

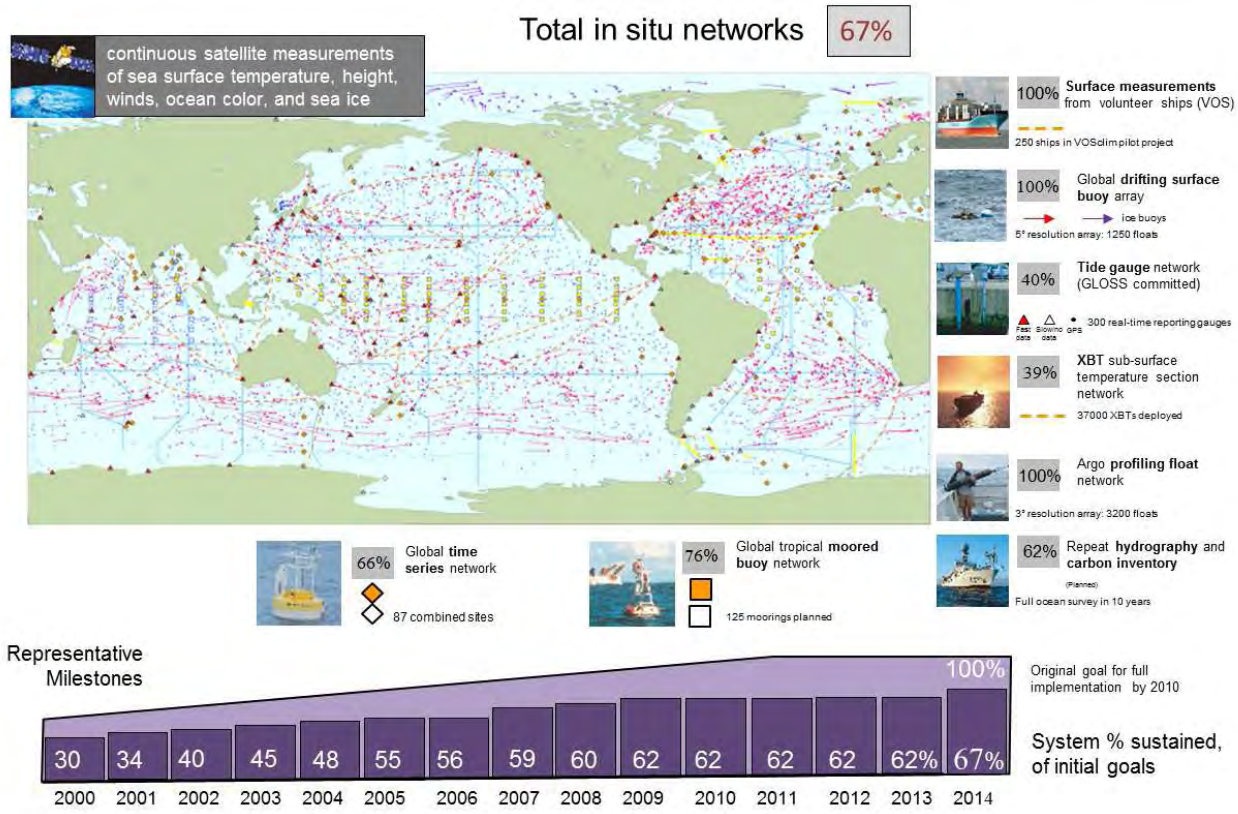


Figure 1.1 – *Système Global d'Observation des Océans (GOOS) et état de développement des différentes composantes in-situ par rapport aux objectifs initiaux définis par la communauté océanographique il y a* 15 *ans.*

Un système global d'observation des océans performant est indispensable pour observer, comprendre et prévoir l'océan et son rôle sur le climat. Le Système Global d'Observation des Océans actuel (GOOS, http ://www.ioc-goos.org/, *Dexter and Summerhayes* [2010]) résulte de la juxtaposition et de l'intégration progressive de systèmes de mesures variés et complémentaires (mesures in-situ et satellites, voir la figure 1.1). Ces systèmes d'observation sont indispensables aux centres d'analyses et de prévisions océaniques. Le centre français Mercator Ocean assimile ainsi en temps réel l'ensemble de ces données dans des modèles océaniques globaux et régionaux.

Cette thèse s'inscrit dans les activités de l'Ifremer [1] en collaboration avec Mercator Océan autour du programme Argo et ses évolutions. Dès le début des années 2000, le programme Argo s'amorçe. L'année 2007 le réseau contient plus de 3000 flotteurs, fournissant des mesures globales de profils de température et de salinité sur les 2000 premiers mètres de l'océan. Le déploiement d'instruments sous-marins autonomes et programmés à dériver sur toutes les mers du globe est sans précédent dans l'histoire des mesures in situ. Les observations Argo possèdent, en outre une forte complémentarité avec les observations satellitaires. C'est une des raisons pour laquelle le réseau Argo est une composante essentielle du système mondial d'observation des océans et du climat permettant de :

- détecter la variabilité climatique aux échelles saisonnières et décennales, et observer à long terme le changement climatique dans les océans ;
- fournir des données essentielles pour contraindre des modèles d'analyse et de prévision océanique (Service Marin de Copernicus [2], Mercator Océan), afin d'initialiser des modèles couplés océan-atmosphère de prévision saisonnière et décennale et de valider les modèles climatiques ;
- fournir l'information nécessaire à l'étalonnage des capteurs et la validation des données satellitaires.

1.2 Apport des données Argo

L'observation opérationnelle de l'océan est ainsi devenue une réalité. La capacité d'observer l'océan en temps réel via les mesures des satellites et via les réseaux in-situ comme Argo ont permis la mise en place de modèles d'analyse et de prévision de tout l'océan analogues aux modèles météorologiques. Ainsi le centre Mercator Océan fournit maintenant en routine des analyses et prévisions océaniques sur l'ensemble des océans. Des cartes de prévisions à 5, 7 ou 14 jours des courants, température, salinité sont actualisées de façon quotidienne ou heb-

1. L'Institut Français de Recherche pour l'Exploitation de la Mer
2. Programme européen d'observation et surveillance de la Terre

domadaire. A ce jour le réseau Argo représente une des plus importantes sources de données in-situ en assurant une couverture globale spatiale et temporelle. Le nombre d'études utilisant les données Argo grandit et des résultats notables ont été obtenus [3]. Les données Argo sont également utilisées de manière systématique au sein des systèmes d'océanographie opérationnelle. Il est donc nécessaire de continuer acquérir une bonne connaissance de "la donnée Argo" et de ses limitations pour une utilisation optimale et une pérennisation du réseau.

Les flotteurs Argo ont permis de combler les manques significatifs présents dans le réseau global d'échantillonnage de l'océan (*Roemmich and the Argo Steering Team* [2009]). Ainsi, ils fournissent des informations très utiles quant à l'estimation de l'état en surface des océans et de la structure 3D de la température et de la salinité. Les cartes globales du niveau de la mer obtenues par altimétrie satellitaire, à des échelles hebdomadaires comme annuelles, peuvent maintenant être interprétées grâce à une meilleure connaissance complète de la structure tridimensionnelle des champs de température et de salinité. Dès lors les modèles océaniques globaux et climatiques peuvent être plus précisément initialisés, testés, et contraints. (*Guinehut et al.* [2012]. De plus *Juza et al.* [2011b,a] ont montré, en particulier, que le réseau Argo est un système d'observations in-situ bien adapté à l'étude d'indices climatiques. Notamment il capture une grande partie de la variabilité saisonnière et interannuelle du contenu thermique de l'océan global.

Depuis le déploiement du réseau Argo la communauté océanographique a démontré les apports essentiels des données Argo dans les systèmes d'assimilation de données et de modélisation numérique, ainsi que leur cohérence avec les données altimétriques et les autres types de données in-situ (documentés dans *Roemmich and the Argo Steering Team* [2009]). Les études conduites par la communauté au sein de GODAE [4] détaillées dans *Oke et al.* [2009] exposent les différentes approches utilisées pour quantifier l'impact de ces observations dans des systèmes avec assimilation de données. Il est usuel de quantifier l'impact des observations au sein d'un système de prévisions/analyse sous la forme d'OSEs [5] (*Oke and Schiller* [2007], *Smith and Haines* [2009] et *Oke et al.* [2014]). Mercator Océan met régulièrement en œuvre ces techniques pour quantifier l'apport des données Argo. Les récentes études menées (*Turpin et al.* [2015]) mettent en évidence l'apport très important des données Argo au niveau global et régional. Une réduction de la couverture du réseau Argo dégrade systématiquement et de façon marquée la qualité des analyses et prévisions océaniques. D'autre part les résultats d'une étude réalisée au CEP-

3. voir l'extrait de la bibliographie des articles publiés jusqu'en 2010 sur le site officiel, http ://www.argo.ucsd.edu
4. Global Ocean Data Assimilation Experiment
5. Observing System Evaluations

MMT [6], basée sur la même approche que celle décrite pour le système de Mercator Océan montrent l'importance des observations Argo pour initialiser des modèles couplés océan/atmosphère pour la prévision saisonnière (*Oke et al.* [2014]). Enfin l'OSE utilisant le système MOVE-G d'assimilation de données japonais souligne que l'impact des données Argo est plus important en salinité qu'en température. Ce résultat est cohérent avec les conclusions des études passées (*Oke and Schiller* [2007]).

1.3 Objectifs de la thèse

Les systèmes d'observations océaniques fournissent aux océanographes un grand nombre d'estimations de diverses quantités géophysiques (température, salinité, niveau de surface de la mer, vitesses des courants ...) à des résolutions spatiales et temporelles hétéroclites. De telles observations peuvent être utilisées pour créer des références communes sous forme d'atlas climatologiques construits sur de longues périodes (World Atlas Ocean documenté dans *Levitus et al.* [1998, 2005, 2013]). La description de l'état complet de l'océan et la prévision de son évolution nécessitent cependant d'intégrer ces observations dans des systèmes d'assimilation. On combine alors les observations avec un modèle océanique. Par conséquent caractériser les erreurs du modèle est alors primordial. Celles-ci peuvent être évaluées par comparaison entre la quantité simulée et la quantité observée. On attend de la part des modèles océaniques qu'ils parviennent à représenter certaines caractéristiques décisives de l'état et de la variabilité physique de l'océan. Grâce aux données Argo il est maintenant possible d'évaluer ces caractéristiques à différentes profondeurs en température et en salinité. Dans cette thèse nous confrontons les observations Argo et une simulation numérique libre dans le but d'évaluer la capacité d'un modèle de circulation générale au 1/4° dont elle est issue, à reproduire correctement les caractéristiques principales de la circulation océanique.

Les études *Kamenkovich et al.* [2009, 2011], *Vecchi and Harrison* [2005], *Guinehut et al.* [2012] et *Oke et al.* [2014] s'intéressent aux capacités de reconstructions de divers champs océaniques à l'aide des observations Argo. Certains facteurs tels que les mouvements des flotteurs, leurs cycles d'échantillonnage, les tourbillons à méso-échelles peuvent influencer la qualité et la précision des résultats obtenus. Analyser de tels facteurs est essentiel à la compréhension des limitations du système d'observations Argo. Certaines étapes préliminaires aux estimations statistiques réalisées dans cette thèse ont été effectuées dans le but d'évaluer l'impact de facteurs tels que le nombre et les mouvements des flotteurs selon les régions ob-

6. Centre Européen pour les Prévisions Météorologiques à Moyen Terme

servées. Nos choix en ce qui concerne l'approche d'estimation en tiendront compte.

Plusieurs études témoignent des nouvelles analyses scientifiques rendues possibles grâce au réseau global Argo. Cependant ce potentiel implique un besoin de validation et de contrôle de qualité au sein de ce large jeu de données. En effet, *Gaillard et al.* [2009] et *Roemmich and Gilson* [2010] soulignent que connaître les variations des échelles spatiales en température et salinité est nécessaire pour améliorer les contrôles de qualité mais aussi pour caractériser au mieux la dynamique océanique et spécifier les schémas d'assimilation de données. L'un des objectifs de cette thèse concerne la caractérisation des échelles spatiales et temporelles en température et salinité de la surface à 1500 m, et il a été motivé par plusieurs faits :

- L'assimilation de données permet en pratique de déterminer la meilleure combinaison entre les observations et un des états à priori de l'océan représenté par un modèle numérique. Une part importante en assimilation de données consiste à déterminer dans quelle mesure le modèle est réaliste. Il s'agit de lui affecter un poids approprié dans le but d'approcher la réalité et d'améliorer les capacités d'analyse et de prévision du modèle. Ce poids doit prendre en compte les échelles spatio-temporelles des erreurs modèles (qui sont souvent approximées par les échelles du signal océanique lui même).
- Le système d'assimilation de Mercator Ocean basé sur la formulation SEEK utilise des informations sur les échelles de corrélation. En effet une caractéristique importante au sein de cette formulation est la spécification de la matrice de covariance d'erreurs de prévision du modèle. Sa définition repose sur une technique de localisation impliquant l'utilisation d'échelles de corrélation spatiales. Parce que la matrice de covariance d'erreurs de prévision du modèle est construite à partir d'un nombre fini d'anomalies de l'état de l'océan, on choisit de mettre à zéro les covariances à partir d'une certaine distance définie comme étant deux fois l'échelle de corrélation spatiale. Ce choix est légitime du fait que l'ensemble fini des anomalies n'est plus significatif à partir d'une certaine distance du point d'analyse. Ces échelles ont été estimées jusqu'à présent en surface via les anomalies des hauteurs d'eau observées par satellites (SLA, SSALTO/DUACS), et à faible profondeur (100 et 300 m) à partir de champs $2D$ en température issus de la ré-analyse globale au 1/4° GLORYS2V1 (GLobal Ocean ReanalYsis ans Simulation, *Ferry et al.* [2012]; *Lellouche et al.* [2013]). Cependant *Kessler et al.* [1996], *Jacobs et al.* [2001] ont souligné des dépendances en fonction des latitudes et de la profondeur au sein des échelles spatiales. Il s'agit donc de poursuivre au mieux la caractérisation des échelles spatiales en température et salinité de la surface à 1500 mètres, ainsi que leur variation en fonction de la profondeur,

et des latitudes et longitudes.
- Par ailleurs ces échelles permettent aussi de définir une "bulle d'influence" au voisinage du point d'analyse et d'y sélectionner les données utilisées (*Lellouche et al.* [2013]). Les observations utilisées dans le cadre de l'assimilation de données possèdent une contribution différentes selon leurs nature et qualité (*Turpin et al.* [2015]). Il est donc intéressant de savoir mesurer objectivement la quantité d'information apportée par les données Argo par comparaison aux observations de natures différentes composants les jeux de données assimilés.

On dispose également d'outils permettant d'évaluer la contribution relative des observations dans n'importe quel système d'assimilation de données : l'un d'eux est le DFS pour Degrees of Freedom. La notion de DFS permet d'estimer la contribution relative des observations mais il est aussi intéressant de comprendre le lien entre les échelles de corrélation spatiales et la contribution relative des observations. Cette dépendance est illustrée par la suite avec un exemple simplifié pour ensuite tester différentes méthodes de calcul approximatif du DFS applicables facilement à des systèmes d'assimilation tel que celui opéré à Mercator Océan.

1.4 Structure du manuscrit

Dans cette thèse j'ai tout d'abord exploité les observations hydrographiques issues du réseau Argo, pour caractériser les erreurs d'un modèle global au 1/4° n'ayant pas assimilé de données Argo (chapitre 3). Dans ce contexte, il a fallu utiliser des méthodes d'interpolation afin d'extraire de la simulation numérique choisie pour l'étude, le pendant exact (en temps et en espace) des observations Argo. Cette étape correspond à la colocalisation au point d'observation des sorties modèles en température et salinité. Dans un deuxième temps j'ai effectué un traitement statistique en temps et en espace dans le but d'obtenir les erreurs modèles moyennes ainsi que les variances d'erreurs en température et salinité en surface et en profondeur. L'analyse de ces résultats permet d'évaluer le comportement de la simulation numérique au regard des observations Argo.

La variabilité océanique possède une large gamme d'échelles spatiales et temporelles. Afin d'étudier plus finement la manière dont cette simulation numérique reconstruit les champs océaniques de température et de salinité, nous avons caractérisé les échelles spatiales des erreurs modèles et du signal océanique lui même (chapitre 4). J'ai développé une méthode pour l'estimation de ces échelles spatiales à partir des observations Argo. Une analyse de l'impact de l'échantillonnage du réseau Argo sur l'estimation de ces échelles grâce à des simulations et un calcul d'erreur formelle ont été menés. Les échelles spatiales de corrélation en température

et en salinité de la surface à 1500 m ont été, enfin, calculées à partir d'anomalies par rapport à la climatologie et par rapport aux sorties modèles sur une période de plusieurs années. Les résultats ont été analysés afin de mettre en évidence les variations géographiques de ces échelles, les variations en profondeur et les variations entre température et salinité.

Dans un troisième temps, j'ai analysé les méthodes basées sur le contenu en information. Ces méthodes permettent de mesurer directement l'impact ou la contribution relative des observations dans un système d'assimilation comme celui mis en œuvre à Mercator Ocean. J'ai mené une étude de sensibilité (bruit de mesure, longueur de corrélation) et testé différentes méthodes permettant d'approximer le contenu en information dans un contexte d'analyse objective. Ces tests ont permis de déterminer la formulation qui conviendrait le mieux à une implémentation dans le système d'assimilation de données de Mercator Ocean.

Chapitre 2

Le programme international Argo et le centre d'océanographie opérationnelle Mercator Ocean

Sommaire

2.1 Le programme international Argo 11
2.1.1 Les objectifs 11
2.1.2 Cycle de plongée et flux de données Argo 11
2.1.3 Résultats prometteurs du réseau Argo 13
2.1.4 Participation très active de la France 14
2.1.5 Les nouveaux défis pour Argo 14
2.2 Mercator Ocean, Centre francais d'analyse et de prévision océanique 15
2.2.1 Histoire et projets européens 15
2.2.2 Modèle océanique et systèmes d'assimilation 16

2.1 Le programme international Argo

2.1.1 Les objectifs

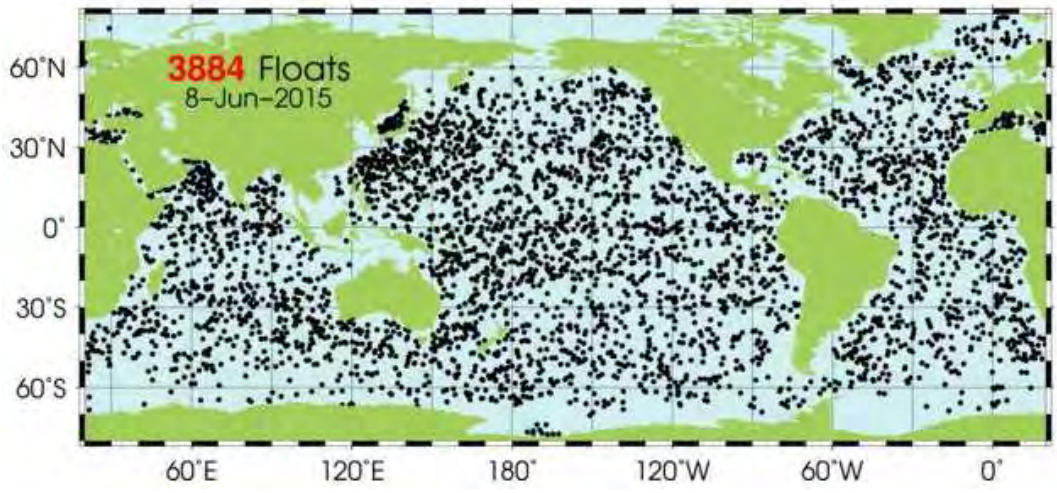
La recherche sur l'évolution du climat de la planète et les effets du changement climatique ont un besoin impératif d'observations globales sur le long terme et de très bonne qualité. C'est également indispensable pour les applications et services de l'océanographie opérationnelle (*Bell et al.* [2010]). L'océanographie opérationnelle dépend essentiellement de la disponibilité en temps quasi-réel de données in-situ et satellites de haute qualité, ceci combiné à un échantillonnage spatio-temporel suffisamment dense. La première exigence pour l'obtention de telles données est la mise en oeuvre d'un système d'observation de l'océan global. Lancé en 2000 par la Commission Océanographique Intergouvernementale (COI) et l'Organisation Météorologique Mondiale (OMM), le programme Argo figure comme un élément essentiel de ce système global d'observation des océans [*Roemmich and the Argo Steering Team*, 2009; *Freeland et al.*, 2010].

La "radiosonde océanographique" Argo est un concept révolutionnaire qui améliore la capacité de mesures des température et salinité en temps réel à travers l'ensemble des océans et ceci jusqu'à 2000 mètres de profondeur. Fin 2007, le projet a atteint son objectif initial avec 3000 flotteurs en opération pour une résolution spatiale de 3°x3°. Argo est le fruit d'une coopération internationale remarquable dans laquelle plus de 30 pays participent directement à la mise en place du réseau. Chaque année 800 à 900 de ces instruments sont mis à l'eau, soit le nombre nécessaire pour maintenir un réseau de 3000 flotteurs actifs compte tenu de leur durée de vie estimée à 4 ans. Par exemple pour le mois de juin 2015 on peut observer sur la figure 2.1 la répartition géographique des flotteurs.

2.1.2 Cycle de plongée et flux de données Argo

Les flotteurs Argo dérivent au grè des courants à une profondeur nominale de 1000 mètres. Tous les 10 jours, ils exécutent un cycle identique préprogrammé de descente, dérive en immersion, descente à l'immersion maximale et remontée. Pendant les quelques heures de remontée, ils mesurent la température et la salinité (via la conductivité) de 2000 mètres à la surface. En surface, les données sont alors transmises aux satellites (ARGOS ou Iridium) obtenant une série des positions des flotteurs et de profils de mesure. Lorsque cette transmission est terminée, le flotteur revient à sa profondeur de parking jusqu'au prochain cycle. Les données Argo sont disponibles selon deux modalités :

- en temps réel, où seules des corrections automatiques sont faîtes,
- en temps différé, où des corrections sont réalisées par des experts scientifiques.



(a)

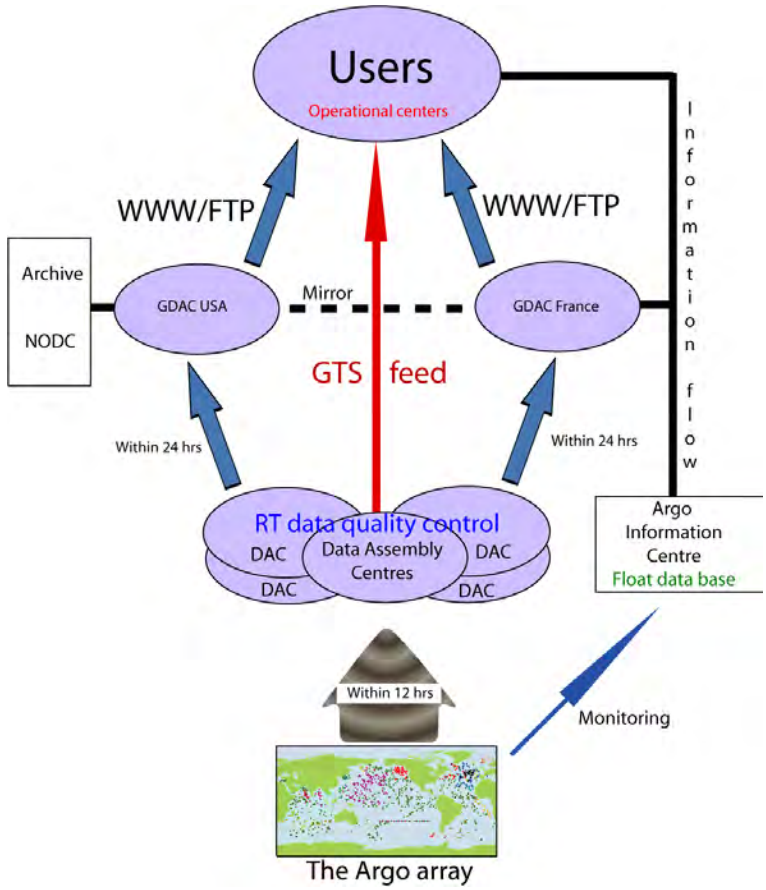
Figure 2.1 – *Positions des flotteurs Argo pour le mois de juin 2015*

Il n'existe qu'une seule base de données où la meilleure version de chaque profil Argo est conservée. Comme le montre la figure 2.2, les acteurs principaux dans la gestion des données Argo sont

- les 11 DACs (Data Assembly Centers) qui reçoivent les données via transmission satellite. Celles-ci sont soumises à 19 tests automatiques en temps réel. Les données erronnées sont flaggées et si possible corrigées. Les données en temps réel sont placées sur le GTS [1] qui livre des données aux centres Argo globaux (GDACs).
- les 2 GDACs [2], localisés à Coriolis en France, et au FNMOC [3] aux Etats-Unis sont chargés de collecter les données Argo traitées par les 11 DACS et de fournir aux utilisateurs un accès unique aux meilleures versions des profils Argo.

Le système de gestion Argo permet aussi de fournir des données climatiques de qualité en temps différé, grâce à l'utilisation combinée des outils statistiques et de la forte implication d'experts scientifiques dans le processus de contrôle de qualité. L'ensemble des procédures de contrôle de qualité (en temps réel et différé) sont décrits dans *Wong et al.* [2014].

1. Global Telecommunication System
2. Global Data Assembly Centers
3. Fleet Numerical Meteorology and Oceanography Center



(a)

Figure 2.2 – *Flux de données Argo et acteurs en temps réel des traitements de données.*

2.1.3 Résultats prometteurs du réseau Argo

Une revue de l'apport des données Argo pour l'océanographie est donnée dans *Freeland et al.* [2010]. Les utilisations de ces données sont nombreuses et variées. Les données Argo ont tout d'abord un apport majeur pour l'analyse et la prévision océanique. Elles sont systématiquement utilisées dans les systèmes d'océanographie opérationnelle et possèdent un fort impact sur la qualité des analyses et prévisions océaniques (*Martin et al.* [2014]; *Turpin et al.* [2015]). Les observations Argo sont aussi très utiles pour l'initialisation de modèles couplés océan/atmosphère pour la prévision saisonnière et, à terme, la prévision décennale. Enfin le réseau Argo est

devenu un outil indispensable pour l'étude du changement climatique et du rôle de l'océan sur le climat. Les données permettent de caractériser par exemple, très précisément le réchauffement de l'océan (*Von Schuckmann and Le Traon* [2011]) et les évolutions de salinité liées au changement du cycle hydrologique (*Durack and Wijffels* [2010]).

2.1.4 Participation très active de la France

La France est très active dans tous les aspects du programme Argo : développement de l'instrumentation (flotteurs Provor et Arvor et leurs dérivés), contribution au réseau par le déploiement de 60 à 80 flotteurs/an (soit environ 8% de l'effort international), centre de données (un des deux centres mondiaux de traitement se trouve au centre Coriolis de Brest), validation des mesures, recherche (circulation océanique, climat, biogéochimie) et océanographie opérationnelle (via Mercator Océan et Copernicus/MyOcean). La contribution française à Argo est coordonnée au sein de la structure inter-organismes Coriolis (composante in situ de l'océanographie opérationnelle) coordonnée par l'Ifremer et qui regroupe les principaux organismes français impliqués dans l'océanographie (CNES,CNRS/INSU, Ifremer, IPEV, IRD, Météo-France, SHOM). La France coordonne également la contribution européenne (Euro-Argo) et accueille depuis 2014 la structure légale européenne Euro-Argo ERIC [4] sur le site de l'Ifremer à Brest.

2.1.5 Les nouveaux défis pour Argo

La première priorité et le premier défi pour Argo sont de maintenir le réseau actuel sur le long terme. Le second défi pour Argo est de faire évoluer le réseau pour répondre à de nouvelles questions scientifiques et élargir son domaine d'applications. L'observation des mers marginales est à la portée de la technologie actuelle, celle des zones polaires, complexe du fait de l'absence de possibilité de transmission directe en zone couverte de glace et des conditions climatiques extrêmes fait l'objet de travaux expérimentaux. Des évolutions majeures sont, en revanche, nécessaires pour l'observation de l'océan sous 2000 mètres et pour l'étude de la biogéochimie et des écosystèmes marins. Les mesures profondes sont nécessaires pour le suivi du climat et la prévision décennale et à plus long terme. Elles sont aussi importantes pour mieux comprendre les mécanismes de la montée du niveau moyen des mers. L'évolution d'Argo vers la biogéochimie est particulièrement prometteuse (*Claustre et al.* [2010]).

4. European Research Infrastructure Consortium

2.2 Mercator Ocean, Centre francais d'analyse et de prévision océanique

2.2.1 Histoire et projets européens

Mercator Océan est un opérateur français de services en information océanique temps réel et différé : c'est le centre français d'analyse et de prévision océanique. Les systèmes numériques et les modèles développés par Mercator Océan sont capables de décrire l'état physique de l'océan à tout instant, en surface comme en profondeur, à l'échelle du globe ou d'une région du globe : température, salinité, hauteur de mer, épaisseur des glaces, des courants. Mercator Océan délivre ses produits océanographiques (analyses et prévisions, bulletins et cartes) et fournit son expertise partout dans le monde.

Les années 2000 ont été marqués par le projet européen d'océanographie opérationnelle, « MyOcean » qui visait la mise en place (définition, conception, développement et validation) d'une capacité européenne intégrée pour la surveillance, l'analyse et la prédiction des océans, en s'appuyant sur l'ensemble des compétences et moyens existants au niveau européen. Le projet « MyOcean » s'est appliqué à uniformiser les capacités propres en océanographie de chaque pays européen. Les challenges au sein de « MyOcean » étaient :

- d'éviter les duplications de moyens
- de rendre tous les sous-systèmes européens interopérables
- de définir et de faire appliquer les mêmes procédures (standards, normes de qualité) de développement et de qualification opérationnelle à chaque système
- de définir une architecture et une organisation qui mettent toutes ces capacités (compétences et moyens) en ordre de marche de façon intégrée
- de prouver que cette organisation est fiable, robuste, durable et répond aux exigences de qualité de service imposées par l'Europe.

Le 31 Mars 2012, le projet « MyOcean » céda la place à « MyOcean2 ». MyOcean2 a duré 30 mois (Avril 2012/septembre 2014). Il a été prolongé par le projet « MyOcean Follow on » jusqu'à fin avril 2015. « MyOcean2 » s'est employé en particulier à travailler davantage à la réponse aux besoins des utilisateurs qui viennent de multiples horizons : agences européennes et nationales, services opérationnels, conventions intergouvernementales, instituts de recherche et laboratoires, entreprises liées au secteur maritime (sécurité, énergie, environnement, pêche...). « MyOcean2 » et « MyOcean Follow on » ont rassemblé 59 partenaires sur 28 pays, c'est à dire près de 350 experts au service d'un service unique au monde.

Fort de la capacité démontrée en mode pilote depuis 2009 via les projets européens de recherche « MyOcean 1 et 2 » et « MyOcean Follow on », Mercator

Océan a été chargé par la Commission Européenne fin 2014 de la mise en place du Service Copernicus de surveillance des océans [5]. Le service européen de surveillance des océans entre maintenant dans sa phase opérationnelle. Le programme Copernicus anciennement le GMES (Global Monitoring for Environement and Security) a pour ambition de donner à l'Union européenne une capacité autonome et opérationnelle en matière d'infrastructures d'observation spatiale et in situ.

2.2.2 Modèle océanique et systèmes d'assimilation

Modeles océaniques

Les systèmes mis en oeuvre à Mercator Océan reposent sur un modèle mathématique d'océan décrit par les équations primitives (NEMO-OPA), avec ou sans assimilation de données, modélisant l'état actuel, futur ou passé de la physique ou la biogéochimie océanique à une résolution horizontale et verticale donnée. Un même système génère différents types de produits en fonction de la configuration choisie. La configuration utilisée pour la simulation étudiée dans cette thèse est basée sur la grille au 1/4° tripolaire ORCA025. La résolution horizontale est de 27 km à l'Equateur et de 6 km vers les pôles de la grille. 75 niveaux verticaux sont utilisés pour discrétiser l'océan ; la discrétisation est de 1m à la surface de la mer jusqu'à 450m dans les couches du fond, avec 22 niveaux sur les 100 premiers mètres. La version 3.1 du code NEMO (*Madec* [2008]) est utilisée. Ce dernier est composé du modèle océanique OPA (*Madec et al.* [1998]) couplé au modèle de glace de mer LIM2 EVP (*Fichefet and Maqueda* [1997]). Tous deux sont forcés par des flux atmosphériques de résolution temporelle de $3h$ issus du Centre Européen de Prévision à Moyen Terme (CEPMMT). Les flux océan-atmosphère sont les conditions limites à la surface océanique des équations primitives que le modèle intègre dans le temps. Ces flux sont calculés à partir de variables atmosphériques à l'aide des formules aérodynamiques globales (Bulk CORE) décrites dans *Large and Yeager* [2009].

Résolution	Horizontale :1/4°, Verticale :50 niveaux
Modèle	ORCA025, NEMO LIM2 EVP, Bulk CORE
Assimilation	SAM, IAU, correction biais 3D-Var
Observations assimilées	SSALTO/DUACS (SLA), SST, profils T/S

Table 2.1 – *Caractéristiques du système d'assimilation de données au sein du système $PSY3V3R3$ de Mercator-Océan*

5. Programme européen de surveillance mondiale pour l'environnement et la sécurité.

Méthodes d'assimilation

L'assimilation de données est le procédé mathématique qui consiste à corriger, à l'aide d'observations (satellitaires et in-situ), l'état de l'océan d'une prévision océanique, afin de produire une analyse et des prévisions de l'état océanique. Le schéma d'assimilation de données SAM2 (Système d'assimilation de données Mercator Océan version 2, voir le tableau 2.1) par méthode séquentielle d'ordre réduit est utilisé dans la plupart des systèmes opérationnels Mercator (*Lellouche et al.* [2013]). Ce schéma d'assimilation permet de combiner de façon optimale des observations (données altimétriques SSALTO/DUACS, données in-situ de température de surface SST, anomalie de hauteurs d'eau SLA, et les profils T/S issus du centre Coriolis) et des prévisions de modèles océaniques pour produire une analyse de l'océan. On estime que cette analyse représente l'état de l'océan le plus probable compte tenu des observations disponibles et de la prévision. La méthode SAM2 repose sur certaines étapes du filtre de Kalman d'ordre réduit basé sur la formulation SEEK (Singular Extended Evolutive Kalman), initialement développé par les équipes MEOM[6] du LGGE[7] (*Pham et al.* [1998]).

Formellement les étapes du filtre de Kalman étendu peuvent s'écrire :

Calcul de la matrice de gain $\mathbf{K}$ au temps t_i :

$$\mathbf{K}_i = \mathbf{P}_i^f \mathbf{H}_i^t (\mathbf{H}_i \mathbf{P}_i^f \mathbf{H}_i^t + \mathbf{R}_i)^{-1} \tag{2.1}$$

Analyse au temps t_i à l'aide de l'opérateur d'observation non-linéaire :

$$\mathbf{x}_i^a = \mathbf{x}_i^f + \mathbf{K}_i \left(\mathbf{y}_i^o - \mathbf{H}_i \mathbf{x}_i^f \right) \tag{2.2}$$

Calcul de la matrice de covariance d'erreur d'analyse au temps t_i :

$$\mathbf{P}_i^a = (\mathbf{I} - \mathbf{K}_i \mathbf{H}_i) \mathbf{P}_i^f \tag{2.3}$$

Prévision au temps t_{i+1} par propagation de l'analyse de t_i à t_{i+1} par le modèle non-linéaire d'évolution :

$$\mathbf{x}_{i+1}^f = \mathbf{M}_{i \to i+1}(\mathbf{x}_i^a) \tag{2.4}$$

Calcul de la matrice de covariance d'erreur de prévision au temps t_{i+1} :

$$\mathbf{P}_{i+1}^f = \mathbf{P}_i^a + \mathbf{Q}_i \tag{2.5}$$

Où $\mathbf{Q}_i$ est la matrice de covariance d'erreurs du modèle. La particularité du filtre de Kalman étendu utilisé dans les systèmes opérationnels Mercator est que la matrice de covariance d'erreur de prévision n'est pas propagé par le modèle d'évolution $\mathbf{M}$.

6. Modélisation des écoulements océaniques multi-échelles
7. Laboratoire de Glaciologie et Géophysique de l'Environnement

Dans l'ensemble des systèmes de prévision utilisés à Mercator Océan, la matrice de covariance d'erreur de prévision est basée sur un ensemble d'anomalies d'états de l'océan 3D. Ces anomalies sont obtenues à partir d'une longue simulation numérique avec ou sans assimilation de l'ordre d'une dizaine d'années, à laquelle on retire une moyenne glissante sur une période de temps fixée en utilisant un filtre de Hanning (plus de détails dans *Lellouche et al.* [2013]). Lors d'un cycle d'assimilation centré sur un jour N d'une année donnée, l'ensemble des anomalies d'états de l'océan est selectionné pour chaque année de la simulation numérique sur une fenêtre de 120 jours centrée en N. Cet ensemble forme alors la matrice de covariance d'erreurs de prévision. Dans le but aussi de réduire le coût numérique, la matrice de covariance d'erreurs de prévision est donc décomposée selon une base de vecteurs d'anomalies fixe et variant saisonnièrement (de type EOFs 3D avec 200 à 300 membres). Par conséquent la solution qui minimise l'écart quadratique entre les observations et leur équivalent modèle est calculée dans une espace réduit, sous forme d'une combinaison linéaire de modes d'erreurs du modèle. La correction calculée, appelée incrément, est appliquée progressivement sur la fenêtre d'assimilation (7 jours) pour limiter le choc d'initialisation au pas de temps où le modèle corrigé est redémarré (Incremental Analysis Update, *Lellouche et al.* [2013]). Le système de prévision basé sur la configuration ORCA025-LIM2 nommé $PSY3V3R3$, est décrit dans *Lellouche et al.* [2013]. Les systèmes d'analyse ont une grande dépendance au type et au nombre d'observations qui sont assimilés, en particulier les données Argo et les données satellites altimétriques. Le système $PSY3V3R3$ opérationnel depuis 2005, assimile systématiquement les données altimétriques, les champs de température de surface de la mer et les profils verticaux de température et salinité issus des différents instruments tels que les profileurs Argo, les bathythermographes (XBT), les bouées (TAO, TRITON, PIRATA, les bathysondes (CTD) ou plus récemment les profils de température et de salinité obtenus à partir de mammifères marins. L'ensemble de ces données sont collectées, vérifiées puis distribuées par Coriolis.

Chapitre 3

Caractérisation des écarts modèle observations pour une simulation sans assimilation

Sommaire

3.1 Méthodes statistiques 20
3.2 Caractéristiques propres du modèle 24
3.3 Présentation des observations in-situ Argo 26
3.4 Evaluation des écarts entre la simulation et les observations in-situ Argo 29
3.4.1 Séries temporelles des écarts moyens 29
3.4.2 Moyennes et variances en température et salinité sur la période 2006-2009 31
3.4.3 Moyennes saisonnières en température et salinité sur la période 2006-2009 37

Les modèles d'océan, libres (i.e sans assimilation de données) et forcés par des états atmosphériques réalistes sont utilisés pour comprendre la dynamique océanique et les processus conduisant à la circulation moyenne et à sa variabilité. Les simulations numériques et les observations océaniques sont complémentaires. Leur utilisation conjointe est très utile pour les investigations de la variabilité océanique observée ainsi que pour les études de reproductibilité de la variabilité des indices climatiques (*Juza* [2011]). Enfin, les simulations, telles que la simulation globale utilisée dans cette thèse, peuvent être utilisées pour guider les déploiements des futurs réseaux d'observations. Dans ce chapitre, nous allons souligner l'intérêt de l'utilisation des données Argo pour valider une simulation numérique. L'analyse présentée permet de caractériser les erreurs modèles afin d'estimer dans quelle mesure la simulation numérique diffère de la réalité observée par Argo. Cette étude passe par un ensemble d'étapes de description et de traitements statistiques décrits dans la section 3.1, puis par la présentation du modèle dans la section 3.2 d'une part et celle des observations dans la section 3.3 d'autre part. Enfin l'analyse des différentes statistiques les plus significatives entre la simulation et le jeu d'observation est faite en section 3.4.

3.1 Méthodes statistiques

Colocalisation 4D du modèle et des observations Pour comparer les données hydrographiques Argo et la simulation numérique réaliste utilisée dans cette thèse, il est nécessaire de calculer la valeur du modèle au point d'observation. Les sorties modèles utilisées sont des champs de température et de salinité T/S(x,y,z,t) journaliers. Ainsi les champs de température et de salinité T/S(x,y,z,t), simulés sont colocalisés dans l'espace des observations. Cette interpolation est faite de façon quadri-linéaire, à partir des points les plus proches de l'espace spatio-temporel (x, y, z, t) du modèle vers l'espace des profils observés. Le but étant de constituer une base de profils hydrographiques simulés *colocalisée*, parsemée dans le temps et l'espace de manière identique aux observations. Ce travail repose sur un code écrit en fortran 90, déjà existant au sein de Mercator Océan (métriques développées par Charly Regnier). Enfin nous utilisons dans cette thèse des données in-situ climatologiques issues du World Ocean Atlas 2009 couvrant la période (1955-2006), documentées dans *Locarnini et al.* [2010] et *Antonov et al.* [2010] et notées Levitus par la suite. L'étape de colocalisation a aussi été réalisée sur les champs climatologiques en température et salinité en prévision de leur utilisation dans les chapitres 4 et 4.4.

Traitement statistique en temps et en espace

Il s'agit à cette étape de moyenner temporellement et spatialement (horizontalement et verticalement) les différences modèle-observation. Les différences modèle-observation sont donc cumulées :

- sur une semaine (pour être ensuite cumulées par saison ou sur toute la durée d'observation)
- par boîtes d'une résolution de 4°x4°
- et échelonnées sur les 26 premiers niveaux verticaux identiques à ceux de la climatologie Levitus 2009

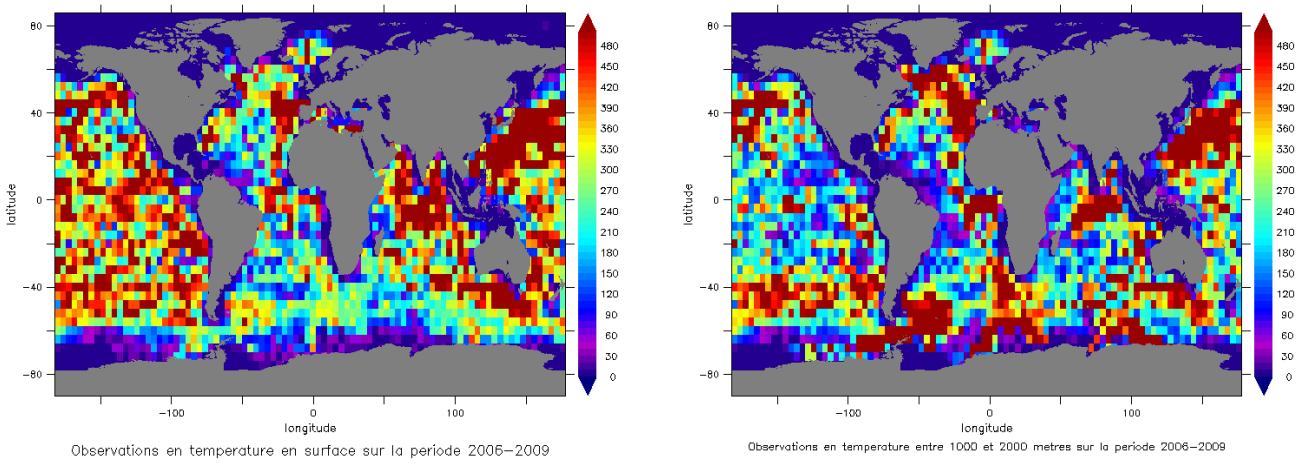
Les profils Argo peuvent s'échelonner de la surface jusqu'en profondeur tous les mètres, voire tous les 5 mètres. La stucture verticale des masses d'eaux est décrite via une échelle non homogène le long de la colonne d'eau. Aussi chaque point vertical est rangé en fonction du niveau standard de la climatologie Levitus auquel il appartient (c'est-à-dire par couche homogène suffisamment fine en surface et à plus grande épaisseur en profondeur). La résolution verticale des boîtes choisie correspond aux 26 premiers niveaux standards de la climatologie Levitus, définis de 0 à 2000 mètres (la profondeur de plongée maximale des flotteurs Argo) :

0	10	20	30	50	75	100	125	150	200	250	300
400	500	600	700	800	900	1000	1100	1200	1300	1400	1500
1750	2000										

La résolution spatiale horizontale choisie est de 4°x4°. L'analyse de la répartition des observations en température et salinité (figures 3.1.a et 3.2.a) permet de garder un regard critique sur les estimations statistiques en fonction des régions observées et de s'assurer de leur robustesse. Le nombre d'observations par boîtes en surface sur la période 2006-2009 est en moyenne supérieure à 500 dans l'océan Pacifique et compris entre 100 et 250 dans l'océan Atlantique. Notons que les années antérieures à 2009 révèlent une forte chute du nombre de profils disposant de données comprises entre 1500 et 2000 mètres (*Cabanes et al.* [2013]).

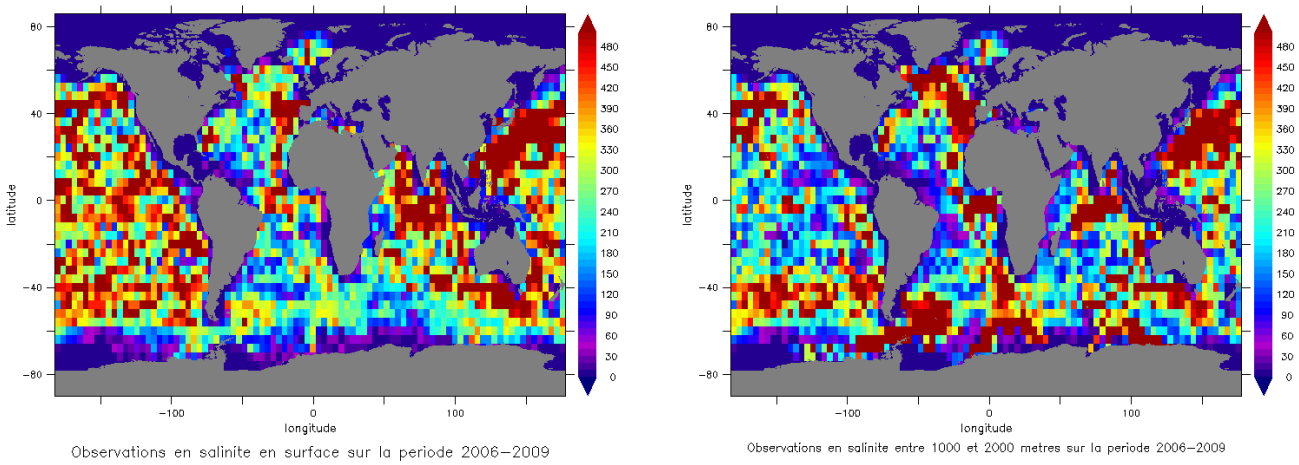
Les figures 3.1.b et 3.2.b représentent la couverture spatiale en température et salinité entre 1000 et 2000 mètres. Dans les autres bassins océaniques le nombre d'observations disponibles dans les couches en profondeur est compris entre 150 et 250, voire supérieure à 500 pour les hautes latitudes. Par conséquent la densité d'observations de la surface à 2000 mètres à cette résolution spatiale et cumulée sur la période $2006 - 2009$ pourra donner lieu à des résultats statistiques robustes. Puis l'analyse des écarts [modèle-observations] en température et salinité passe par le choix des indicateurs statistiques les plus représentatifs. On utilisera les écarts moyens [modèle-observations] en moyenne annuelle et saisonnière sur la période 2006-2009 (figures 3.8 à 3.14 et celles comprises dans l'annexe A). Enfin on s'affranchit de la dimension horizontale pour observer les séries temporelles des écarts moyens [modèle-observations] sur la période 2006-2009 (figures 3.6 et 3.7).

Certaines régions sont connues pour être difficilement reproduites par la simulation numérique (figures 3.8). Ce comportement sera aussi décrit et comparé entre plusieurs bassins océaniques comportant ces biais de simulation.



(a) (b)

Figure 3.1 – *Nombre total d'observations en température sur la période* 2006 – 2009 *dans des boîtes de* 4°x4° *: en surface (a) , entre 1000 et 2000 mètres (b).*



(a) (b)

Figure 3.2 – *Nombre total d'observations en salinité sur la période* $2006 - 2009$ *dans des boîtes de* $4°\text{x}4°$ *: en surface (a) , entre 1000 et 2000 mètres (b).*

3.2 Caractéristiques propres du modèle

Configurations, résolutions et forcage atmosphérique

La simulation $T335$ réalisée par Mercator Ocean sur les machines du CEP-MMT, Centre Européen pour les Prévisions Météorologiques à Moyen terme [1], repose sur la configuration océanique et de glace de mer ORCA025-LIM2EVP, basée sur le système NEMO [2] (version 3.1, *Madec* [2008]). La simulation globale couvre l'océan sur la période Novembre 1989-Décembre 2009. Nous travaillons ici sur la période $2006-2009$. La résolution verticale s'échelonne sur 75 niveaux, celle de la grille horizontale est au $1/4°$. La résolution spatiale du modèle océanique correspond à une maille de 28 km à l'équateur devenant plus fine avec la latitude jusqu'à être voisine de 6 km dans l'océan Arctique et le long du continent Antarctique. Cette simulation libre, sans assimilation de données est pilotée par le forçage atmosphérique ERA/interim. L'initialisation des variables de température et de salinité est produite via la climatologie levitus 1998.

Kallberg [2011] montre un biais systématique dans les prévisions à court terme de certaines variables du modèle généré dans ERA/interim. Il compare notamment les prévisions du flux de précipitation provenant de la réanalyse ERA/interim avec les données issues du GPCPV2.1 (Global Precipitations Climatology Project, *Huffman et al.* [2009]). De manière générale la réanalyse ERA/interim révèle des précipitations plus élevées que celles estimées à partir des données GPCPV2.1 sur l'ensemble des océans tropicaux (*Dee et al.* [2011]). Un déséquilibre est aussi présent dans le flux radiatif net de chaleur à la surface, biais issus principalement de la représentation des nuages, notamment sur les zones "d'upwelling" côtières. Une méthode de correction a donc été appliquée lors de l'utilisation des flux radiatifs descendants de chaleur et de précipitations issus de la réanalyse ERA/interim afin de forcer le modèle. Cette méthode de correction utilise les données satellites GPCPV2.1 et Gewex SRB (Surface Radiation Budget, *Stackhouse et al.* [2011]) pour corriger uniquement la composante grande échelle du forçage sans impacter la variabilité synoptique (cyclone) et la variabilité interannuelle du signal. Cette méthode a permis de corriger partiellement le biais d'eau douce positif présent sur les 20 ans de la simulation $T335$. En effet la figure 3.3 représente les conséquences de la méthode de correction sur la salinité de surface. On peut observer que la salinité de surface (courbe orange) est supérieure à celle estimée par une expérience sans correction des précipitations (courbe bleue). Cependant la salinité de surface obtenue après correction est aussi supérieure à celle estimée à partir de la

1. en anglais ECMWF, European Centre for Medium-Range Weather Forecasts
2. Nucleus for European Modelling of the Ocean, www.lodyc.jussieu.fr/NEMO

climatologie Levitus (courbe en pointillés rouges). La méthode de correction a permis cependant une meilleure estimation de la salinité de surface notamment dans le gyre subtropical Atlantique (non montré). Les principales paramétrisations de la simulation numérique $T335$ sont présentées dans le tableau (3.1).

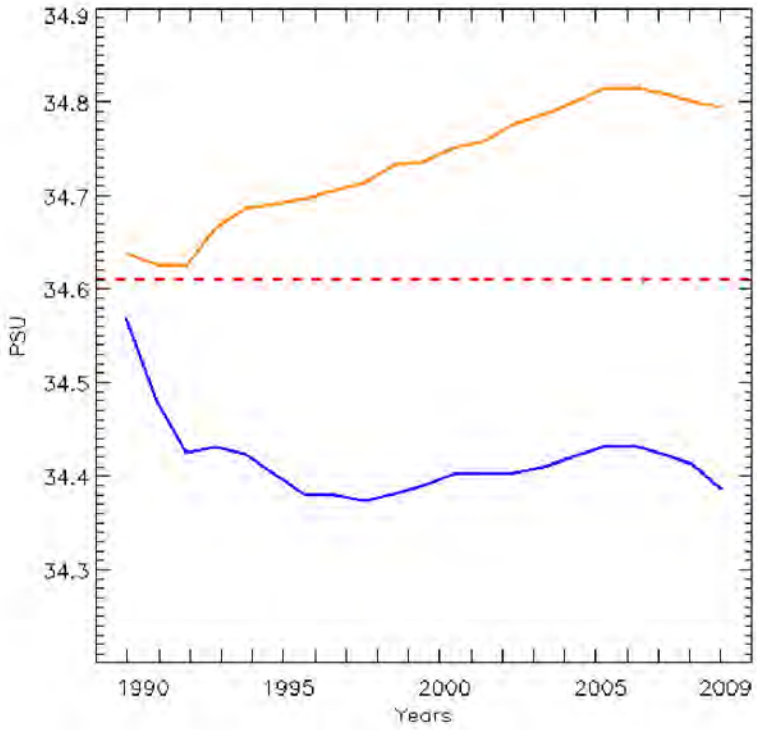


Figure 3.3 – *Application de la méthode de corrections : série temporelle de la salinité de surface avec correction des flux de précipitations (orange) et sans (bleu), en comparaison à celle issue de la climatologie Levitus (pointillés rouges). Figure réalisée par G. Garric.*

Configuration	
Résolution	Horizontale :1/4°, Verticale :75 niveaux
Modèle	ORCA025, NEMO version 3.1 LIM2 EVP, Bulk CORE
Bathymétrie	Etopo1, Gebco1
Paramétrisations	Surface libre filtrée, Mélange vertical TKE
Schéma d'advection	Diffusion isopycnale pour les traceurs, biharmonique pour le moment
Forcage atmosphérique	
	ERA/Interim 3h, cycle diurne analytique appliqué au flux de chaleur solaire journalier, correction grande échelle des flux descendants de chaleur et de précipitations

Table 3.1 – *Caractéristiques de la simulation numérique T*335 *réaliste réalisée à Mercator-Océan*

3.3 Présentation des observations in-situ Argo

Nous utilisons des observations hydrographiques *in situ*, qui permettent d'étendre la connaissance de l'océan dans sa profondeur, à la différence par exemple des données altimétriques de la hauteur de la mer. La base de données CORA fournit des profils instantanés de température et salinité, T/S(x,y,z,t) issus notamment des profileurs Argo (*Cabanes et al.* [2013]). Cette base de données CORA correspond à une extraction des profils de température et salinité in situ de la base de donnée de Coriolis en un temps différé, sur une longue période passée. Toutes les données sont alors re-validées. Ici, ces dernières sont exclusivement issues, après contrôle de qualité, des profils Argo (décrit au chapitre 2). Le fonctionnement des flotteurs et leur cycle de plongée jusqu'à 2000 mètres de profondeur a été détaillé dans le chapitre 2, section 2.1.2. La figure 3.4 représente la moyenne annuelle du nombre de flotteurs Argo disponibles sur la période $[2006 - 2009]$ par boîtes 4°x4° et souligne l'hétérogénéité spatiale du réseau Argo. Aussi on observe une plus grande densité dans le Pacifique Nord-Ouest, le Nord de l'Océan Indien et une bonne couverture moyenne dans tout l'Océan Pacifique. La figure 3.4 servira de référence dans le chapitre 4. Enfin afin d'illustrer l'inhomogénéité et le caractère sporadique du réseau Argo dans certaines régions, notamment les régions côtières, la figure 3.5 présente à titre d'exemple pour l'année 2009 un snapshot des positions occupées par les profileurs. Chaque point représente une position prise par l'un des flotteurs profileurs à la remontée en surface (un profileur tous les 10 jours est visible).

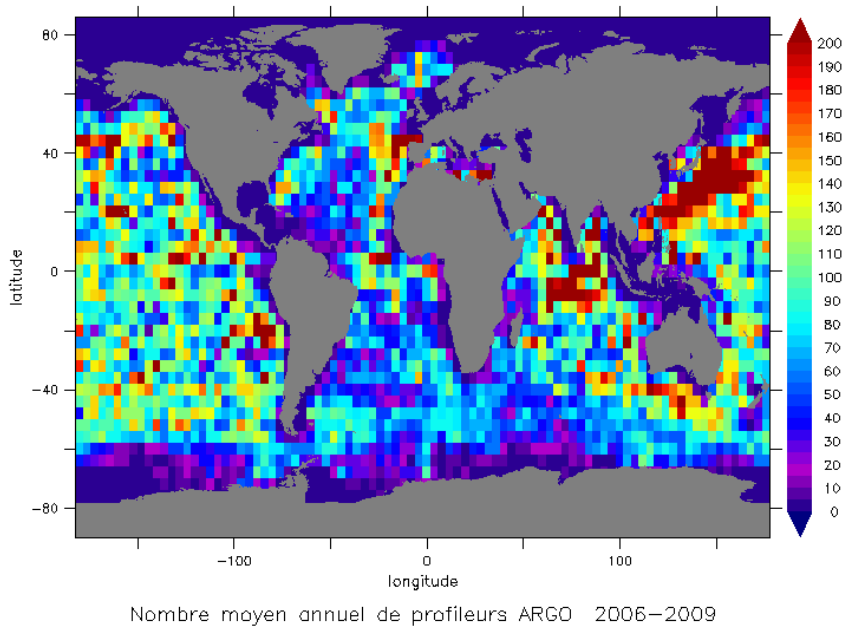
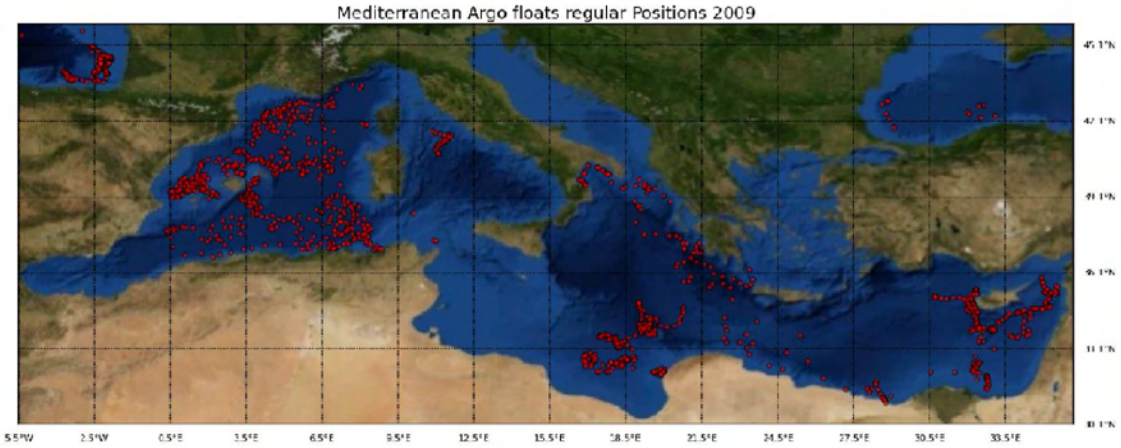
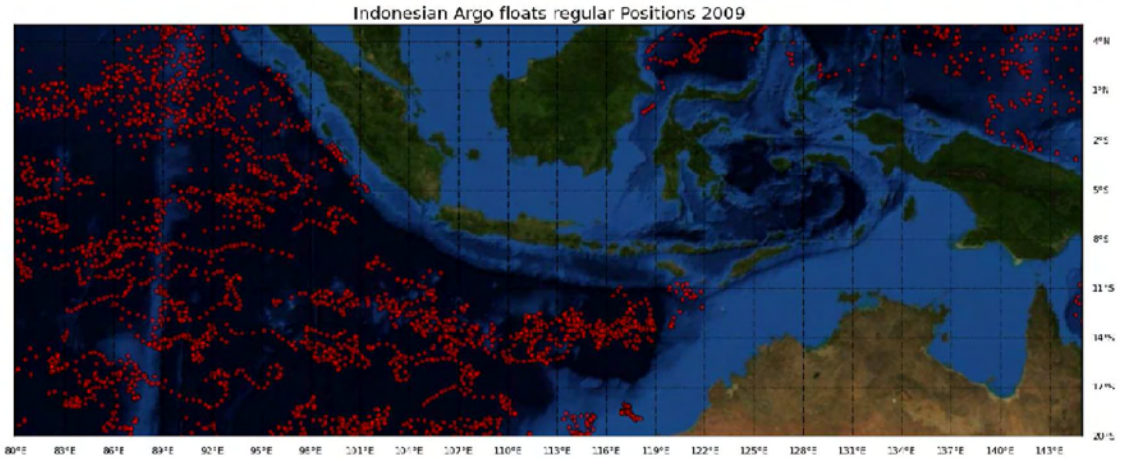


Figure 3.4 – *Nombre moyen annuel de flotteurs Argo disponibles dans des boîtes de 4°x4° sur la période 2006-2009 .*



(a) Mer Méditerranée



(b) Mers Indonésiennes

Figure 3.5 – *Echantillonnage pour plusieurs grands bassins des profileurs Argo, Année 2009.*

3.4 Evaluation des écarts entre la simulation et les observations in-situ Argo

Il s'agit ici de décrire le comportement de la simulation numérique $T335$ par rapport à celui des observations Argo en surface et en profondeur, d'une part d'un point de vue temporel sur la période étudiée 2006-2009, d'autre part d'un point du vue spatial sur l'ensemble des océans (sections 3.4.2 et 3.4.3). L'ensemble des erreurs modèles est étudié en regard des bulletins de validation Quo Vadis réalisés à Mercator Océan [3].

3.4.1 Séries temporelles des écarts moyens

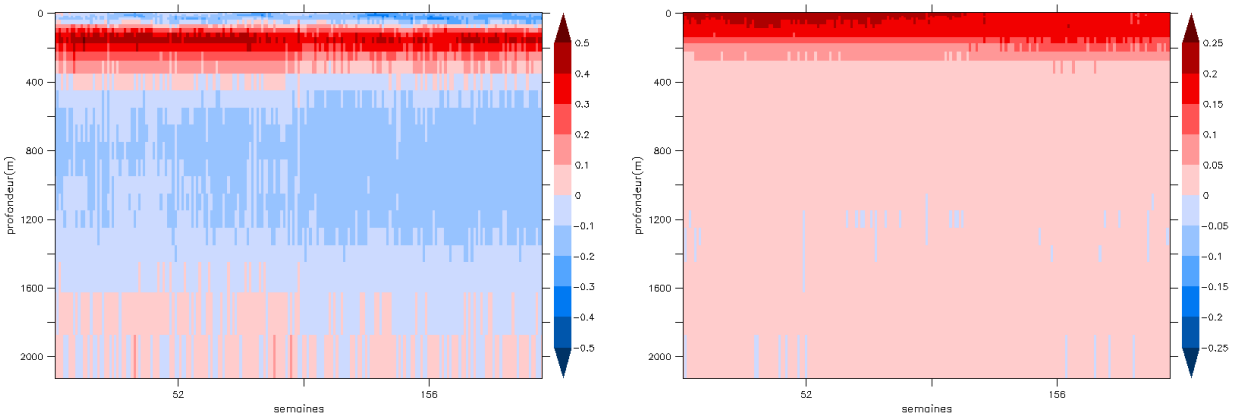
Les écarts moyens [modèle-observations] en température et en salinité moyennés sur l'ensemble des bins 4°x4° (détails en section 3.1) sont représentés sur les figures 3.6 et 3.7.

Bilan temporel en température : Les figures 3.6a et 3.7a permettent d'observer plus précisément la structure stratifiée des écarts modèle observations dans la couche de mélange et en dessous au fil des 4 ans. En moyenne globale la simulation T335 n'est pas assez stratifiée, en d'autres termes le gradient de température en fonction de la profondeur est trop faible. En effet, la température de surface issue de la simulation est inférieure à celle observée jusqu'à 80 mètres (bleu sur la figure 3.7a). Puis le modèle est trop chaud sur la couche [100 à 200] mètres, et révèle une erreur de positionnement de la profondeur de la thermocline. A partir des 400 mètres de profondeur le modèle se refroidit. Dans cette simulation le mélange semble trop important dans les couches de surface, la chaleur est donc injectée dans la thermocline, mais ne gagne pas les couches plus profondes. Cependant il faut garder à l'esprit que le modèle a un comportement différent selon les régions, et que certains biais se compensent en moyenne globale. On peut aussi remarquer un biais froid du modèle en surface pour l'ensemble des inter-saisons 2006-2009 (alternance bleu foncé et clair sur la figure 3.7a). La profondeur de la couche de mélange est plus fine en été qu'en hiver, à l'inter-saison le modèle ne parvient pas de manière précise à représenter la réelle variation saisonnière observée.

Bilan temporel en salinité : De même les figures 3.6b et 3.7b soulignent les variations des erreurs modèles en salinité sur la période étudiée. Sur les 200 premiers mètres, la salinité de surface issue de la simulation est surestimée (rouge sur les figures 3.6b et 3.7b). Le bilan évaporation-précipitation semble être en déficit

3. http ://www.mercator-ocean.fr/science/qualification

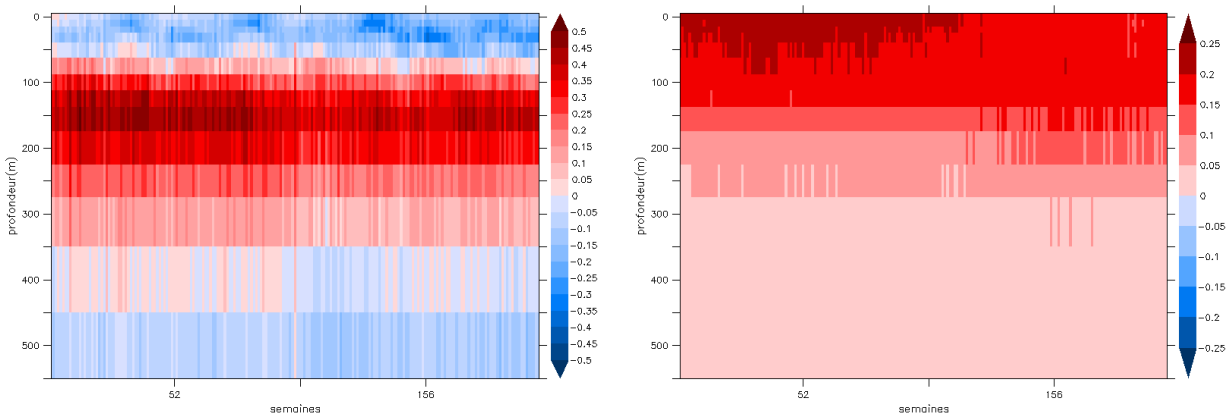
d'eau douce (erreur sur les forcages E-P). L'évaporation serait plus importante que les apports des pluies et des fleuves, ce qui concentre la teneur en sel. Pour des profondeurs supérieures à 400 mètres l'équilibre paraît globalement préservé, la simulation est proche en moyenne globale de la salinité observée (écarts modèle-observations de l'ordre de 0.05 Psu). Le biais salé dans les 200 premiers mètres contribue à déstratifier la simulation.



(a) Série temporelle des écarts en température

(b) Série temporelle des écarts en salinité

Figure 3.6 – *Série temporelle des écarts moyens globaux modèle-observations sur la classe de profondeur [0-2000]m en température (en °C) et en salinité (en psu) sur la période 2006-2009.*



(a) Série temporelle des écarts en température

(b) Série temporelle des écarts en salinité

Figure 3.7 – *Série temporelle des écarts moyens globaux modèle-observations sur la classe de profondeur [0-600]m en température (en °C) et salinité (en psu) sur la période 2006-2009.*

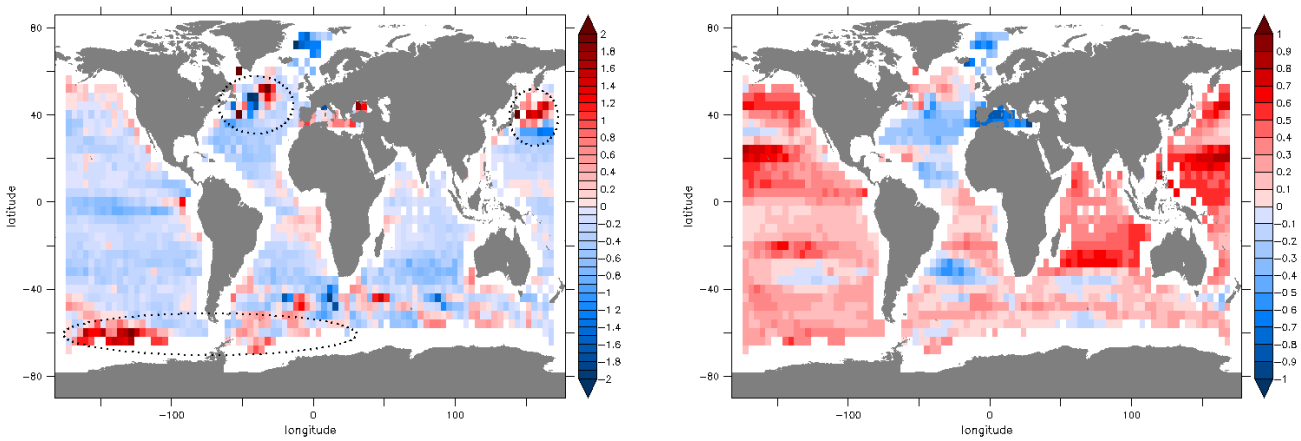
3.4.2 Moyennes et variances en température et salinité sur la période 2006-2009

Evaluation des erreurs modèles moyennes en surface

Bilan spatial en température Les écarts moyens [modèle-observations] en température de surface représentés sur la figure 3.8a révèlent un dipôle chaud/froid au niveau des courants de bord Ouest (Gulf stream, Kuroshio) dans le Nord Ouest de l'océan Pacifique, le Nord de l'océan Atlantique. Puis on observe un biais chaud le long du courant circumpolaire et enfin des biais froids à l'est du Groënland et dans l'océan Pacifique équatorial. A l'exception de ces sites où circulent de forts courants, la simulation numérique sous-estime la température de surface sur le reste des océans. Les observations Argo montrent que la simulation reste encore en deçà de la température observée dans la couche de surface $0 - 10$m.

Bilan spatial en salinité A l'échelle globale la simulation tend à représenter la salinité de surface avec un biais salé et dessalé dans l'océan Nord Atlantique (figure 3.8b). Il réside un biais dessalé de l'ordre de -0.5 Psu dans l'océan Nord Atlantique (latitudes [20°N-40°N]). Par ailleurs les cartes climatologiques (WOA 2009, non montrées) montrent que l'Océan Atlantique est plus salé en surface que les autres bassins océaniques (écart de l'ordre de 2 Psu dans la gyre subtropicale).

Aussi la présence d'un biais dessalé dans l'océan Nord Atlantique montre que le modèle NEMO représente moins bien les contrastes de salinité de surface.



(a) Ecarts en température sur la section 0-10 m

(b) Ecarts en salinité sur la section 0-10 m

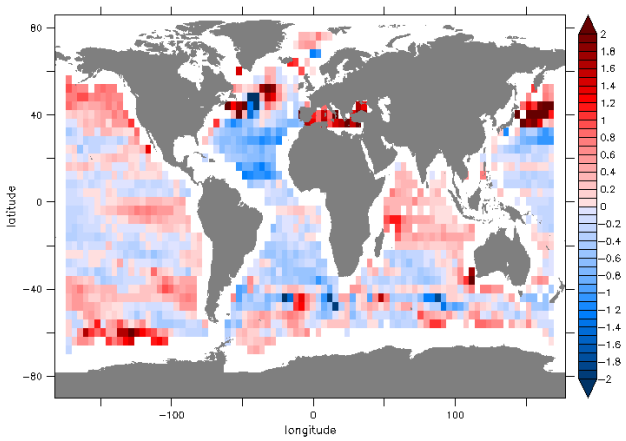
Figure 3.8 – *Ecarts moyens en température (en °C) et salinité (en psu) de surface sur la période* 2006 – 2009

Evaluation des erreurs modèles moyennes et de leur variabilité en profondeur

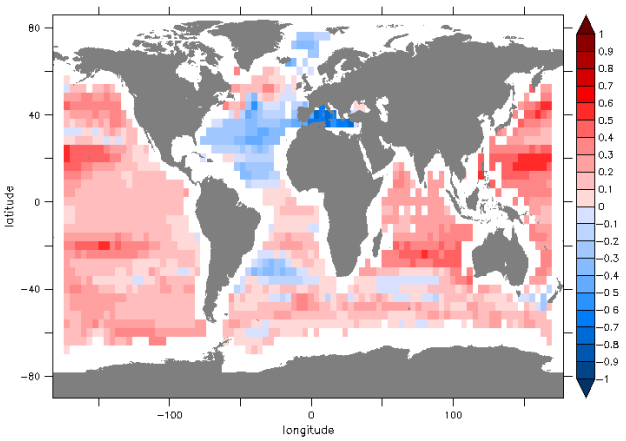
Dans cette section, je vais cibler l'analyse sur certaines régions où les biais de simulation en température et salinité sont forts (voir figures 3.8) et ceci jusqu'en profondeur. En effet les profileurs Argo permettent d'observer la variabilité océanique en subsurface, nous attachons alors plus de détails aux couches de profondeur comprises entre $dp1 = [0 - 300]m$, $dp2 = [300 - 700]m$ et enfin $dp3 = [700 - 2000]m$. Les figures 3.9 et 3.10 représentent les erreurs modèles en température et salinité sur ces différentes couches de profondeur. Enfin la figure 3.11 correspond aux cartes des variances des erreurs modèles citées ci-dessus (erreur moyenne quadratique exprimée à partir des écarts [modèle-observations]).

Nord Pacifique A l'ouest de l'Océan Nord Pacifique (région du Kuroshio), on observe au sein de la couche de surface $dp1$, des biais thermiques forts (supérieurs à 2°C, voir la figure 3.9a). Les erreurs modèles en température s'atténuent progressivement avec la profondeur (de l'ordre de 1.5°C à 0.5°C sur la couche $dp2$ et

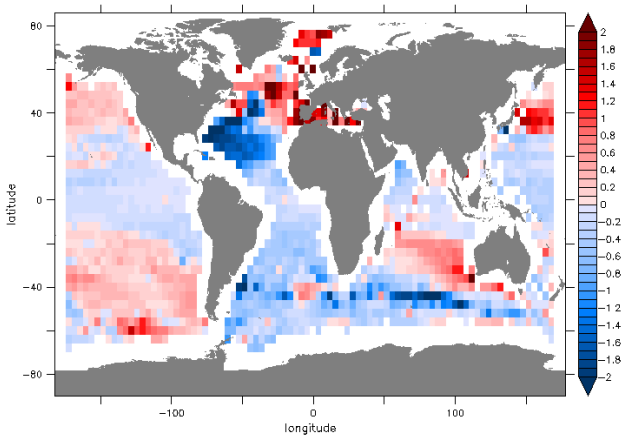
proches de 0 sur la couche $dp3$ sur les figures 3.9c et 3.10a). En parallèle, les biais halins sont élevés (de l'ordre de 0.4 Psu) sur la couche $dp1$ puis diminuent le long de la colonne d'eau (de l'ordre de 0.1 Psu, voir les figures 3.9b, 3.9d et 3.10b).



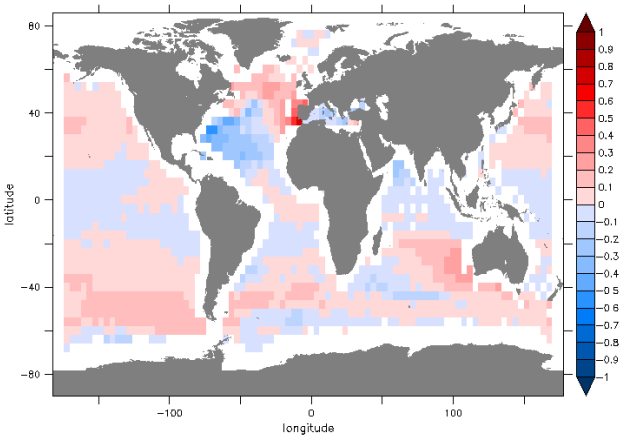
(a) Ecarts en température sur la couche $dp1 = 0 - 300$ m



(b) Ecarts en salinité sur la couche $dp1 = 0 - 300$ m

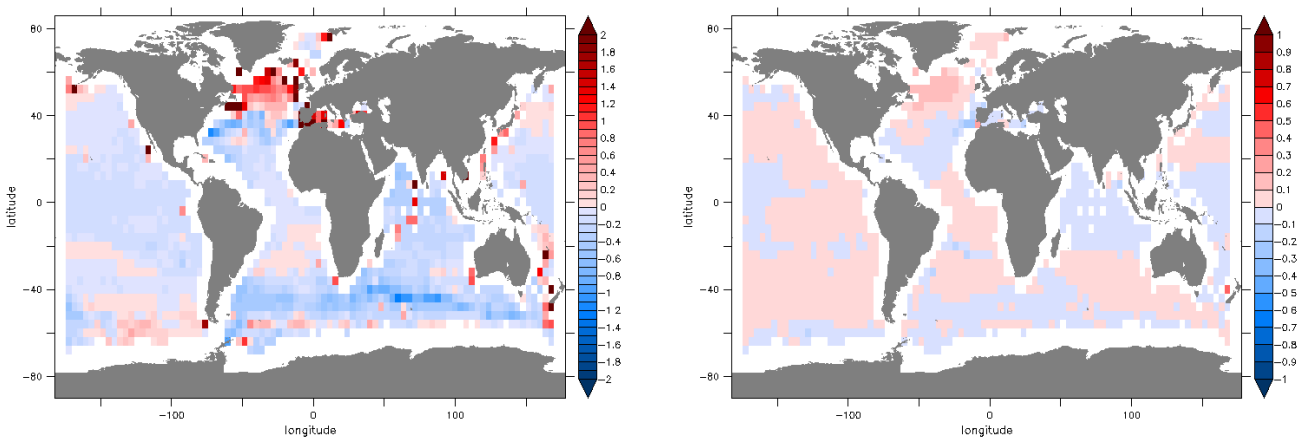


(c) Ecarts en température sur la couche $dp2 = 300 - 700$ m



(d) Ecarts en salinité sur la couche $dp2 = 300 - 700$ m

Figure 3.9 – *Erreurs modèle moyennes en température (en °C) et en salinité (psu) sur la période 2006-2009, sur les couches de profondeur* $dp1 = 0 - 300m$ *et* $dp2 = 300 - 700$ m



(a) Ecarts en température sur la couche 700-2000 m

(b) Ecarts en salinité sur la couche 700-2000 m

Figure 3.10 – *Erreurs modèle moyennes en température (en °C) et en salinité (psu) sur la période 2006-2009, sur la couche de profondeur* $dp3 = 700 - 2000$ *m*

Nord Atlantique Le Nord-Ouest de l'Océan Atlantique présente une alternance de biais thermiques positifs et négatifs sur les couches $dp1$, $dp2$ et $dp3$. Ces dipôles positifs-négatifs traduisent un possible mauvais positionnement des courants présents (Gulf Stream, courant nord atlantique, courant du labrador). Plus au sud (aux latitudes [0°-25°N]) les biais thermiques sont négatifs de l'ordre de -0.5°C sur la couche $dp1$ puis augmentent fortement sur la couche $dp2$ (jusqu'à -2°, voir la figure 3.9a) et enfin ces biais s'estompent en profondeur. Le bassin nord atlantique comporte des biais halins positifs (de l'ordre de 0.2 à 0.6 Psu) sur la bande de latitudes [40°N-60°N] et négatifs plus au sud aux latitudes [20°N-40°N] et ce, sur les couches $dp1$ et $dp2$. Puis ces biais halins sur la couche $dp3$ oscillent autour de zero. A proximité du bassin méditerranéen on observe aussi un biais salé au niveau du détroit de Gilbraltar sur la couche $dp2$. Le flux d'eau continental influe sur les salinités et les températures de ce bassin. Une hypothèse possible serait que la simulation ne représente pas correctement le flux d'eau Atlantique entrant en surface, et la profondeur du flux d'eau salée sortant (*Beuvier* [2011]). De plus, à cette résolution il est difficile de conclure sur le comportement de la simulation au sein du bassin méditerranéen.

Courant Circumpolaire Antarctique Le long du courant circumpolaire Antarctique les biais thermiques sont positifs sur la couche $dp1$, dans le bassin Pacifique et

alternent entre positif et négatif dans les bassins atlantique et indien (figure 3.9a). Au sein de la couche $dp2$ les biais thermiques négatifs s'intensifient dans le bassin indien et diminuent dans le bassin pacifique (figure 3.9a). Plus en profondeur les biais thermiques et halins s'atténuent. Enfin seule la couche $dp1$ présente des biais halins positifs de l'ordre de 0.3 à 0.4 Psu le long du courant circumpolaire antarctique.

Les variances des erreurs modèles en température et salinité sont élevées et sont associées à des biais thermiques et halins (figure 3.11) dans la plupart des bassins et régions abordées ci-dessus. Notamment le long de la couche $dp1$ les courants de bords Ouest (Gulf Stream, Kuroshio), et le bassin méditerranéen présentent à la fois de fortes erreurs moyennes et variances d'erreurs en température et salinité. Enfin on note un biais chaud dans l'océan Pacifique Equatorial sur la couche $dp1$ (voir figure 3.9a) aussi associé à des variances d'erreurs modèles élevées (figure 3.11). Ces tendances s'atténuent sur la couche $dp2$. Ceci signe la présence d'une forte variabilité au sein de la circulation océanique, de la surface jusqu'à 700 mètres. Au sein de ces régions énergétiques et tourbillonnaires, la simulation numérique reposant sur un modèle au $1/4°$, reste éloignée des températures et salinités observées par Argo.

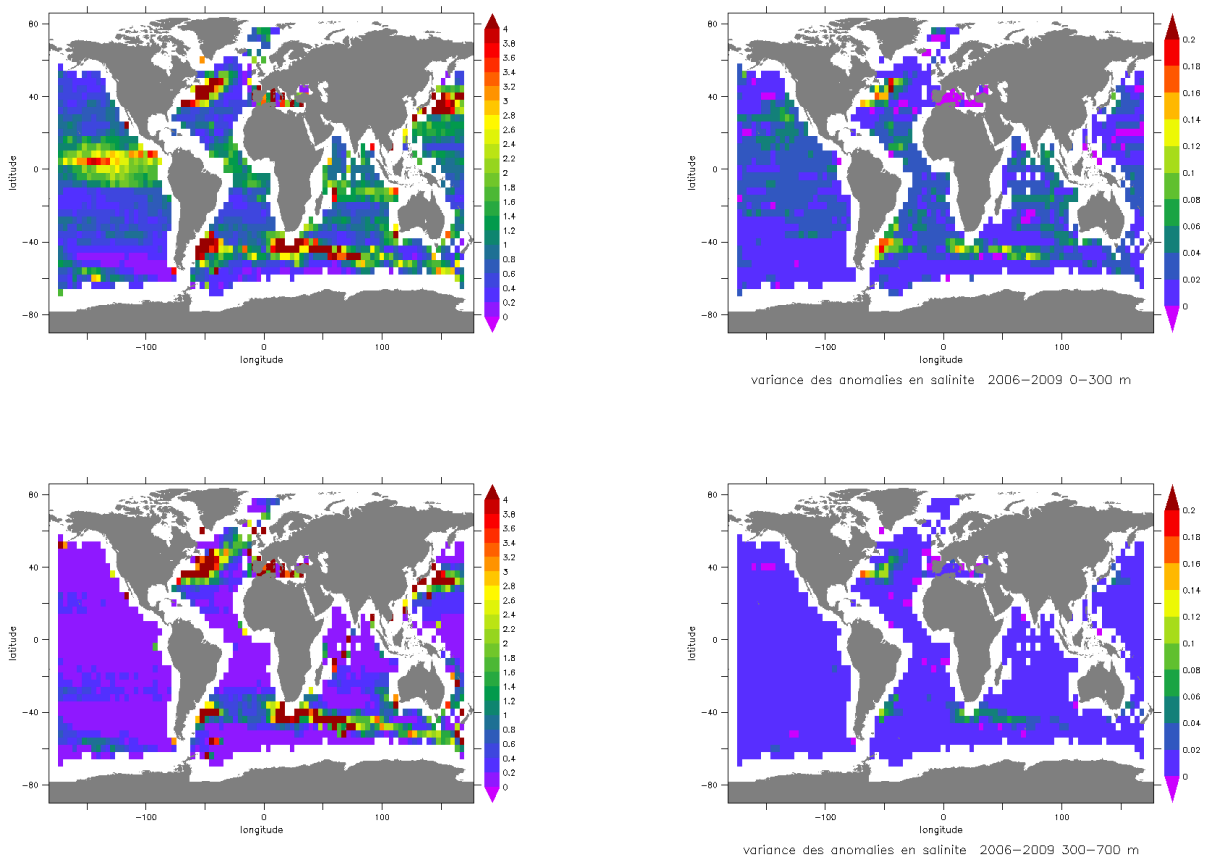
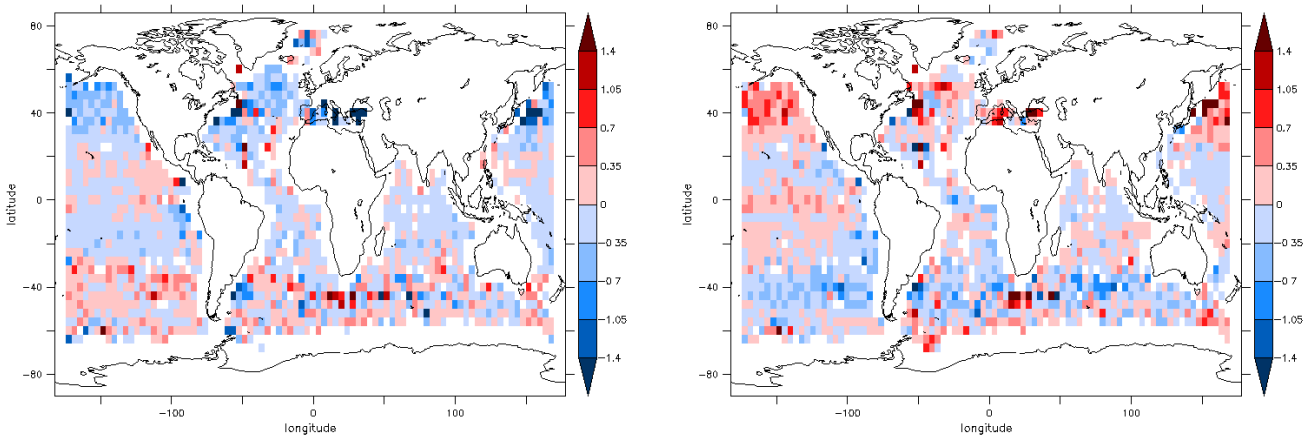


Figure 3.11 – *Variance des erreurs modèle en température (en* $°C^2$*) et salinité (en* psu^2*), pour les couches* $dp1 = [0 - 300]m$*,* $dp2 = [300 - 700]m$

3.4.3 Moyennes saisonnières en température et salinité sur la période 2006-2009

Erreurs modèle à l'échelle saisonnière

L'objectif est de confirmer la présence des biais déjà observés à la section 3.4.1, particulièrement ceux possédant un cycle saisonnier décrits à partir de la figure 3.7a. Pour cela les erreurs modèles en température et salinité sont représentées par saison (voir l'annexe A) et comme précédemment sur des couches de profondeur définies à la section 3.4.2. Dans cette section, à titre d'exemple les figures des erreurs modèles en hiver [4] et en été [5] boréal sont représentées (figures 3.12, 3.13 et 3.14).



(a) été sur la section 0-10 m

(b) hiver sur la section 0-10 m

Figure 3.12 – *Ecarts moyens en température (en °C) entre le modèle et les données Argo, en surface pour les étés et hivers en* 2006 – 2009

Evaluation du cycle saisonnier en surface Les erreurs modèle en température de surface en été (figure 3.12a) et en hiver (figure 3.12b) suivent un cycle saisonnier. Les régions du Kuroshio et du Gulf Stream présentent des biais thermiques négatifs en été et positifs en hiver. La simulation numérique ne parvient pas à représenter la variation saisonnière observée par le réseau Argo (sous-estimation du cycle saisonnier de la température de surface).

4. basé sur les mois de décembre à fevrier
5. basé sur les mois de juin à août

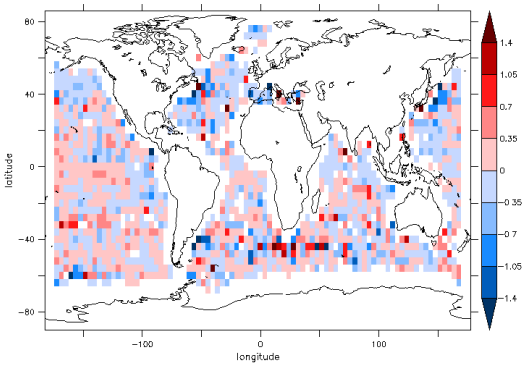
Evaluation du cycle saisonnier en profondeur Sur les figures 3.13, les biais thermiques dus au cycle saisonnier situés sur les régions du Kuroshio et du Gulf Stream sont encore présents sur la couche $dp1$ mais de manière moins marquée qu'en surface. La simulation numérique en ces zones énergétiques de l'hémisphère nord présente une tendance froide en été et chaude en hiver.

Sur les couches $dp2$ et $dp3$ les zones où sont situés les courants de bord Ouest (Gulf Stream, Kuroshio, courant des Aiguilles) montrent des dipôles chauds/froids ainsi que salés/dessalés cependant sans variation saisonnière. Ces dipôles salés/dessalés s'estompent avec la profondeur. Par contre en température ils sont plus nombreux et parsemés le long des côtes notamment sur la couche $dp3$. Au 1/4° la circulation côtière ne peut pas être correctement représentée. Pour finir, on n'observe pas de biais à grande échelle saisonnier en salinité le long de la colonne d'eau (voir les figures 3.14).

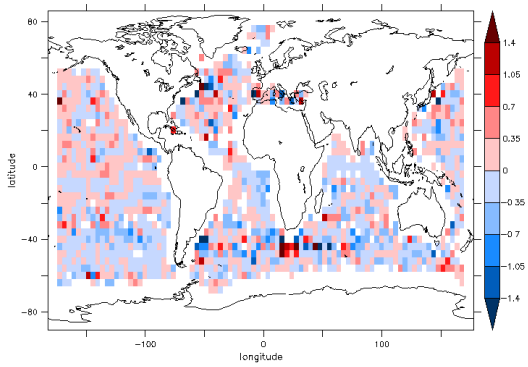
Conclusion

Dans ce chapitre nous avons étudié la capacité de la simulation numérique à reproduire la variabilité océanique de la surface jusqu'à 2000 mètres, ainsi que son réalisme face aux données Argo. La simulation numérique présente des erreurs modèle à grande échelle trop salées de la surface jusqu'à 300 mètres et s'atténuent en profondeur. Par ailleurs, l'océan Nord Atlantique reste encore trop peu salé malgrè les corrections en précipitation. On observe aussi une alternance des erreurs modèle froides/chaudes/froides le long de la colonne d'eau. De plus, certains biais thermiques et halins révélés en surface persistent jusqu'en profondeur (700 mètres) dans les régions de forts courants (Gulf Stream, Kuroshio, ACC, Aiguilles). Puis, nous avons pu mettre en évidence des biais saisonniers en température présents en surface seulement.

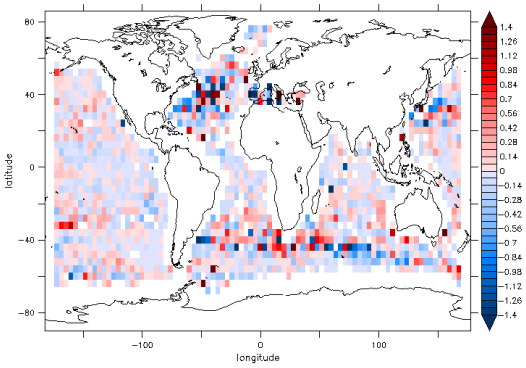
Ces erreurs modèle peuvent être en partie liées aux forcages, et à l'initialisation des variables de température et salinité via la climatologie Levitus 1998. La variabilité océanique s'échelonne selon un large spectre d'échelles spatio temporelles. L'étude, la modélisation et la prévision de la circulation océanique nécessitent de bien reproduire une large gamme d'échelles spatiales et temporelles. Outre les estimations des biais et écarts-types entre le modèle et les observations analysées dans ce chapitre, il est aussi très important de caractériser les échelles spatiales et temporelles associées. Dans les chapitres suivants nous allons donc nous focaliser sur l'estimation des échelles spatiales des erreurs modèle en température et salinité. Nous estimerons également les échelles spatiales de variabilité de la température et de la salinité (c'est-à-dire à partir des écarts climatologie-observations Argo).



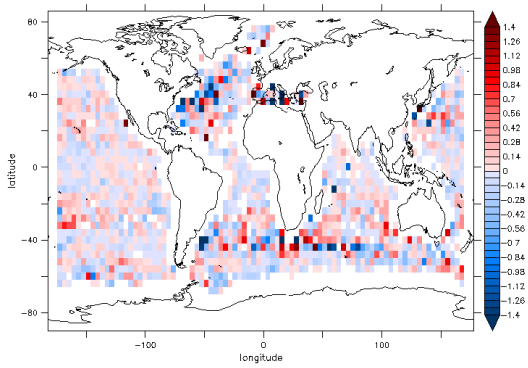
été, $dp1 = [0 - 300]m$



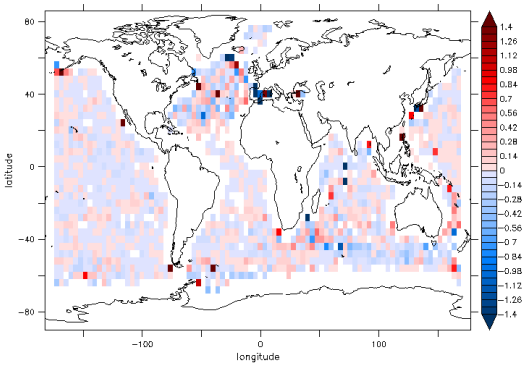
hiver, $dp1 = [0 - 300]m$



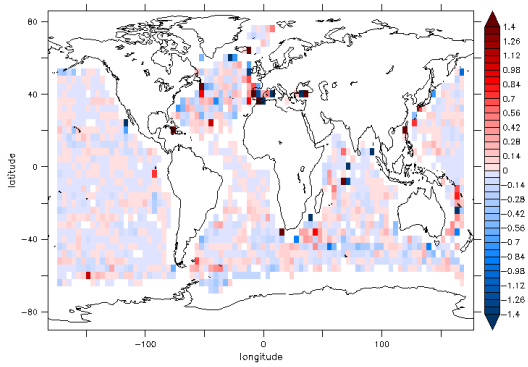
été, $dp2 = [300 - 700]m$



hiver, $dp2 = [300 - 700]m$

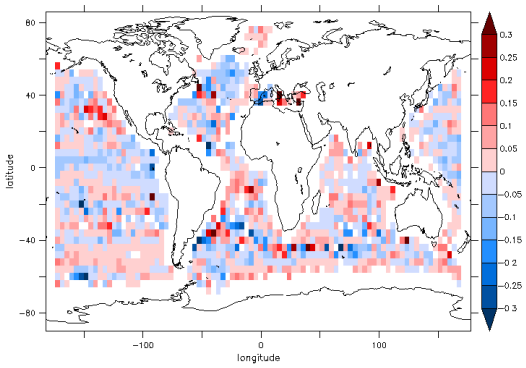


été, $dp3 = [700 - 2000]m$

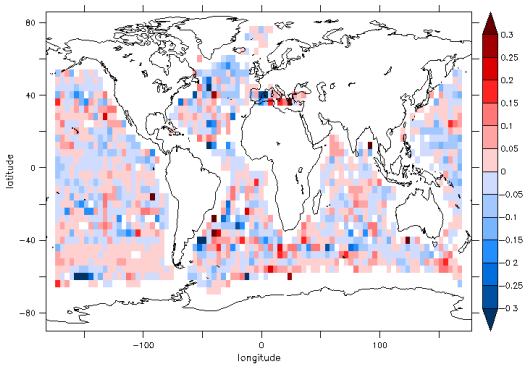


hiver, $dp3 = [700 - 2000]m$

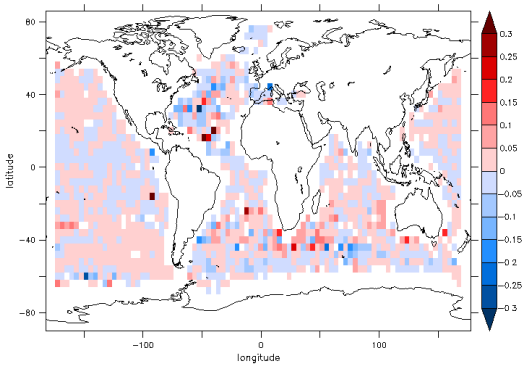
Figure 3.13 – *Ecarts moyens en température entre le modèle et les données Argo sur les couches* $dp1 = [0 - 300]m$, $dp2 = [300 - 700]m$ *et* $dp3 = [700 - 2000]m$ *pour les étés (gauche) et hivers (droite) 2006-2009*



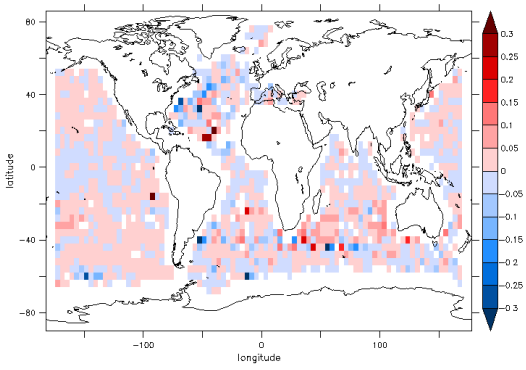
été, $dp1 = [0 - 300]m$



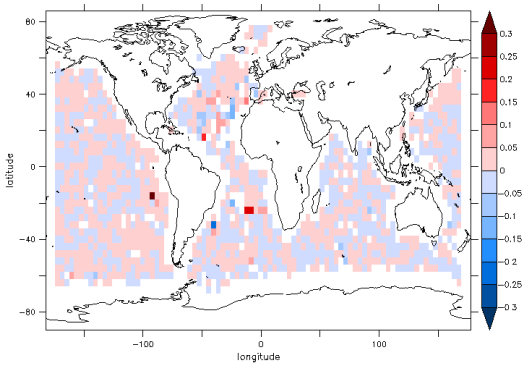
hiver, $dp1 = [0 - 300]m$



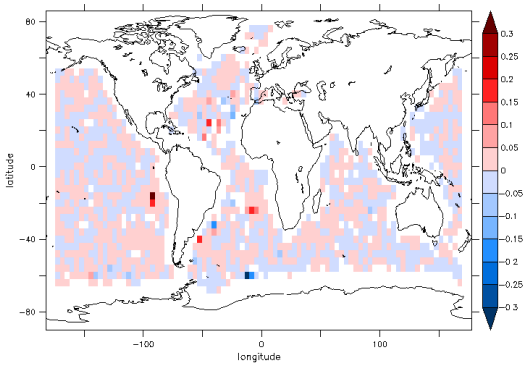
été, $dp2 = [300 - 700]m$



hiver, $dp2 = [300 - 700]m$



été, $dp3 = [700 - 2000]m$



hiver, $dp3 = [700 - 2000]m$

Figure 3.14 – *Ecarts moyens en salinité entre le modèle et les données Argo sur les couches* $dp1 = [0 - 300]m$, $dp2 = [300 - 700]m$ *et* $dp3 = [700 - 2000]m$ *pour les étés (gauche) et hivers (droite) 2006-2009*

Chapitre 4

Caractérisation des échelles spatiales des erreurs modèles et de la variabilité océanique pour la température et la salinité

Sommaire

4.1 Introduction 43

4.2 Méthodologie pour l'estimation des échelles spatiales . 44

4.2.1 Discrétisation des covariances 44

4.2.2 Ajustement d'une fonction analytique de corrélation . . 46

4.2.3 Validation de la méthodologie à partir de données simulées 47

4.2.4 Conclusion 55

4.3 Caractérisation des échelles spatiales des erreurs modèles et du signal à partir des données Argo 56

4.3.1 Structures et caractéristiques des régions examinées . . 56

4.3.2 Utilisation d'une nouvelle formulation analytique du modèle de corrélation 59

4.3.3 Analyse globale des échelles spatiales des erreurs modèles en température et salinité 59

4.3.4 Comparaison des échelles spatiales des erreurs modèles par rapport aux échelles de la variabilité océanique . . . 63

4.4 Estimation des échelles spatiales des variations de température et salinité à partir des données Argo . . . 71

4.4.1 Article paru dans le journal *Ocean Science* 71

4.4.2 Erreur formelle sur l'estimation de la covariance 87

4.5 Conclusion 89

4.1 Introduction

Ce chapitre constitue une étape préliminaire pour parvenir à affiner les exigences d'échantillonnage du réseau Argo, et définir un système d'observations permettant de représenter ou contraindre via l'assimilation de données les principales échelles de la variabilité océanique. L'échantillonnage actuel du réseau Argo (un flotteur tous les $3°x3°$) vise principalement à résoudre les variations à grande échelle des champs de température et de salinité. La combinaison des données Argo avec les données altimétriques et satellitaires de température de surface permet, par contre d'accéder aux variations à moyenne échelle que ce soit via des méthodes de cartographie de type analyse objective (*Guinehut et al.* [2012]) ou via l'assimilation dans les modèles océaniques (*Lellouche et al.* [2013], *Turpin et al.* [2015]). La qualité de la restitution des champs de température et de salinité de la grande échelle à la moyenne échelle reste dans tous les cas fortement dépendante de l'échantillonnage du réseau Argo.

Une meilleure caractérisation des échelles spatiales et temporelles de la température et de la salinité ainsi que leurs variations géographiques est nécessaire pour améliorer les méthodes de cartographie de type analyse objective et donne des indications utiles pour optimiser l'échantillonnage du réseau Argo. Une échelle de corrélation spatiale L (L définie comme le premier passage à zero de la fonction de corrélation) est ainsi une bonne estimation du pas d'échantillonnage minimal à rechercher pour un réseau Argo optimisé. Les échelles spatiales sont, en outre, à priori différentes selon les régions (notamment en latitude) et les contraintes d'échantillonnage ne sont donc pas les mêmes. Bien entendu il n'est pas concevable de définir un réseau Argo résolvant seul la mésoechelle et c'est via la combinaison avec l'altimétrie notamment que l'on peut essayer de résoudre ces échelles. Les contraintes sur l'échantillonnage du réseau Argo peuvent donc être relachées (par exemple un pas d'échantillonnage égal à 2L ou 3L) mais la connaissance de L reste essentielle pour définir un échantillonnage optimisé. Lorsqu'il s'agit de contraindre un modèle via l'assimilation des données Argo conjointement à d'autres données, l'objectif doit être de réduire les erreurs sur l'ébauche modèle, i.e les écarts entre les observations et la prévision du modèle. Ce sont alors les échelles de cette erreur sur l'ébauche modèle qui sont importantes pour améliorer les schémas d'assimilation et pour optimiser l'échantillonnage. En assimilation de données, les erreurs sur l'ébauche modèle sont souvent approximées à partir d'une caractérisation des erreurs d'un run libre du modèle. Notre objectif ici est donc de caractériser à partir des données Argo, les échelles spatiales des anomalies de température et de la salinité par rapport à une climatologie et par rapport à une simulation issue d'un run libre du modèle global au 1/4°. On pourra ainsi caractériser à la fois les échelles spatiales du signal et celles des erreurs modèles.

Dans ce chapitre nous allons détailler à la section 4.2, la méthode exploitée lors de l'estimation statistique des échelles spatiales, l'ajustement non linéaire d'un modèle analytique de corrélation ainsi que la validation de cette méthodologie à partir de données simulées. En section 4.3.3 nous abordons l'analyse globale des échelles spatiales de corrélation pour l'ensemble des variables, profondeurs et zones étudiées. Les échelles spatiales du signal sont comparées à celles des erreurs modèles en section 4.3.4. Enfin nous allons travailler à la fois sur des anomalies Argo de température et de salinité par rapport à une climatologie (Levitus 2009) et par rapport à un modèle (simulation numérique $T335$) : cela nous permettra d'estimer à la fois les échelles spatiales du signal [observations-climatologie] et des erreurs modèles et d'analyser leurs différences.

4.2 Méthodologie pour l'estimation des échelles spatiales

4.2.1 Discrétisation des covariances

L'estimation des échelles spatiales de corrélation des erreurs modèles ou du signal passe par celle du champ de covariance des erreurs en température et salinité. Pour l'ensemble des calculs réalisés dans ce chapitre, ces derniers sont extraits des fichiers des profils Argo (voir chapitre 2) eux même sélectionnés sur une fenêtre temporelle de 7 jours.

La première étape du traitement consiste à identifier pour chaque couple de profils la classe de distance à laquelle ils appartiennent. Toutes les paires de profils sont donc classées en fonction des distances zonale et méridionale qui les séparent. Pour chaque classe de distance $\triangle x$ et $\triangle y$ on obtient un ensemble N de paires de profils qui permettent d'estimer la covariance spatiale $\mathbf{COV}(\triangle x, \triangle y)$ selon :

$$\mathbf{COV}(\triangle x, \triangle y) = \mathbf{E}(zz') - \mathbf{E}(z)\mathbf{E}(z') \tag{4.1}$$

$$\mathbf{COV}(\triangle x, \triangle y) = \frac{1}{N}\sum_{i=1}^{N} z_i z'_i - \frac{1}{N}\sum_{i=1}^{N} z_i \frac{1}{N}\sum_{i=1}^{N} z'_i \tag{4.2}$$

$$\text{tel que} \quad z, z' \in \mathbf{C}_{(\triangle x, \triangle y)} \tag{4.3}$$

Où z, z' sont les anomalies de température ou de salinité à une profondeur donnée et $N = N_{\mathbf{C}_{(\triangle x, \triangle y)}}$ est le nombre de paires de profils appartenant à la classe de distance donnée.

$\mathbf{E}(zz')$ est la moyenne du produit des anomalies [1]. $\mathbf{E}(z)\mathbf{E}(z')$ est le produit moyen des anomalies au sein de la classe de distance. Cette quantité est difficilement connue, et nous avons donc choisi d'utiliser les champs moyens saisonniers par bins de 4°x4° calculés au sein du chapitre 3. On rappelle qu'ils sont issus d'un traitement de colocalisation au point d'observation, et ce dans le cas de la simulation $T335$ et des données climatologiques. Les champs moyens saisonniers sont ensuite associés à chaque profil z_i' en fonction de la localisation et de la date. La résolution spatiale et temporelle des champs moyens est plus lâche (4°) que celle des anomalies au point d'observation. La méconnaissance du champ moyen est un des facteurs responsables du bruit au sein de la matrice de covariance. Pour cette raison il est préférable d'estimer le variogramme plutôt que la fonction de covariance. L'utilisation du variogramme est donc nécessaire car l'estimation de la fonction de covariance reste biaisée du fait de la méconnaissance de l'espérance $\mathbf{E}(\bullet)$, et ce même si l'hypothèse de stationnarité d'ordre 2 est vérifiée [2] ([*Garet and Kurtzmann*, 2011], *Allard* [2012]). La définition (4.1) peut encore s'écrire comme suit

$$\begin{aligned} \mathbf{COV}(\triangle x, \triangle y) &= \mathbf{E}[(z - \mathbf{E}(z))(z' - \mathbf{E}(z'))] \\ \mathbf{COV}(\triangle x, \triangle y) &= \mathbf{E}[(z - m)(z' - m)] \end{aligned} \tag{4.4}$$

On note l'espérance mathématique $\mathbf{E}(z) = m$ de même pour $\mathbf{E}(z')$, qui ne sont pas connues. La remplacer par une estimation $\hat{m}$ entraîne un biais qu'il est difficile de corriger. Alors le variogramme tel que z, $z' \in \mathbf{C}_{(\triangle x, \triangle y)}$, est la fonction :

$$\gamma(\triangle x, \triangle y) = \mathbf{E}[(z - z')^2] \tag{4.5}$$

$$\gamma(\triangle x, \triangle y) = \frac{1}{N}\sum_{i=1}^{N}[z_i - z_i']^2 \quad tel \quad que \quad z, z' \in \mathbf{C}_{(\triangle x, \triangle y)} \tag{4.6}$$

On exprime la fonction de covariance à partir de la définition du variogramme comme suit :

$$\mathbf{COV}(\triangle x, \triangle y) = \mathbf{var}(\triangle x, \triangle y) - \frac{1}{2}\gamma(\triangle x, \triangle y, \triangle t) \tag{4.7}$$

Il suffit normaliser le champ de covariance à l'aide de la variance des profils autocorrélés (i.e par $\mathbf{COV}(0,0)$) pour obtenir le champ de corrélation. Aussi la matrice de covariance est normée par la variance définie elle aussi par classe de distance

1. **E** est l'espérance mathématique dans le cas de variables aléatoires discrètes pour un ensemble fini
2. Existence des moments d'ordre 2

$\mathbf{C}_{(\triangle x,\triangle y)}$. Ceci permet de normer le champ de covariance par une variance locale et d'obtenir une matrice de corrélation plus lisse. Finalement, le fonction de correlation discrétisée s'exprime comme suit,

$$\mathbf{COR}(\triangle x,\triangle y)=\begin{cases} 1-\frac{1}{2}\gamma(\triangle x,\triangle y)\mathbf{var}(\triangle x,\triangle y)^{-1/2} & si \quad (\triangle x,\triangle y)\neq(0,0) \\ 1+\mathfrak{e} & si \quad (\triangle x,\triangle y)=(0,0) \end{cases}$$

où $\mathfrak{e}$ le bruit du aux erreurs de mesures et d'observations et $\mathbf{var}(\triangle x,\triangle y)$ variance définie par classe de distance. Nous utilisons aussi comme solution analytique selon les cas plusieurs fonctions de corrélation analytiques. Les fonctions de corrélations permettent de mettre en avant deux types d'information indépendantes : d'une part l'amplitude de l'incertitude correspondant au maximum de l'information, et d'autre part l'extension spatiale caractérisée par la longueur de corrélation.

4.2.2 Ajustement d'une fonction analytique de corrélation

Dans la section 4.2.3 nous allons procéder à des tests de validation de l'algorithme statistique d'estimation des champs de corrélation et de l'ajustement des échelles spatiales dans un cas simplifié. Pour cela on utilise un ensemble de distributions normales multivariées afin de simuler les anomalies $z=[\bullet-observation]$. L'algorithme d'estimation de covariance à partir des erreurs simulées gaussiennes peut être généralisé par l'expression analytique de la fonction de covariance d'une loi normale :

$$\mathbf{f}(r)=\begin{cases} (1+\mathfrak{e})\exp(-r) & si \quad r\neq 0 \\ 1+\mathfrak{e} & si \quad r=0 \end{cases} \tag{4.8}$$

où $r=\sqrt{(\frac{dx}{L_x})^2+(\frac{dy}{L_y})^2}$. On dispose d'un certains nombres de mesures sous la forme (r_i,z_i) , où $1\leq i\leq N$, N comme précédemment défini. Ces mesures dépendent de la taille de la zone observée et de la résolution de la classe distance utilisées. La minimisation au sens des moindres carrés consiste à estimer les paramètres $\mathbf{P}=(L_x,L_y,e)$ tels que la fonction $\mathbf{f}$ décrive au mieux les résultats de mesure, z_i. L_x,L_y sont les échelles spatiales de corrélations et e le niveau de bruit [3] présent au sein des mesures.

3. parfois défini aussi comme l'effet pépite en géostatistique.

On cherche donc $\mathbf{P}$ de sorte que la somme des carrés des déviations :

$$\mathbf{S}(\mathbf{P}) \quad = \quad \sum_{i=1}^{N} [z_i - f(r_i|\mathbf{P})]^2 \tag{4.9}$$

soit minimale.

Pour obtenir une solution numérique au problème de minimisation non linéaire nous utilisons l'algorithme de Levenberg-Marquardt (*Levenberg* [1944]; *Gavin* [2013]). La procédure de l'algorithme est itérative. On choisit un vecteur de paramètres initiaux, que l'on suppose assez proche d'un minimun qui constituera le vecteur $\mathbf{P}_0 = (100km, 100km, 0.1)$ de départ. Si on a effectué plus d'un certain nombre d'itérations, ou bien que l'on s'est approché suffisamment d'un minimum, la procédure se termine et renvoie le paramètre $\mathbf{P}$ comme estimation de la solution. Nous estimons aussi les barres d'erreurs (1σ) attribuées à chaque paramètre. Elles permettent d'avoir un avis critique sur l'estimation du vecteur paramètre $\mathbf{P}$.

4.2.3 Validation de la méthodologie à partir de données simulées

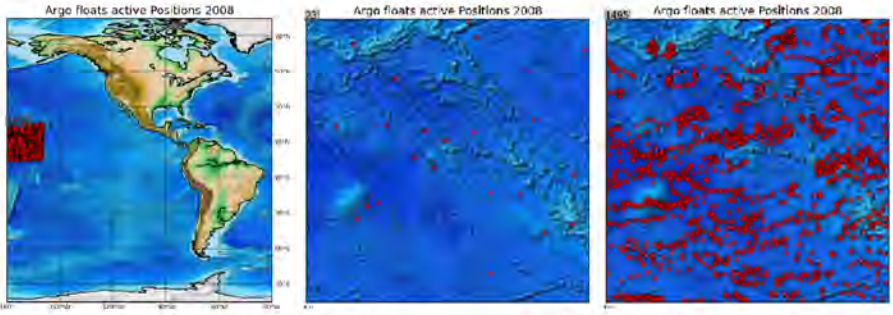
Mettre en pratique l'estimation statistique des échelles de corrélation spatiales a entrainé la réalisation de plusieurs tests afin de valider la méthodologie utilisée. Pour cela nous allons comparer l'estimation statistique de covariances spatiales à partir du réseau Argo des années 2006-2009, noté $Argo_{0609}$ et celle à partir d'un réseau d'observations à géométrie régulière noté $Argo_{ideal}$. Afin qu'il n'y ait pas d'erreurs d'interprétation et d'analyse lors de cette comparaison les données simulées sont issues de distributions normales multi-variées. Nous pourrons ainsi souligner l'influence de l'échantillonnage du réseau d'observations Argo sur ce type d'estimation statistique. Nous avons choisi une zone géographique arbitraire située dans l'Océan Pacifique Equatorial loin des côtes de longueur de côté de 10°à 20°. Les dimensions de la zone test doivent permettre la capture de suffisamment d'observations afin d'obtenir des statistiques robustes. Par la suite nous utilisons deux champs d'erreurs gaussiens reflétant de petites échelles spatiales et des échelles spatiales moyennes. La description des géométries des deux réseaux d'observations utilisés ainsi que celle de la modélisation des champs des erreurs gaussiennes et leurs caractéristiques statistiques sont faites aux sections A et B. Enfin l'exploitation des résultats pour l'ensemble des cas traités est faite en sections C et D.

A- Géométrie des réseaux utilisés

On notera par la suite $\mathbf{R_{zone10}}$ la zone de calcul de 10°x10° et $\mathbf{R_{zone20}}$ celle de 20°x20° .

Réseau $Argo_{0609}$

On dispose ici des coordonnées réelles des profileurs dériveurs. Chaque fichier journalier est exploité afin de sélectionner les profils appartenant à la zone géographique définie (exemple sur la figure 4.1a). Différents snapshots de la $\mathbf{R_{zone20}}$, hebdomadaire et annuel en 2008 sont représentés sur les figures 4.1 b) et c) permettant à titre indicatif d'observer l'évolution temporelle de la répartition spatiale. Connaissant la résolution nominale du réseau Argo (3°x3°, cycle de 10 jours), le nombre moyen de profileurs par semaine dans la zone $\mathbf{R_{zone20}}$ est de 31. La moyenne hebdomadaire au cours de la période étudiée est du même ordre de grandeur que le nombre moyen de profileurs par semaine pour la $\mathbf{R_{zone20}}$ (non montré). Les apports annuels en observations pour la zone test sur la période 2006-2009 sont similaires entre les années 2007 à 2009 (en moyenne 1450 observations), seule l'année 2006 est inférieure (1005 observations). La géométrie hebdomadaire observée sur la figure 4.1b) est irrégulière et la densité numérique observée (ici de 23) confirme le nombre moyen de profileurs attendu.



(a) Zone test année 2008 (b) Zoom semaine 2008 (c) Zoom année 2008

Figure 4.1 – *Echantillonnage sur la zone test de calcul des profileurs Argo, année 2008.*

Réseau $Argo_{ideal}$

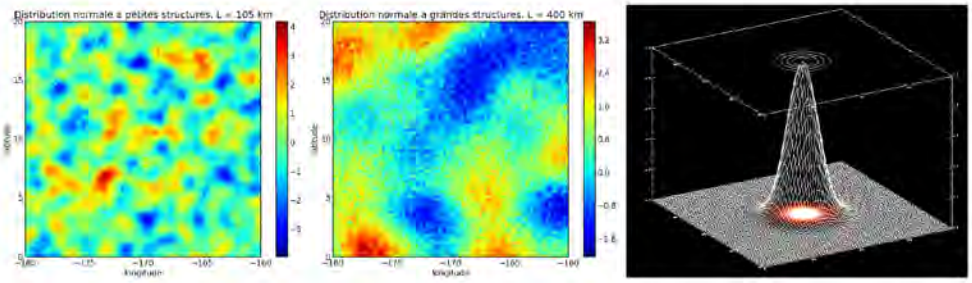
La construction d'un réseau d'observations idéal au 1/2° permet d'augmenter le nombre de profils appartenant à la zone $\mathbf{R_{zone}}$ par rapport à la réalité. Il est important de connaître la densité numérique moyenne par semaine car les statistiques sont réalisés sur cette période temporelle, puis cumulés. Pour la zone $\mathbf{R_{zone20}}$, 1680 profils sont plaçés régulièrement et représentent les profils disponibles pour une semaine soit environ 50 fois plus que pour le réseau réel ; également pour la zone $\mathbf{R_{zone10}}$, on dispose de 440 profils hebdomadaires, soit environ 10 fois plus que le réseau $Argo_{0609}$.

B- Simulation des champs d'erreurs gaussiens

On modélise un champ 3D issu d'une distribution normale multivariée sur une grille régulière horizontale au 1/4° située au point d'origine $(x_0, y_0) = (-180, 0)$. Ce champ répond au besoin d'une modélisation des écarts |modèle-observation| pour chaque latitude, longitude de la zone choisie. Soit z un vecteur 2D tel que $\mathbf{z} \sim \mathcal{N}(0, \mathbf{B})$. où B la matrice de covariance du signal, symétrique semi-définie positive de type

$$B(i,j) = \begin{cases} \exp(\frac{-d(i,j)^2}{L^2}) & si \quad i \neq j. \\ 1 + \mathfrak{e} & si \quad i = j \end{cases}$$

où **d** est la distance entre deux points de grille i et j en km, **L** représente la longueur caractéristique de la fonction de corrélation et $\mathfrak{e}$ le bruit du aux erreurs de mesures et d'observations. Par hypothèse, la fonction de corrélation gaussienne est **homogène**, autrement dit elle est possède la même forme en tout point du domaine. Le champ est **isotrope** : l'information apportée par la fonction de correlation au point (i,j) sera répartie de manière identique de part et d'autre du point. On se place dans deux cas, l'un représentant de petites structures spatiales (exemple sur la figure 4.2a), le second, de grandes structures spatiales (exemple sur la figure 4.2b) : pour cela la longueur de corrélation sera respectivement $L_p = 105km$ et $L_g = 400$km. Enfin l'ensemble des champs modélisés sera noté $\mathcal{Z} = \{ z^t, \mathrm{t} \in [0, N_{heb}]\}$, où N_{heb} est le nombre de semaines nécessaires pour couvrir la fenêtre temporelle voulue. Chaque membre k de cet ensemble possède un générateur aléatoire différent et indépendant n'entrainant aucune corrélation temporelle.



(a) z^t, L_p (b) z^t, L_g (c) Matrice de covariance de z^t

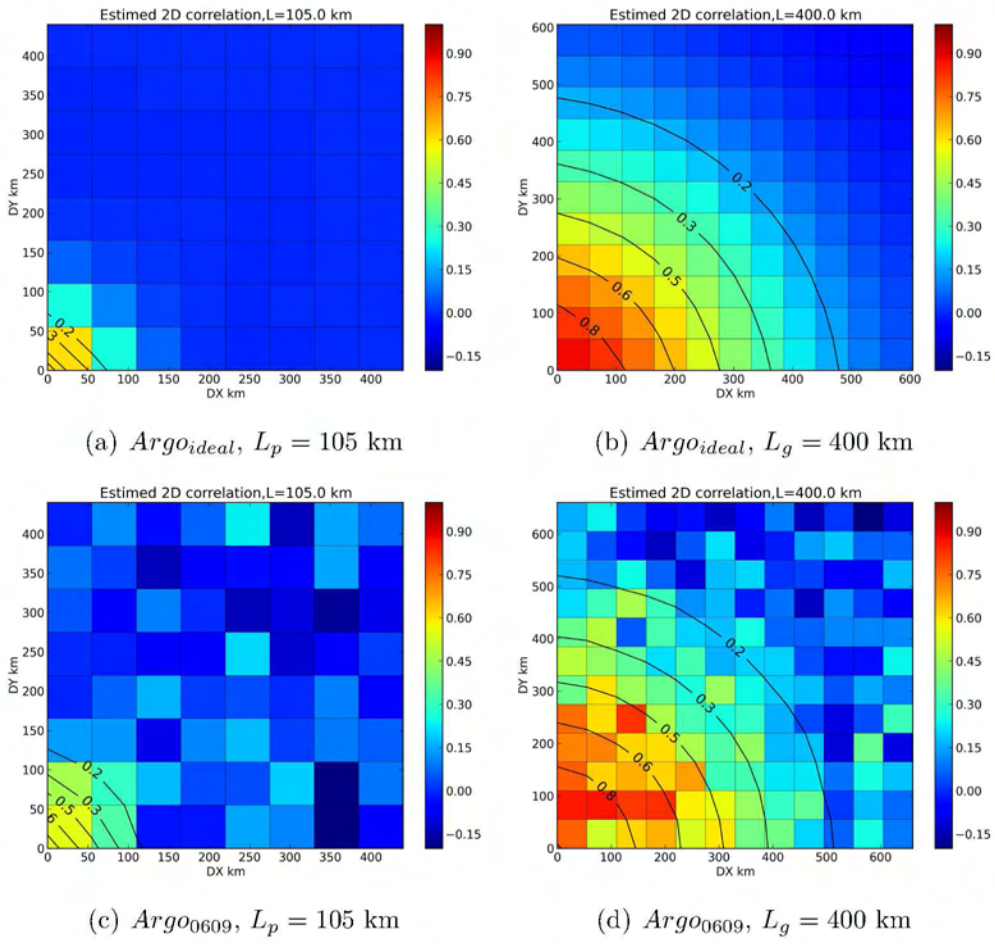
Figure 4.2 – *a) et b) Exemples de simulation pour un membre z^t, c) la matrice de covariance associée*

C- Corrélation spatiale issue du réseau $Argo_{ideal}$ sur les zones $\mathbf{R_{zone10}}$ et $\mathbf{R_{zone20}}$

Soit Δx_m la résolution spatiale des classes de distances. Du fait de la résolution spatiale du réseau $Argo_{ideal}$ et de la latitude maximale de 20°, il n'y a aucun profils espacés de moins de 50 km. On choisit alors un pas spatial égal à $\Delta x_m = 55$km. Les corrélations sont estimées via l'algorithme de calcul statistique de covariance décrit à la section 4.2.1. Nous avons réalisé un premier test d'estimation statistique sur la zone $\mathbf{R_{zone10}}$. L'estimation statistique pour de petites structures spatiales (L_p) est en accord avec la théorie. Cependant dans le cas de structures spatiales plus grandes (L_g) les résultats d'estimations sont en deçà de la théorie (non montrés). Une investigation auprès de la convergence des moments statistiques (moyenne et variance) des champs d'erreurs gaussiens utilisés a permis de confirmer que le nombre de semaines $N_{heb} = 209$ est insuffisant dans le cas de la zone $\mathbf{R_{zone10}}$. Les limitations numériques dues à la dimension de la zone $\mathbf{R_{zone10}}$ ne permettent pas une estimation correcte sur la période temporelle 2006-2009. En conséquence l'étude suivante sera faite sur la zone $\mathbf{R_{zone20}}$.

Estimation des corrélations spatiales 2D sur la zone $\mathbf{R_{zone20}}$

Les estimations des covariances statistiques à partir du réseau $Argo_{ideal}$ sont calculées pour deux types de structures spatiales correspondant aux longueurs de corrélation imposées : $L_p = 105$ km et $L_g = 400$ km. Les figures (4.3a et b) illustrent une bonne reconstruction du champ de corrélation 2D. On peut observer les contours de la fonction analytique justifiant une approximation correcte des estimations statistiques. L'ajustement non linéaire de la fonction analytique de corrélation donne ainsi des longueurs de corrélation proches de la réalité.. La connaissance des échelles de corrélation spatiales prescrites au sein des champs gaussiens $\mathcal{Z}$ nous fournit une référence afin d'exprimer des erreurs dites expérimentales (non montrées). Elles proviennent de la différence entre la fonction de covariance analytique et les estimations statistiques, et permettent ainsi de discuter de la cohérence des estimations statistiques. Le tableau 4.2 contient le bilan des ordres de grandeur des erreurs expérimentales. Pour la suite, ces résultats seront utiles comme comparaison au travail réalisé sur le réseau des positions des profileurs Argo, $Argo_{0609}$.



(a) $Argo_{ideal}$, $L_p = 105$ km (b) $Argo_{ideal}$, $L_g = 400$ km

(c) $Argo_{0609}$, $L_p = 105$ km (d) $Argo_{0609}$, $L_g = 400$ km

Figure 4.3 – *Champs de corrélation estimés et théoriques (contours en noir) sur la zone* $\mathbf{R_{zone20}}$ *pour les réseaux* $Argo_{ideal}$ *et* $Argo_{0609}$

	réseau $Argo_{ideal}$		réseau $Argo_{0609}$	
	L_p	L_g	L_p	L_g
L	107.86 ± 0.48 km	404.71 ± 0.49 km	110.80 ± 13.20 km	406.48 ± 11.62 km
$\mathfrak{e}$	0.22 ± 0.009	0.10 ± 0.001	0.61 ± 0.27	0.16 ± 0.04

Table 4.1 – *Longueur de corrélation et niveau de bruit de mesure pour chaque réseau utilisé, avec comme valeurs théoriques :* L_p=105km, L_g=400km *et* $\mathfrak{e}_{\text{th}} = 0.10$

D- Corrélation spatiale issue du réseau $Argo_{0609}$ sur la zone $\mathbf{R_{zone20}}$

Le réseau de profils possède des coordonnées dites discrètes car ce sont les positions actives des profileurs irrégulièrement dispersées dans le temps comme

$L_{p,g}$	réseau $Argo_{ideal}$	réseau $Argo_{0609}$
105 km	10^{-3}	10^{-1}
400 km	10^{-2}	10^{-1}

Table 4.2 – *Ordre de grandeur des écarts entre les corrélations estimées et théoriques dans le cas des réseaux $Argo_{ideal}$ et $Argo_{0609}$*

dans l'espace (voir la figure (4.1c). Le réseau comprend les données des années [2006-2009]. On rappelle que les valeurs aux points d'observations sont celles issues de l'ensemble des champs d'erreurs gaussiennes $\mathcal{Z}$ définis à la section B.

Confrontation des estimations à partir des deux réseaux utilisés

Nous allons procéder à une comparaison systématique des estimations réalisées, des erreurs expérimentales puis des ajustements non linéaires de la fonction analytique de corrélation. Les champs de corrélations reconstruits via le réseau $Argo_{0609}$ sont moins lisses que dans le cas du réseau $Argo_{ideal}$ et donnent lieu à des échelles de corrélation spatiales sensiblement plus grandes que celles prescrites en amont $L_{p,g}$ (voir figures 4.3). Les incertitudes portant sur l'ajustement des échelles de corrélation spatiales semblent elles aussi impactées par la géométrie du réseau utilisé. Pour le réseau $Argo_{0609}$, dans le cas des petites structures spatiales (L_p) et respectivement des grandes structures (L_g) elles sont de 12% et de 3%. Alors que pour le réseau $Argo_{ideal}$ et pour chaque configuration spatiale elles sont inférieures à 1% (valeurs numériques dans le tableau 4.1). La diminution du nombre d'observations et leur hétérogénéité au sein du réseau $Argo_{0609}$ ont pour conséquence l'augmentation des erreurs expérimentales (voir le tableau 4.2).

Analyse de sensibilité de l'estimation des corrélations spatiales

La résolution numérique du problème posé passe par une discrétisation afin d'obtenir les valeurs approchées de la fonction de corrélation analytique en un certain nombre de points. Le raffinement numérique est motivé par le but d'améliorer l'approximation des corrélations spatiales. Nous allons comparer les différents résultats issus des ajustements statistiques permettant d'obtenir les échelles de corrélation spatiales et le bruit de mesure, puis les champs de corrélation pour chacun des cas $\triangle x_m = 25$km et $\triangle x_m = 55$km.

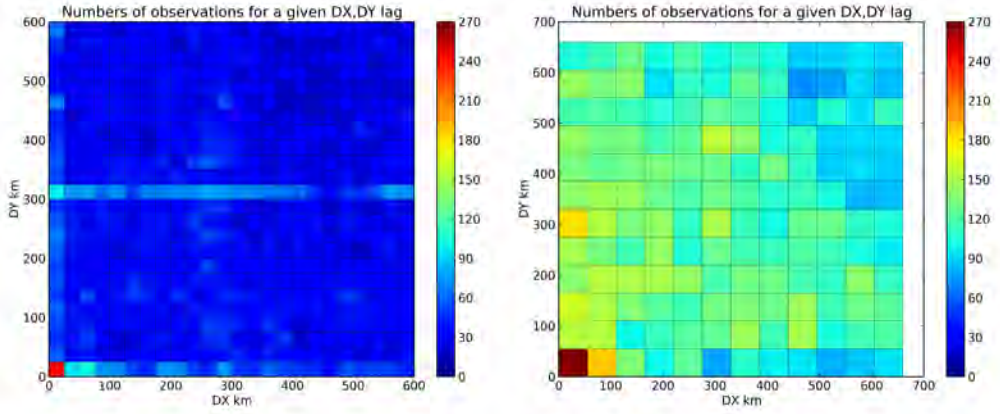
- La répartition des observations par classe pour les deux résolutions spatiales est homogène. Mais comme on peut s'y attendre elle est en moyenne environ 4 fois plus faible dans le cas $\triangle x_m = 25$ km (voir les figures 4.4, 4.5 b. et d.).
- Dans les deux cas $L_p = 105$ km et $L_g = 400$ km, les échelles de corrélation spatiales estimées sont cohérentes avec celles prescrites. Les incertitudes des

ajustements sont elles aussi similaires : de 12 à 13% pour $L_p = 105$ km, et de 2.5 à 3% pour $L_g = 400$ km (voir le tableau 4.3).

- Le bruit de mesure $\mathfrak{e}$ est inférieur pour les différentes $L_{p,g}$ lorsque $\triangle x_m = 25$km (et plus proche de celui prescrit en amont $\mathfrak{e}_{th} = 0.10$). Dans le cas de petites structures spatiales (L_p), le bruit de mesure estimé est très supérieur à la valeur attendue. On peut penser qu'un phénomène de petites échelles soit lié à ce résultat dans la mesure où dans le cas L_g, $\mathfrak{e}$ est raisonnablement proche de la valeur prescrite. Aussi nous avons aussi réalisé le cas $\triangle x_m = 100$ km pour $L_g = 400$ km permettant de souligner la croissance du bruit de mesure à mesure que $\triangle x_m$ augmente.

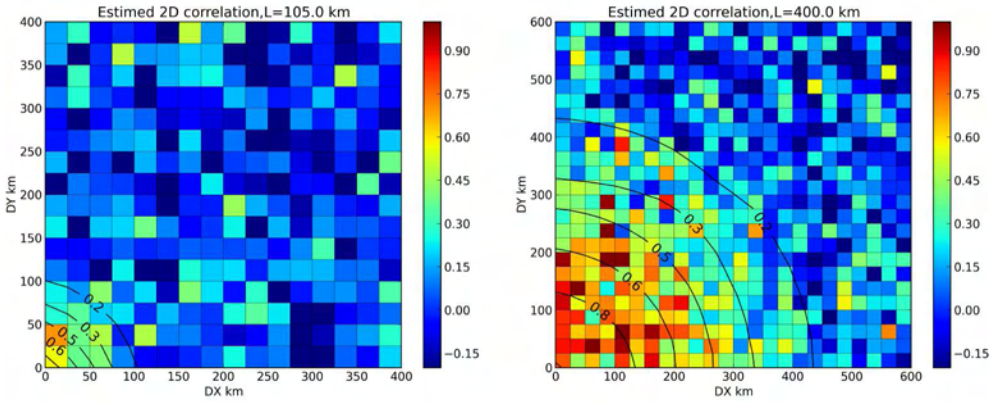
$\triangle x_m$	L_p	$\mathfrak{e}$
25	111.58 ± 13.47 km	0.56 ± 0.26
55	110.80 ± 13.20 km	0.61 ± 0.43
$\triangle x_m$	L_g	$\mathfrak{e}$
25	402.34 ± 11.24 km	0.14 ± 0.04
55	406.48 ± 11.62 km	0.16 ± 0.04
100	411.78 ± 11.61 km	0.19 ± 0.04

Table 4.3 – *Analyse de sensibilité réalisée selon les différentes résolutions spatiales ($\triangle x_m$) en fonction du type de structures spatiales observées $L_p = 105$ km, $L_g = 400$ km et $\mathfrak{e}_{th} = 0.10$*



(a) $\triangle x_m = 25$km (b) $\triangle x_m = 55$km

Figure 4.4 – *Nombre de profileurs répartis par classe distance sur la zone* $\mathbf{R_{zone20}}$, *pour a)* $\triangle x_m = 25km$ *b)* $\triangle x_m = 55km$ *pour le réseau* $Argo_{0609}$



(a) Corrélation, $L_p = 105$km (b) Correlation, $L_g = 400$km

Figure 4.5 – *Champ de corrélation pour* $\triangle x_m = 25km$*, dans les cas* $L_p = 105km$ *et* $L_g = 400km$*.*

4.2.4 Conclusion

Comme on a pu le constater lors de l'estimation statistique des échelles de corrélation spatiales d'une zone située aux basses latitudes, de dimension 20°x20° la densité numérique (certes plus faible que dans le cas idéalisé choisi), et la fréquence spatiale des observations non homogène du réseau $Argo_{0609}$ sont des facteurs limitants. L'estimation des échelles spatiales pour de petites structures (cas pour L_p) demande à la fois une résolution spatiale fine et une densité d'observations suffisante, ce qui est difficile pour cette zone. On obtient une estimation statistique satisfaisante à partir de cinquante paires d'observations, ce qui représente une limite minimale pour obtenir des statistiques robustes. Afin d'affiner la discrétisation des estimations des covariances, la résolution spatiale des classes de distance sera par la suite égale à $\triangle x_m = 25km$. Ce choix associé à la limite minimale d'effectifs par classe de distance amène à travailler sur des zones de plus grande dimension (voir section 4.3). Pour finir, l'ensemble de ces tests constitue un travail préparatoire à la caractérisation des échelles de corrélation spatiales des erreurs modèles et de la circulation océanique. Nous conserverons les choix de résolution spatiale et limite minimale d'effectifs par classe de distance lors des estimations réalisées en section 4.3.

4.3 Caractérisation des échelles spatiales des erreurs modèles et du signal à partir des données Argo

Nous utilisons les sorties modèles de la simulation $T335$ colocalisées au point d'observation (détaillé au chapitre 3). La colocalisation a aussi été réalisée pour les données climatologiques en température et salinité (Levitus 2009). Ainsi les fonctions de corrélation spatiales exploitées dans ce chapitre sont obtenues à partir des anomalies $z = [modele - observations]$ et des anomalies $z = [climatologie - observations]$. Dans un premier temps, les profils d'observations Argo sont sélectionnés sur chaque zone définie sur la figure 4.6, et en fonction de leurs dates, espacées sur une fenêtre temporelle de 7 jours. Selon le même raisonnement décrit en section 4.2, l'estimation statistique de la matrice de covariance est faite à partir de couples de profils appartenant à des classes de distance $\mathbf{C}_{(\triangle x, \triangle y)}$. On les regroupe en plusieurs classes de telle sorte que les profils d'un même classe soient les plus semblables possibles et que les classes soient les plus distinctes possibles. L'ensemble du calcul est fait avec une résolution spatiale zonale et méridionale égale à 25km

4.3.1 Structures et caractéristiques des régions examinées

Les échelles spatiales de corrélation sont estimées sur l'ensemble des océans par le biais de 14 zones grande échelle.

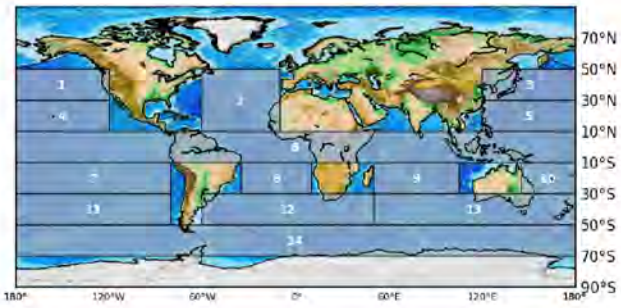
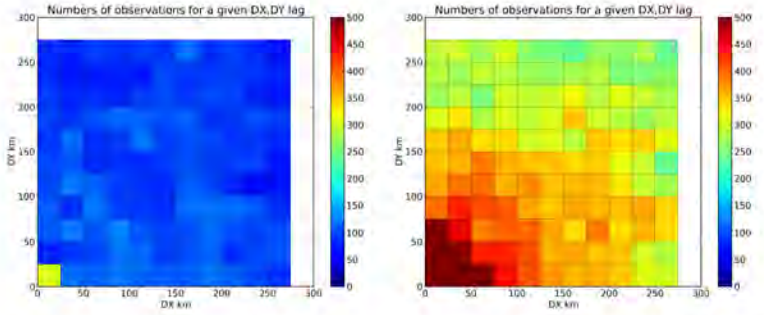


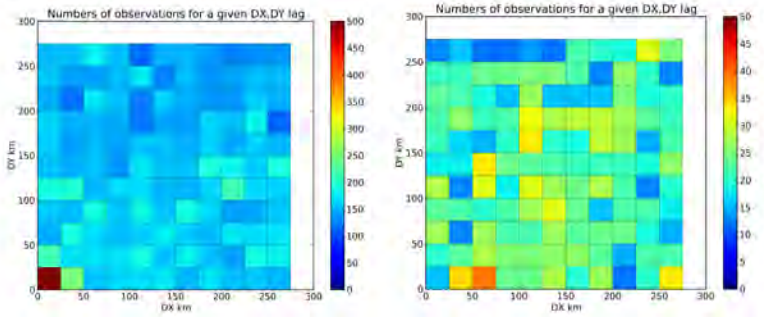
Figure 4.6 – *Ensemble des zones où les échelles spatiales de corrélation sont estimées.*

Les figures 4.7 et 4.8 représentent le nombre de paires de profils par classe de distance (N, défini en section 4.2) pour plusieurs zones situées dans l'océan Pacifique, Atlantique, au latitudes tropicales et aux hautes latitudes de l'hémisphère

sud cumulés sur la période 2006 – 2009. On note une densité plus élévée dans le Pacifique Nord-Ouest que dans le Pacifique Nord-Est (résultat déjà observé dans le chapitre 2, figure 3.4). Globalement la répartition des paires d'observations dans le Pacifique Nord est dense et uniforme. Ce type de distribution reste similaire dans le Pacifique Sud-Est. Cependant les couples de profils situés dans le sud du Pacifique Ouest sont plus parsemés, le nombre de paires d'observations est inférieur à 60 (zone 10, figure 4.7). L'océan Atlantique Nord possède une densité d'observation homogène et moyennement dense (N compris entre 100 et 150, figure 4.8). La zone 8 située dans l'hémisphère sud est moins bien échantillonnée (nombre de paires de profils inférieur à 60). Par ailleurs dans la région tropicale et aux hautes latitudes dans l'hémisphère sud, N est respectivement compris entre 300 et 500, 150 et 250. Compte tenu des dimensions très grande échelle de la zone 14, la densité d'observation reste moyennement basse (360°x20°, figure 4.6). Enfin la densité d'observations dans les bassins Sud-Indien et Austral est homogène et abondante (100 à 200 paires de profils, figures en annexe B.4). Pour finir, à l'ex-



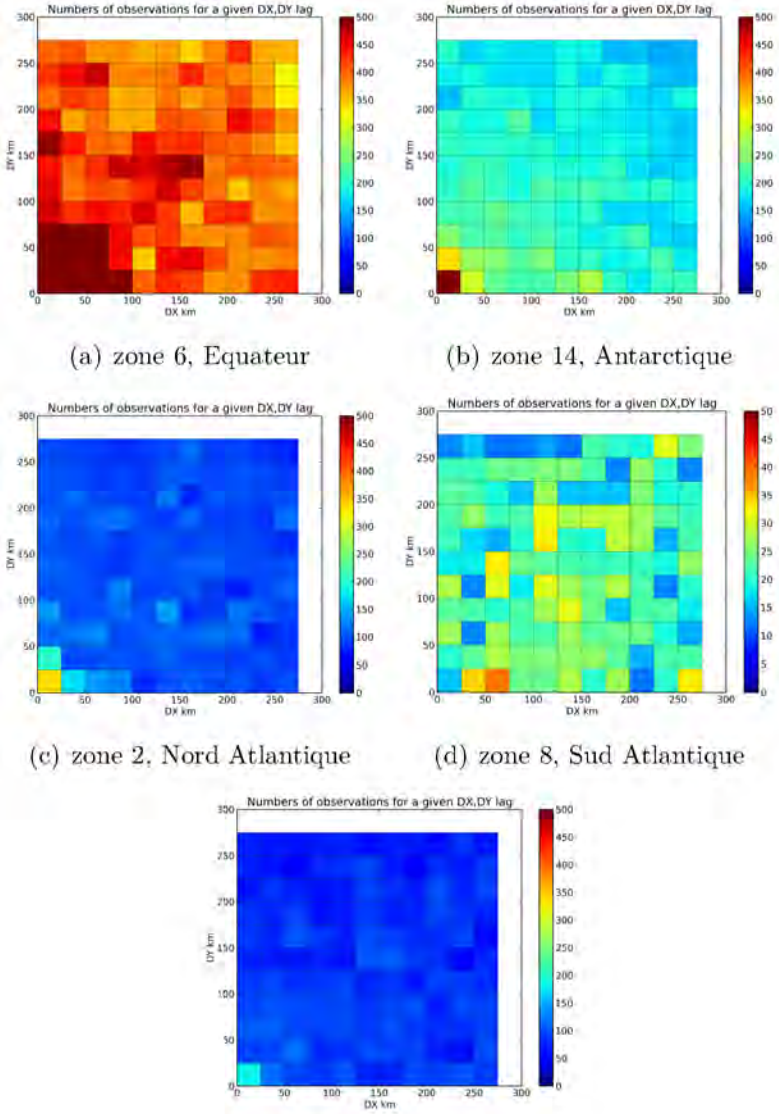
(a) zone 1, Pacifique Nord-Est (b) zone 3, Pacifique Nord-Ouest



(c) zone 7, Pacifique Sud-Est (d) zone 10, Pacifique Sud-Ouest

Figure 4.7 – *Répartition des paires d'observations selon les classes de distances pour les régions situées dans l'océan Pacifique, à* 200 *m sur la période* 2006 – 2009

ception des zones 8 et 10 dont la densité d'observations est faible, la répartition des paires de profils des zones examinées est continue et décroissante lorsque la distance augmente. Les zones 8 et 10 ne seront pas prises en compte lors de l'étude, et confirment la nécessité de travailler avec des zones de dimensions minimales de 60°x20°, du moins sur la période 2006 – 2009 exploitée dans ce chapitre.



(a) zone 6, Equateur

(b) zone 14, Antarctique

(c) zone 2, Nord Atlantique

(d) zone 8, Sud Atlantique

(e) zone 12, Sud Atlantique

Figure 4.8 – *Répartition des paires d'observations selon les classes de distances pour les régions situées dans l'océan Atlantique, aux tropiques et aux hautes latitudes de l'hémisphère sud, à 200 m, sur la période 2006 – 2009*

4.3.2 Utilisation d'une nouvelle formulation analytique du modèle de corrélation

Lors des tests réalisés à partir des données simulées (voir section 4.2.3) nous avons utilisé un modèle gaussien qui s'est avéré moins adapté lors de l'ajustement aux données réelles Argo. C'est pourquoi le modèle de fonction de corrélation analytique utilisé ici repose sur celui proposé par *Arhan and Verdière* [1985]. Ce modèle avait été ajusté sur les mesures in-situ (température et salinité) de l'expérience océanographique Tourbillon (*groupe Tourbillon* [1978]), les détails théoriques sur l'utilisation d'un tel modèle sont présents dans la thèse de Pierre-Yves Le Traon, *Le Traon* [1990]. Il a ensuite été utilisé pour la cartographie des données altimétriques (*Le Traon et al.* [1998], *Ducet et al.* [2000] et *Le Traon et al.* [2003]). C'est donc ce modèle qui sera ajusté aux observations Argo selon la méthode décrite au paragraphe 4.2.2.

$$\mathbf{f}(r) = \begin{cases} (1+\mathfrak{e}).[1+ar+\frac{1}{6}(ar)^2-\frac{1}{6}(ar)^3]\exp(-ar) & si \quad r \neq 0 \\ 1+\mathfrak{e} & si \quad r = 0 \end{cases} \tag{4.10}$$

où $r = \sqrt{(\frac{dx}{L_x})^2 + (\frac{dy}{L_y})^2}$, a est une constante égale à $3,337$ et $\mathfrak{e}$ le bruit du aux erreurs de mesures et d'observations.

Dans les sections suivantes nous allons rendre compte des tendances globales des variations des échelles spatiales en température et en salinité issues des erreurs modèles (section 4.3.3). Ces estimations seront ensuite comparées à celles issues des anomalies par rapport à la climatologie mensuelle Levitus 2009 (section 4.3.4).

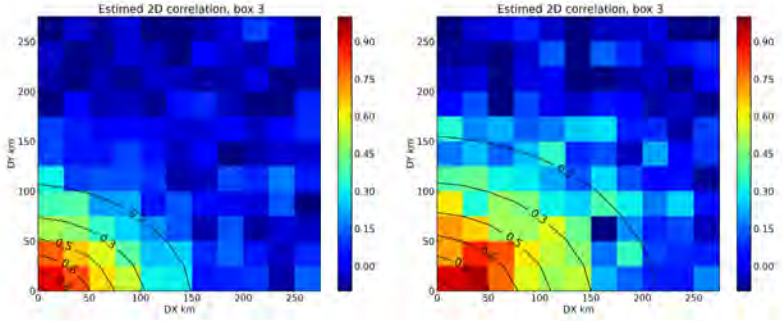
4.3.3 Analyse globale des échelles spatiales des erreurs modèles en température et salinité

A titre d'exemple, les figures 4.9a et 4.9b présentent les champs de corrélation spatiales en température pour une zone située dans l'océan Nord Pacifique à 200 et 1000 m. Comme décrit en section 4.3.1 la zone 3 dispose d'un large nombre de paires de profils permettant d'obtenir des estimations statistiques correctes (N compris entre 200 et 400, figure 4.7). Les échelles spatiales zonales et respectivement méridionales sont de 150 km et 115 km à 200 m de profondeur. Ceci semble en accord avec les résultats connus et obtenus via les analyses de données altimétriques (*Jacobs et al.* [2001]; *LeTraon et al.* [2003]). On note aussi l'augmentation des échelles lorsque la profondeur augmente : respectivement $L_x = 210$

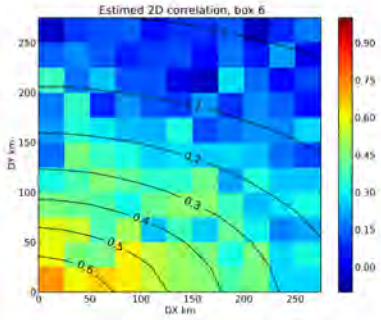
km et $L_y = 160$ km à 1000 m. Le tableau 4.4 regroupant les valeurs des échelles spatiales zonales et méridionales selon plusieurs bassins et profondeurs, confirme cette tendance à l'exception de la zone 6 située dans la région tropicale (ceci sera détaillé en section 4.3.4). Comme on peut le constater sur la figure 4.10, l'ensemble des échelles spatiales diminuent lorsqu'on s'éloigne de l'Equateur. Les figures 4.9c et 4.10 soulignent l'augmentation des échelles spatiales dans la région tropicale et aussi le caractère anisotropique des champs de corrélation spatiales plus marqué à ces latitudes (zone 6). Les échelles spatiales de corrélation sont estimées à $L_x = 500$ km et $L_y = 270$ km à 200 m.

Les figures 4.12, 4.15 et 4.16 présentent les distributions verticales des échelles zonales et méridionales estimées à partir des erreurs modèles (courbe bleu) pour plusieurs zones. Pour des profondeurs supérieures à 800-1000 m, les échelles spatiales zonales et méridionales en température ont tendance à fortement augmenter. Par exemple pour la zone 11, les échelles spatiales atteignent à 1400 m, $L_x = 770$ km et $L_y = 800$ km (voir figure 4.16), et ce malgrè des champs de corrélation relativement bruités (perte du signal). La troisième colonne du tableau 4.4 contient les valeurs du bruit de mesure pour les différents zones et profondeurs étudiées. Le bruit de mesure engloge à la fois le bruit réel de mesure e l'effet des petites structures non ou mal résolues par notre pas d'échantillonnage (25 km) et notre modèle analytiqur de corrélation. Le bruit de mesure est présent sur l'ensemble des zones étudiées, variant de 0.1 à parfois 0.9 ce qui peut s'interpréter comme si les observations étaient entachées d'erreurs mais cela reflète aussi le fait que dans certaines régions le modèle de covariance analytique n'est sans doute pas bien adapté. En outre les hypothèses d'homogénéité sur la zone d'estimation ne sont pas nécessairement respectées sur l'ensemble des régions étudiées et cela peut bruiter nos estimations. On observe des valeurs très faibles entre 200 et 1000 m (colonne 3 du tableau 4.4). Puis le bruit de mesure est compris entre 0.3 et 0.6 à partir de 1000 m.

Enfin il est intéressant de noter que les échelles spatiales en salinité présentent des différences significatives par rapport à celles en température. On n'observe pas notamment une augmentation des échelles en salinité à partir de $800-1000$ mètres de profondeur contrairement au cas en température. Plusieurs zones montrent aussi des variations rapides des échelles en salinité sur la verticale (zones $2, 4, 6$ et 11). Certaines régions (par exemple la zone 2 située dans l'océan Atlantique Nord avec la présence des eaux méditerranéenes en profondeur) présentent des signaux très spécifiques de salinité. Il n'est pas étonnant dans ces conditions d'observer des différences entre les échelles spatiales en salinité et en températures ; elles reflètent vraisemblablement des dynamiques différentes.



(a) zone 3, à 200 m (b) zone 3 à 1000 m



(c) zone 6, à 200m

Figure 4.9 – *Corrélation estimée en température à partir des erreurs modèles à* 200m *et* 1000m *dans l'Océan Nord Pacifique sur la période* 2006 – 2009

Zone	L_x		L_y		$\mathfrak{e}$	
	200m	1000m	200m	1000m	200m	1000m
2	150 ± 25 km	210 ± 25 km	100 ± 20 km	110 ± 15 km	0.01	0.3
3	160 ± 10 km	220 ± 15 km	115 ± 8 km	160 ± 10 km	10^{-11}	10^{-10}
6	500 ± 30 km	350 ± 30 km	270 ± 10 km	300 ± 20 km	0.6	0.25
12	120 ± 15 km	130 ± 20 km	110 ± 15 km	140 ± 20 km	10^{-8}	10^{-7}
13	110 ± 10 km	135 ± 10 km	100 ± 10 km	130 ± 10 km	10^{-9}	10^{-9}
14	100 ± 15 km	180 ± 20 km	80 ± 10 km	170 ± 20 km	0.1	0.3

Table 4.4 – *Echelles spatiales zonales* (L_x) *et méridionales* (L_y) *pour plusieurs bassins (zones), et bruit de mesure* ($\mathfrak{e}$) *des observations estimés à* 200 *et* 1000 m.

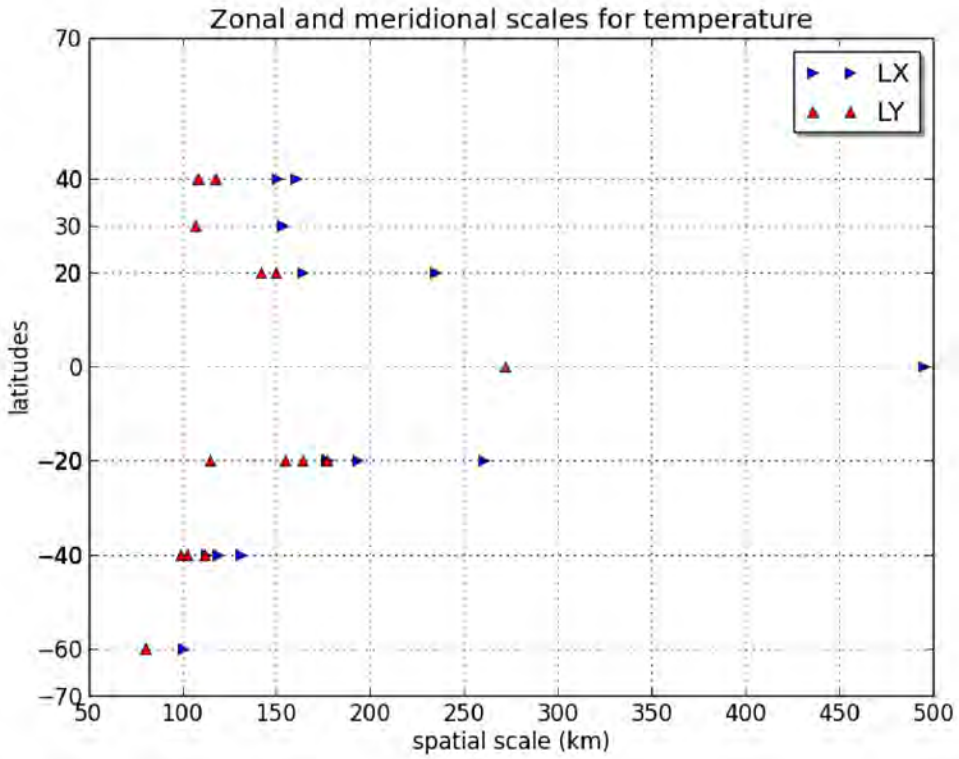


Figure 4.10 – *Echelles spatiales zonales (bleu) et méridionales (rouge) des erreurs modèles en température en fonction des latitudes à* 200 *m sur la période* 2006−2009.

4.3.4 Comparaison des échelles spatiales des erreurs modèles par rapport aux échelles de la variabilité océanique

Nous allons réaliser une description comparative pour chaque bassin océanique en prenant comme référence les échelles spatiales de corrélation estimées à partir des anomalies [observations-climatologie], en température et salinité. L'ensemble des figures 4.11 à 4.16 représentent les échelles spatiales des erreurs modèles (courbe bleu) et des anomalies par rapport à la climatologie (courbe verte) en fonction de la profondeur. Les échelles spatiales de corrélation représentatives de la circulation océanique sont plus précisemment détaillées au cours du chapitre 4.4.

Echelles spatiales en température

Sur les premiers 200 m, les résultats d'estimations à partir des erreur modèles ne présentent pas de grandes échelles spatiales en température à contrario du cas climatologique. En effet les échelles spatiales de corrélation des erreurs modèles en surface ou au sein de la couche de mélange sont comprises entre 100 et 200 km à l'exception de la zone tropicale (courbe bleu, figure 4.13). Cela est sans doute lié à la capacité du modèle contrairement à la climatologie, de représenter certaines structures grande échelle de température, ainsi que des échelles intrasaisonnières et interannuelles. Aux latitudes tropicales les échelles zonales et méridionales en température atteignent respectivement 400 à 550 km et 200 à 300 km. En deçà de la couche de mélange les échelles spatiales en température sont relativement en accord avec celles désignées comme référence (cas climatologique, courbe verte sur les figures). On note cependant des écarts entre les deux cas d'estimations supérieurs à 100 km pour les zones situées aux latitudes tropicales (zone 6, figure 4.13), et dans l'océan Pacifique (zones 4 et 11, figures 4.15, 4.16). Ces écarts augmentent avec la profondeur mais les tendances globales des échelles spatiales des erreurs modèles restent en accord avec la variabilité observée.

Echelles spatiales en salinité

Nous pouvons constater que les estimations des échelles spatiales en température dans le cas des erreurs modèles sont relativement proches de celles estimées dans le cas de référence. Aussi les échelles spatiales de corrélation des erreurs modèles en salinité, de la surface à l'océan profond, des zones situées aux hautes latitudes de l'hémisphère sud, dans l'océan Atlantique Sud et Austral (zone 13, figure B.10 en annexe) sont proches de celles déduites des écarts à la climatologie. Par ailleurs aux latitudes tropicales on observe entre 0 et 200 m, des écarts entre les deux cas d'estimations supérieurs à 100 km, augmentant avec la profondeur et plus conséquents selon la direction zonale (figure 4.13). La même tendance en surface

est présente au sein de l'océan Pacifique (figures 4.15, 4.16). Enfin dès 200 m, les échelles spatiales en salinité estimées au sein des zones situées dans les océans Pacifique et Nord Atlantique sont supérieures à celles obtenues via les anomalies par rapport à la climatologie. Ceci est observé pour la zone 2, et pour des couches plus profondes (300 à 400 m) pour les zones 1 et 7 (figures en annexe B). Ce comportement semble indiquer que le modèle ne représente pas la variabilité en salinité telle qu'elle a pu être perçue par les observations Argo (i.e différences significatives entre les échelles dans le cas des erreurs modèles et celles issues du signal [observations-climatologie]).

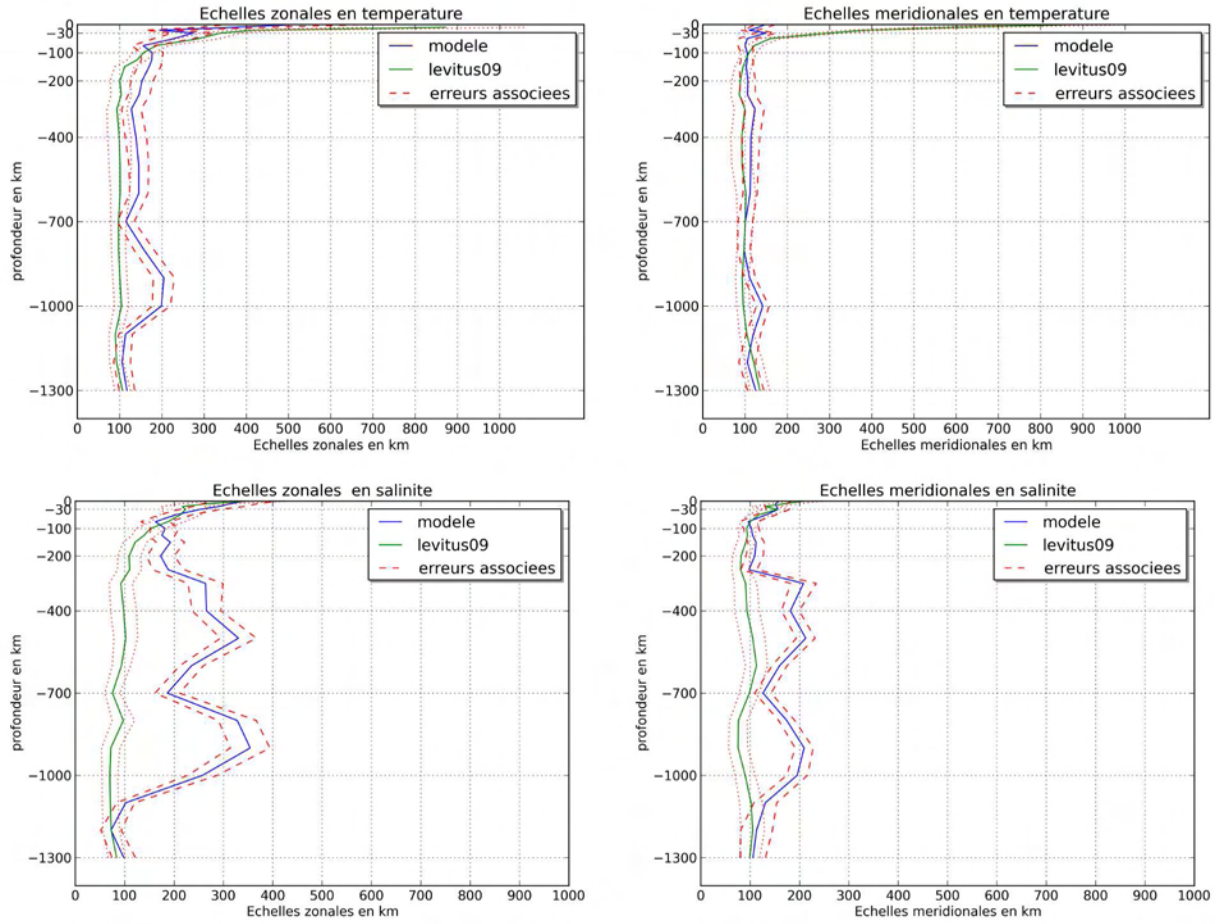


Figure 4.11 – *Corrélation estimée en température (haut) et en salinité (bas) à partir des erreurs modèles (bleu) et pour le signal ([climatologie − observations], en vert) dans l'océan Atlantique Nord (zone 2). Les incertitudes liées à l'ajustement des échelles sont aussi représentées (pointillés rouge)*

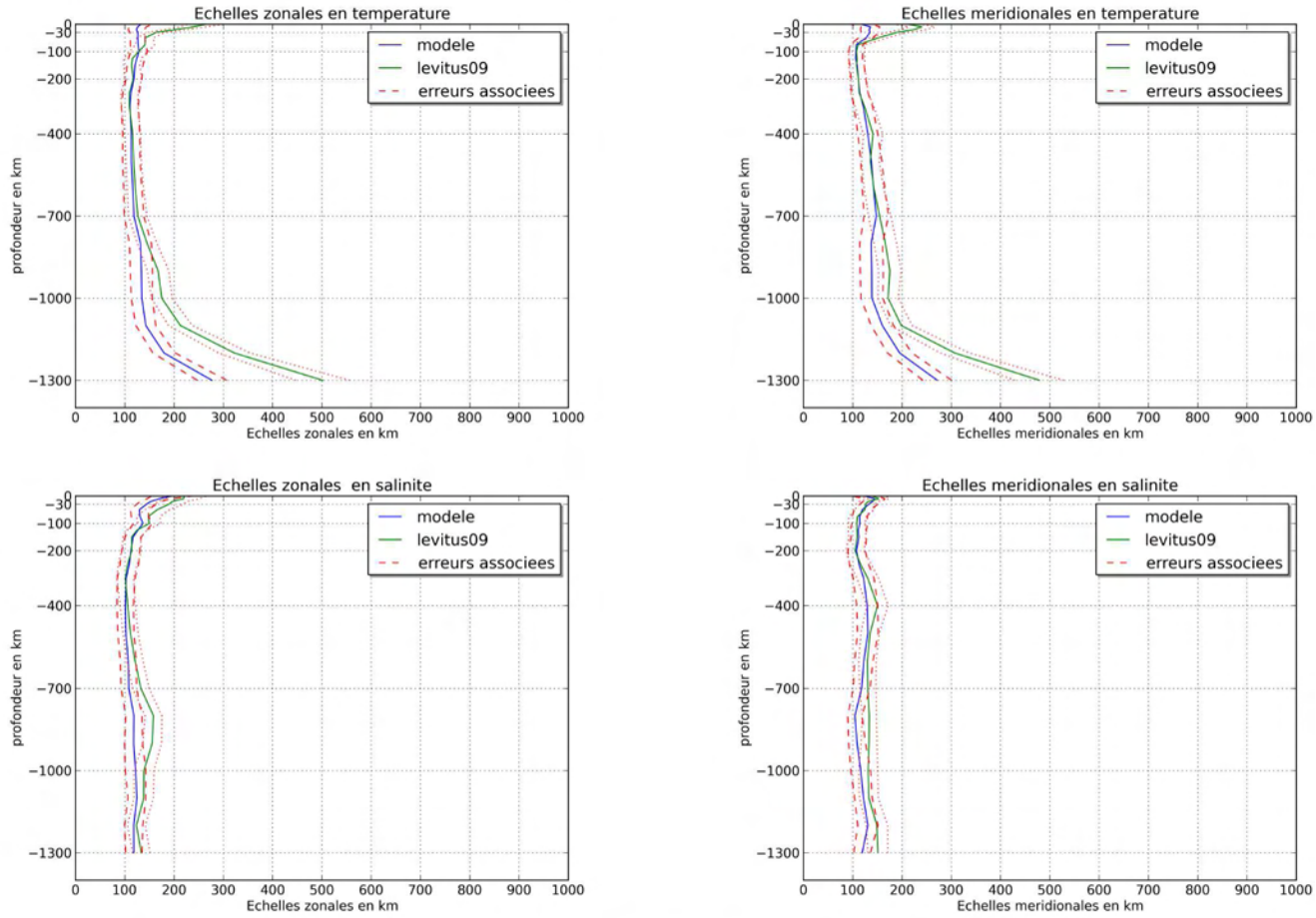


Figure 4.12 – *Corrélation estimée en température (haut) et en salinité (bas) à partir des erreurs modèles (bleu) et pour le signal ([climatologie − observations], en vert) dans l'océan Atlantique Sud (zone* 12*). Les incertitudes liées à l'ajustement des échelles sont aussi représentées (pointillés rouge)*

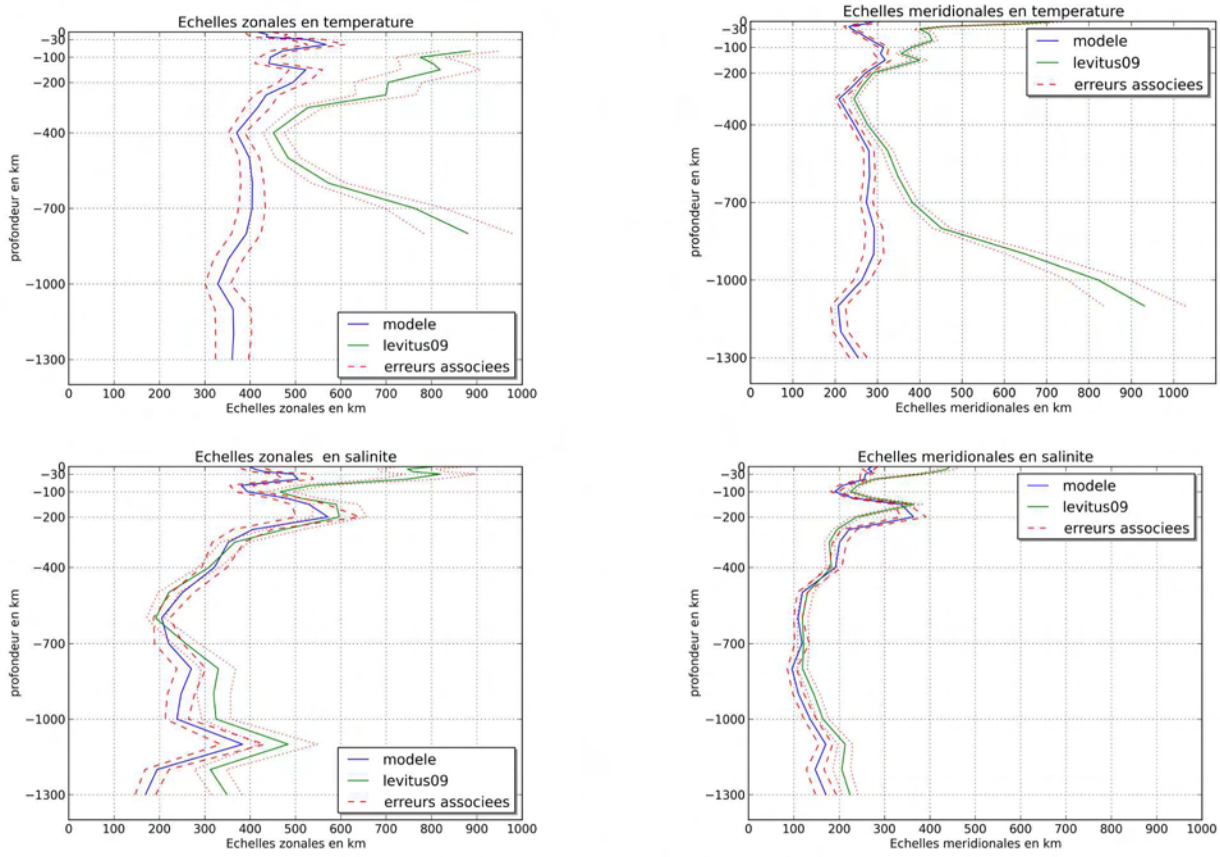


Figure 4.13 – *Corrélation estimée en température (haut) et en salinité (bas) à partir des erreurs modèles (bleu) et pour le signal ([climatologie – observations] en vert) dans la région tropicale (zone 6). Les incertitudes liées à l'ajustement des échelles sont aussi représentées (pointillés rouge)*

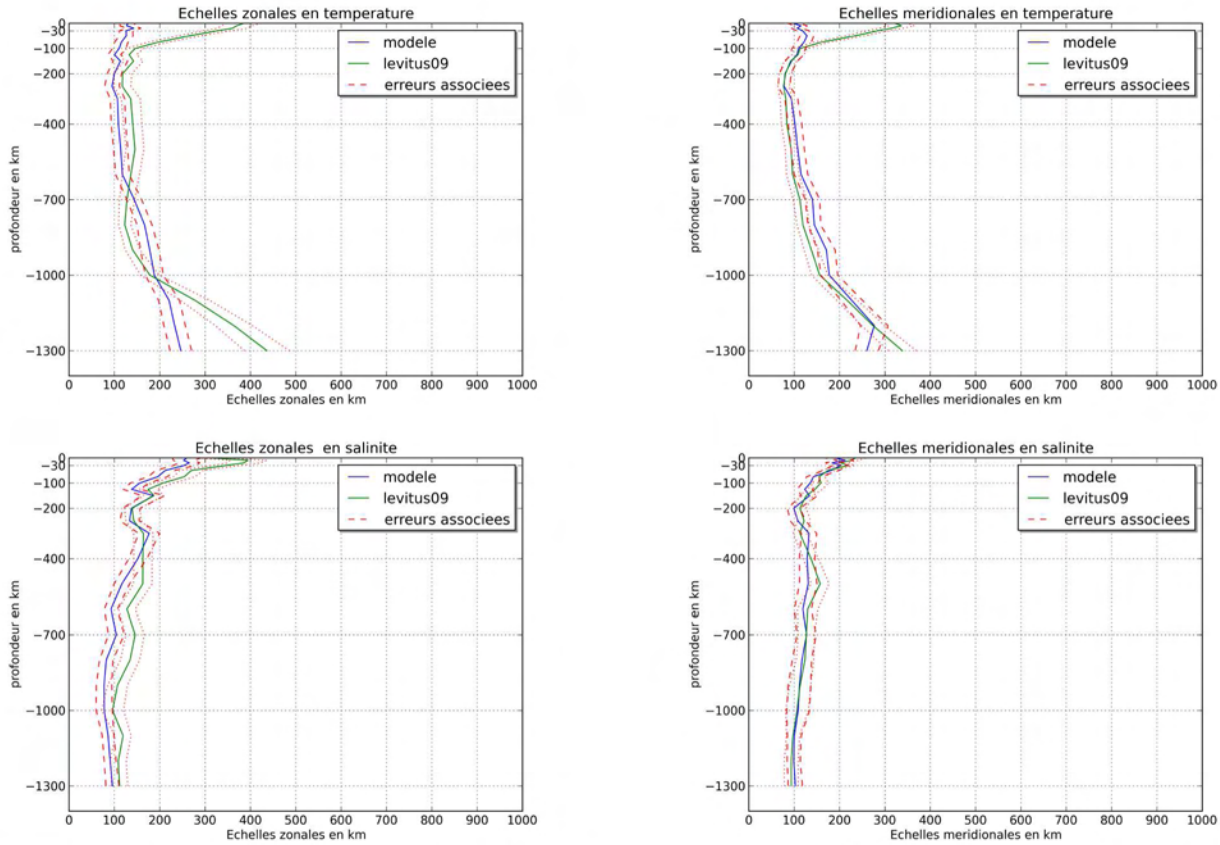


Figure 4.14 – *Corrélation estimée en température (haut) et en salinité (bas) à partir des erreurs modèles (bleu) et pour le signal ([climatologie − observations] en vert) aux hautes latitudes de l'hémisphère Sud (zone* 14*). Les incertitudes liées à l'ajustement des échelles sont aussi représentées (pointillés rouge)*

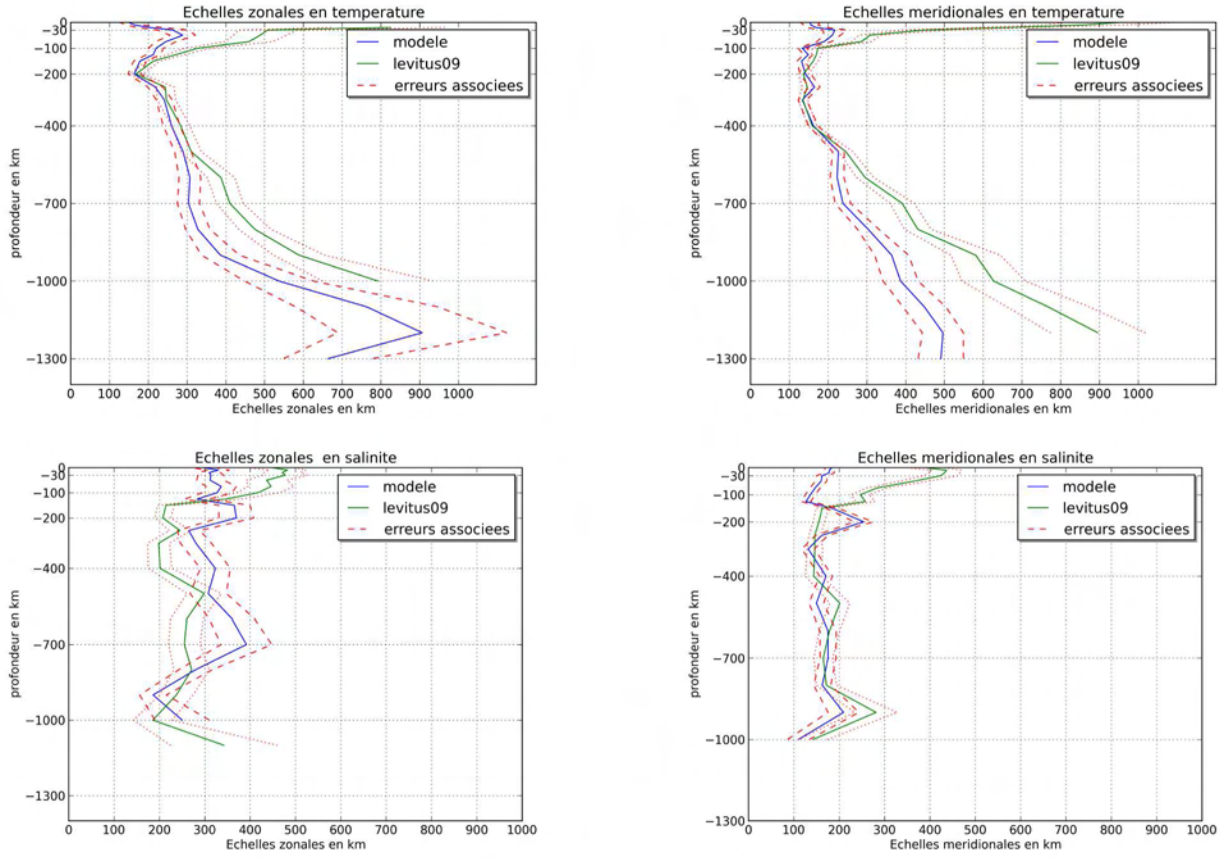


Figure 4.15 – *Corrélation estimée en température (haut) et en salinité (bas) à partir des erreurs modèles (bleu) et pour le signal [climatologie – observations] (vert) dans l'océan Nord-Est Pacifique (zone 4). Les incertitudes liées à l'ajustement des échelles sont aussi représentées (pointillés rouge)*

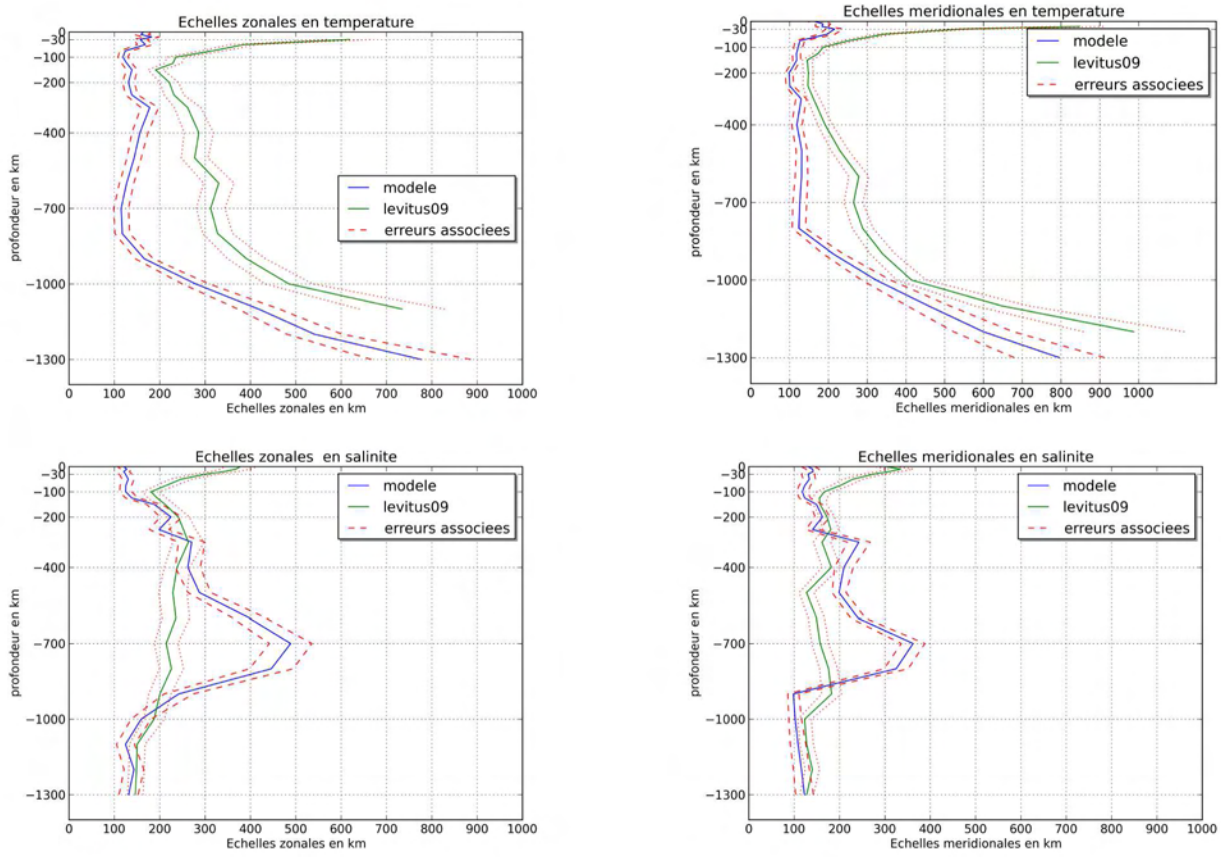


Figure 4.16 – *Corrélation estimée en température (haut) et en salinité (bas) à partir des erreurs modèles (bleu) et pour le signal [climatologie – observations] (vert) dans l'océan Sud-Est Pacifique (zone* 11*). Les incertitudes liées à l'ajustement des échelles sont aussi représentées (pointillés rouge)*

4.4 Estimation des échelles spatiales des variations de température et salinité à partir des données Argo

Le travail réalisé en section 4.3 est ici poursuivi et a permis la parution dans le journal *Ocean Science* de l'article intitulé *Spatial scales of temperature and salinity variability estimated from Argo observations*. Lors de cette étude nous utilisons les observations Argo sur la période 2005-2013 afin de caractériser les échelles spatiales des variations en température et salinité de la surface jusqu'à 1500 m de profondeur. Nous avons aussi calculé les erreurs formelles sur l'estimation des covariances (détaillé en section 4.4.2) permettant d'obtenir une estimation plus précise des erreurs sur les échelles spatiales. Enfin la section 4.3 a permis de réaliser une étape préparatoire quant au découpage de l'ensemble de l'océan en plusieurs zones de grande échelle (la zone 6 équatoriale en section 4.3 dynamiquement hétérogène a été redécoupé comme on peut le voir en figure 3 de l'article).

4.4.1 Article paru dans le journal *Ocean Science*

Dans un premier temps nous avons réalisé une étude de sensibilité à partir de données simulées afin d'analyser l'influence de l'échantillonnage du réseau Argo sur ces estimations. En conséquence les échelles spatiales de la variabilité océanique sont estimées sur des zones de grande échelle. Les échelles spatiales zonales et méridionales varient de 100 km aux hautes latitudes à plus de 700 km dans les régions équatoriales et tropicales. Les échelles méridionales sont supérieures (facteur de 2 à 3) aux échelles zonales particulièrement dans les régions aux latitudes tropicales. On observe des échelles plus grandes en surface, et une tendance croissante en profondeur pour les échelles en température. Enfin on note des différences significatives entre les échelles spatiales en température et en salinité, notamment en profondeur.

Spatial scales of temperature and salinity variability estimated from Argo observations

F. Ninove[1], P.Y. Le Traon[1,2], E. Remy[2] and S. Guinehut[3]

[1] {Ifremer, Plouzané, France}

[2] {Mercator Ocean, Ramonville Saint Agne, France}

Correspondence to: P.-Y. Le Traon (pierre-yves.letraon@mercator-ocean.fr)

Accepted to Ocean Science

June 24, 2015

Abstract

Argo observations from 2005 to 2013 are used to characterize spatial scales temperature and salinity variations from the surface down to 1500 m. Simulations are first performed to analyze the sensitivity of results to Argo sampling; they show that several years of Argo observations are required to estimate of spatial scales of ocean variability over 20°x20° boxes. Spatial scales are then computed over several large scale areas. Zonal and meridional spatial scales (Lx and Ly which are also zero crossing of covariance functions) vary as expected with latitudes. Scales are of about 100 km at high latitudes and more of 700 km in the Indian and Pacific equatorial/tropical regions. Zonal and meridional scales are similar expect in these tropical/equatorial regions where zonal scales are much larger (by a factor of 2 to 3) than meridional scales. Spatial scales are the largest close to the surface and have a general tendency for temperature to increase in deeper layers. There are significant differences between temperature and salinity scales, in particular, in the deep ocean. Results are consistent with previous studies based on sparse in-situ observations or satellite altimetry. They provide, however, for the first time a global description of temperature and salinity scales of variability and a characterization of their variations according to depths.

1 Introduction

Thanks to outstanding international cooperation, Argo the global array of profiling floats (Roemmich et al., 2009) reached its initial target of 3000 floats in operation in 2007. Argo floats measure every 10 days temperature and salinity from the surface down to 2000 m and deliver their data both in real time for operational users and after scientific quality control for climate change research and monitoring.

Argo has revolutionized oceanography by providing for the first time a near real time global description of the ocean state that is fully complementary to satellite observations. An overview of Argo achievements is given in Freeland et al. (2010). Argo data have been used to better understand global and regional sea level rise and ocean heat content variations (e.g. von Schuckmann and Le Traon, 2011), to analyze large scale ocean circulation and mesoscale variations (e.g. Roemmich et al; 2007; Dong et al., 2014) and large scale salinity variations related to the global hydrological cycle (Durack and Wijffels, 2010). Argo has strong complementarities with satellite altimetry and Argo data are now systematically used together with altimeter data for ocean analysis and forecasting (e.g. Guinehut et al., 2012; Le Traon, 2013; Oke et al., 2015).

The availability of global temperature and salinity data sets over several years is a unique opportunity to better characterize the statistics of ocean mesoscale variability at global scale. Although Argo does not resolve mesoscale variability due to its 3°x3° spatial sampling, it is very well suited to estimate its main statistical characteristics. Guinehut et al. (2012) derived, for example, statistical relationships between surface and subsurface fields to infer the 3D mesoscale T&S fields from altimetry and sea surface temperature (SST) and Argo observations. We focus here on the spatial scales of temperature and salinity variations. Over several years and in given region, there are many nearly simultaneous pairs of floats with different separation distances allowing an estimation of such scales. These estimations are important to better characterize and understand ocean dynamics and to improve quality control, mapping or data assimilation schemes (e.g. Gaillard et al., 2009; Roemmich and Gilson, 2009). They are also essential to refine the sampling requirements for the Argo global array as an optimal sampling should reflect the actual spatial (and time) scales of ocean variability.

The paper is organized as follows. Data and methods are presented in section 2. The capability of Argo sampling to estimate spatial correlation scales is analyzed with simulated data in section 3. Section 4 provides a global calculation of spatial scales and discusses the main results. Conclusions and perspectives are given in section 5.

2 Data sets and methods

We used Argo observations from 2005 to 2013 as obtained from the Coriolis data center. Data from 2005 to 2012 are delayed mode quality controlled data from the CORA data base (Cabanes et al., 2013). Data from 2013 are near real time data from the Coriolis Argo Global Data Assembly Center (one of the two Argo GDACs). An additional quality control with regional climatology checks was applied to these near real time data sets.

After several tests (see discussion in section 3), correlation scales were calculated over several large scale areas to provide sufficient pair of observations at different zonal and meridional distances.

Correlations were computed both for temperature and salinity and for the surface down to 1300 m. The following steps are used for the calculation:

1. The Levitus 2009 seasonal climatology is removed from Argo profile observations.

2. All temperature and salinity Argo data (from 2005 to 2013) within a given box (e.g. 20° latitude x20° longitude up to 20° latitude x 100° longitude) are gathered and a large scale (4°x4°) seasonal mean of observations is computed and removed from the observations. This allows removing possible biases in the climatology. Data are then stored in weekly files.

3. The covariance for a given zonal (dx) and meridional (dy) distance is then calculated as:

   $Cov(dx, dy) = Var - 0.5\ \gamma(dx, dy)$,

   $\gamma(dx, dy) = 1/N \sum_{all\ weeks} \sum_{ij} (z'(xi,yi,ti)-z'(xj,yj,tj))^2$,

   where var is the variance, γ is the variogram and z' is the anomaly of temperature or salinity at a given depth and N is the number of pairs of Argo profiles whose zonal (xi-xj) and meridional (yi-yj) distances are comprised between dx ± 12.5 km and dy ± 12.5 km and whose time separations are comprised between ± 3.5 days. The calculation is done with a spatial zonal and meridional resolution of 25 km. Note that covariances were derived from a variogram calculation to reduce sensitivity of results to unknown mean fields. Since we remove large scale fields (see point 2 above) prior to the calculation, this only has a small impact on the calculation.

4. Covariances are then normalized by the variance to get correlation values:

   $Cor(dx, dy) = Cov(dx, dy)/Cov(0,0)$.

5. The formal error variance on the correlation, noted Ɛ here, is then derived following the Isserlis theorem (Bendat and Piersol, 1986) and expressed as:

   $Var(Ɛ) = 1/N^2 \sum_{all\ weeks} \sum_{ij} Cor(dx_{ij}, dy_{ij})^2 + Cor(dx+dx_{ij}, dy+dy_{ij}).\ Cor(dx-dx_{ij}, dy-dy_{ij})$.

   dx_{ij} and dy_{ij} are the zonal and meridional distances between Argo profile i and Argo profile j. In practice this calculation is done in an iterative way after an analytical model (see below) is fitted to covariance observations. Note that if we assume that the N pairs of observations provide uncorrelated estimations of the covariance for a given dx and dy lag (which is the case if $dx_{ij} >> dx$ or $dy_{ij} >> dy$), the formal error variance on the correlation is simply equal to the following expression:

   $Var(Ɛ) = 1/N\ [1+cor(dx,dy)]$.

   It shows that RMS errors on correlations vary as $1/\sqrt{N}$ (e.g. 100 observation pairs should lead to an error of 0.1 to 0.2).

6. An analytical correlation model is then fitted to the discrete correlation estimations through a non-linear weighted least square curve fitting method based on the Levenberg-Marquardt algorithm. Formal errors (see point 5 above) are taken into account in the adjustment (weights). The correlation model follows the covariance model proposed by Arhan and Colin de Verdière (1985).

   $Cor(dx, dy) = (1/1+E) [1 + ar + (ar)^2/6 - (ar)^3/6] \exp(-ar)$ if $r \neq 0$

   $Cor(dx, dy) = 1$ if $r=0$, $r = [(dx/Lx)^2 + (dy/Ly)^2]^{1/2}$.

   Lx and Ly are the zonal and meridional scales (zonal and meridional zero crossings of the correlation function). E is the noise variance that represents both measurement and representativity errors. a is a constant equal to 3.337 calculated so that $[1 + a + a^2/6 - a^3/6] = 0$, i.e. Cor (dx,dy)=0 when r is equal to 1. This ensures that Lx and Ly scales correspond to zero crossing correlation scales. The fitting procedure provides estimation of Lx and Ly and of their formal errors. Another calculation of error ("standard fitting errors") is also carried out by using a least square fitting with unit weights to characterize the consistency of correlation estimations with our correlation model.

3 Sensitivity of results to sampling: a simulation study

To analyze the sensitivity of results to Argo sampling, a simulation study was performed. The main objective is to test the impact of realistic Argo sampling by using actual Argo float positions in 2005 and 2013 in the North Pacific. Over a 20°x20° box, 1600 Argo profiles were available in 2005 and about 2700 for 2013. A nominal 3°x3° Argo sampling would yield about 1800 profiles per year for a 20°x20° box. 2005 is thus close to a nominal Argo sampling and 2013 corresponds to an improved Argo sampling.

We generated 52 weekly (i.e. 1 year) simulated temperature 2D fields on a 20°x20° grid that follows the Arhan and Colin de Verdière (1985) covariance model. The 2D temperature fields were then sampled at the float positions in 2005 and 2013 and an observation noise of 10% was added (E=0.1). From these simulated Argo data, we analyzed how well covariance functions can be reconstructed following the method outlined in section 2. The calculations were done both for L=100 km and L= 400 km (isotropic field with L=Lx=Ly). Figures 1a and 1b show the estimated 2D correlation fields for the L=100 km simulation for the 2005 and 2013 sampling and the associated formal error. Figures 2a and 2b show the same results but for the L=400 km simulations. Table 1 summarizes the results for correlation scale and associated error estimations.

Results show that the estimations of correlation functions are highly sensitive to the Argo sampling. Typical error for a covariance or correlation value is about 0.25-0.4 for the 2005 sampling and 0.15-0.25 for the 2013 sampling over a one year time period. Correlation scales (assuming an a priori knowledge of the covariance function shape) can be determined with an accuracy of about 20 to 30 km

for 2005 and 10 to 20 km for 2013. These results are obviously dependent on the number of observation pairs available for a given spatial dx and dy lag. Correlation errors are also larger for the 400 km simulation because there are less independent observations of correlation.

These results show that one year of Argo observations over a 20°x20° box does not allow estimating precise enough correlation functions. When the sampling is improved as in 2013, results are, however, significantly improved. These results can easily be extrapolated to longer time series (and/or larger boxes) as correlation error RMS are proportional to the number of observation pairs at a given spatial lag (see equation in section 2). RMS errors for a four year time period will thus be divided by a factor of two. In that case, we expect errors on correlation of about 0.1 to 0.2 and an error on correlation scales below 10 km.

4 Results and discussion

A preliminary calculation of spatial scales (Lx and Ly which are also zero crossing of correlation functions) was carried out over several large areas (figure 3). Calculations were done both for temperature and salinity and all depths from the surface down to 1300 meters.

As an illustration, results for one box (box 3) in the North Pacific are shown on figures 4 and 5 for temperature at two different depths (200 and 1000 m). In that box, correlations are well estimated with a typical error below 0.1 due a large number of observation pairs N for a given zonal and meridional spatial lag (N comprised between 200 and 400). The estimated zonal and meridional correlation scales are 130 km and 110 km respectively. This is consistent with results derived from altimeter data analysis in mid latitude regions (e.g. Kurugano and Kamachi, 2000; Jacobs et al., 2001; Le Traon et al., 2003). Correlation scales are significantly larger at 1000 m (figure 5) and Lx and Ly are estimated to 185 km and 160 km respectively. Salinity scales (not shown) are very close to temperature ones although the estimation is slightly noisier.

Zonal and meridional spatial scales vary as expected with latitudes. Compared to mid latitude regions, scales are much larger in the tropical and equatorial regions. Figure 6 shows, for example, the correlation function for temperature at 200 m in the whole Equatorial Pacific. Zonal and meridional scales are estimated to about 900 km and 350 km. The zonal scales are smaller than those derived from TAO observations and larger than those derived from altimeter data (e.g. Kessler et al., 1996; Jacobs et al., 2001). This may be due to both the techniques used to compute scales (e.g. removing of large scale signals before computing altimeter spatial scales) and the sparse spatial sampling of TAO observations. As expected and well observed from altimetry and in-situ observations, there is a strong anisotropy with zonal scales two to three times larger than meridional scales. It is interesting to note that, compared to the Pacific ocean, smaller zonal scales are observed in the Indian (box 9 – zonal

scale of 780 km at 200 m for temperature) and Atlantic (box 8 - zonal scale of 360 km at 200 m for temperature) tropical/equatorial oceans.

They are also interesting variations of scales according to depth. Figures 7, 8 and 9 shows the vertical distribution of scales both for temperature and salinity for several areas (boxes 2, 9 and 18). At the surface or in the mixed layer, scales are much larger because they reflect large scale atmospheric forcing (heat flux, evaporation and precipitation). Note, however, that a mean seasonal cycle is removed priori the calculation. Below the mixed layer, scales are more representative of mesoscale dynamics and are consistent with scales derived from satellite altimetry. There is a general tendency (not systematic though) for an increase of temperature scales at depths larger than 800-1000 m although the correlation functions are noisier there because of lower signals. This may reflect a smaller influence of mesoscale variability at deeper depths but this should be investigated further.

There are significant differences between salinity and temperature scales (see figures 7, 8 and 9). At the surface and in the mixed layer where we observe large spatial scales, differences may reflect differences in scales between E-P (in particular precipitation) and heat flux forcing. At mid depth and depending on regions, differences may reflect the different dynamical nature of temperature and salinity signals. It is interesting to note, in particular, that the increase of scales for depths deeper than 800-1000 m for boxes 9 and 18 is not observed for salinity as it is for temperature. In many ocean regions (in particular tropics and subtropics) and for the deep ocean, temperature variations are more important than salinity in changing density. Temperature variations are thus more representative of ocean dynamics and salinity is more acting as a tracer of circulation. In 2D ocean turbulence, a tracer would exhibit smaller scales than density (or temperature) with a less steep wavenumber spectrum (e.g. Vallis, 2006). Although we tend to observe smaller salinity scales, this should be analyzed further and globally.

A similar calculation was done by Resnyanskii et al. (2010) but with a more limited Argo data set (2005-2007). Our results are in a qualitative agreement with theirs although they found larger scales. This may be due to the differences in data sets but also to differences in the way spatial scales were computed. They did not remove, in particular, biases in the Levitus climatology and did not adjust a covariance model taking into account noise level.

5 Conclusions and perspectives

This study was a first attempt to estimate spatial scales of temperature and salinity at different depths from the Argo global ocean observing system. A careful error analysis was carried out and it shows that several years of Argo observations are required for a precise enough (error on correlation below 0.1 to 0.2) estimation of correlation functions over 20°x20° boxes. Correlation functions and associated zonal and meridional spatial scales were then calculated over several large areas over the

global ocean. Scales vary from 350 to 900 km in the equatorial regions down to less than 100 km at high latitudes. Zonal and meridional scales are similar except in the Pacific and Indian tropical/equatorial regions where zonal scales are much larger (by a factor of 2 to 3) than meridional scales. These results are consistent with previous studies based on sparse in-situ observations or satellite altimetry but they allow for the first time a global characterization, an analysis of differences of scales between temperature and salinity and the variations of scales according to depths. As the Argo array develops, more precise and/or higher resolution estimations can be derived. We plan to extend soon these analyses to time and space/time correlation estimations. Similar calculations will also be applied to characterize global eddy resolving model errors, i.e. instead of analyzing Argo observations minus climatology analyzing Argo observations minus a model guess. This is essential to improve data assimilation systems.

Acknowledgments

These data were collected and made freely available by the International Argo Program and the national programs that contribute to it. (http://www.argo.ucsd.edu, http://argo.jcommops.org). The Argo Program is part of the Global Ocean Observing System.

References

Arhan, M. and Colin De Verdière, A.: Dynamics of Eddy Motions in the Eastern North Atlantic. J. Phys. Oceanogr., 15, 153–170, 1985.

Bendat, and Piersol, A.G.: Random data analysis and measurement procedures. J.J. Wiley-Interscience Publication John Wiley and Sons, New York 1986, 566 pp., 1986.

Cabanes, C., Grouazel, A., Von Schuckmann, K., Hammon, M., Turpin, V., Coatanoan, C., Paris, F., Guinehut, S., Boone, C., Ferry, N., de Boyer Montégut, C., Carval, T., Reverdin, G., Pouliquen, S., and Le Traon, P.Y.: The CORA dataset : validation and diagnostics of in-situ ocean temperature and salinity measurements. Ocean Science, 9, 1-18, doi: 10.5194/os-9-1-2013, 2013.

Dong, C., McWilliams, J.C., Liu, Y., and Chen, D.: Global heat and salt transports by eddy movement, Nat Commun, 5, http://dx.doi.org/10.1038/ncomms4294, 2014.

Durack, P.J., and Wijffels, S.E.: Fifty-year trends in global ocean salinities and their relationship to broad-scale warming. J. Climate, 23, 4342–4362, 2010.

Freeland, H. J., Roemmich D., Garzoli, S.L, Le Traon, P.-Y., Ravichandran, M., Riser, S., Thierry, V., Wijffels, S., Belbeoch, M., Gould, J. , Grant, .F, Ignazewski, .M, King, B., Klein, B., Mork, K.A., Owens, B., Pouliquen, S., Sterl, A., Suga, .T, Suk, M.-S., Sutton, P., Troisi, A., Velez-Belchi, P. J. and

Xu, J.: Argo - A Decade of Progress. Proceedings of OceanObs'09: Sustained Ocean Observations and Information for Society (Vol. 2), Venice, Italy, 21-25 September 2009, Hall, J., Harrison D.E. & Stammer, D., Eds., ESA Publication WPP-306.

Gaillard F., E. Autret, V. Thierry, P. Galaup, C. Coatanoan, and T. Loubrieu, 2009. Quality Control of Large Argo Datasets. J. Atmos. Oceanic Technol., 26, 337–351. doi: http://dx.doi.org/10.1175/2008JTECHO552.1, 2010.

Guinehut, S., Dhomps, A.-L., Larnicol, G., and P.Y. Le Traon: High resolution 3-D temperature and salinity fields derived from in situ and satellite observations, Ocean Sci., 8, 845-857, doi:10.5194/os-8-845-2012, 2012.

Jacobs, G. A., Barron, C. N., and Rhodes, R. C.: Mesoscale characteristics, J. Geophys. Res., 106, 19,581–19,595, doi:10.1029/2000JC000669, 2001.

Kessler, W.S., Spillane, M. C., McPhaden, M. J., and Harrison, D. E: Scales of Variability in the Equatorial Pacific Inferred form Tropical Atmosphere-Ocean Buoy Array. J. Climate, 9, 2999–3024. doi: http://dx.doi.org/10.1175/1520-0442, 1996.

Kuragano, T., and Kamachi, M.: Global statistical space-time scales of oceanic variability estimated from the TOPEX/POSEIDON altimeter data, J. Geophys. Res., 105, 955–974, doi:10.1029/1999JC900247, 2000.

Le Traon, P.Y., Faugère, Y., Hernandez, F., Dorandeu, J, Mertz, F, and Ablain, M.: Can we merge GEOSAT Follow-On with TOPEX/Poseidon and ERS-2 for an improved description of the ocean circulation? Journal Of Atmospheric And Oceanic Technology, 20(6), 889-895. http://dx.doi.org/10.1175/1520-0426, 2003.

Le Traon, P.Y.: From satellite altimetry to Argo and operational oceanography: three revolutions in oceanography. Ocean Science, 9(5), 901-915. Publisher's official version: http://dx.doi.org/10.5194/os-9-901-2013, 2013.

Oke, P. R., Balmaseda, M. A., Benkiran, M., Cummings, J. A., Fujii, Y., Guinehut, S., Larnicol, G., Le Traon, P.-Y., Martin, M. J., and Dombrowsky, E.: Observing System Evaluations using GODAE systems. Oceanography, 22(3), 144-153, 2009.

Resnyanskii Y.D., M.D. Tsyrulnikov, B.S. Strukov, and A.A. Zelenko: Statistical structure of spatial variability of the ocean thermohaline fields from Argo profiling data, 2005-2007. Oceanology, Vol. 50, No. 2, 149-165. DOI: 10.1134/S0001437010020013, 2010.

Roemmich, D., Gilson, J., Davis, R., Sutton, P., Wijffels, S., and Riser, S.: Decadal Spinup of the South Pacific Subtropical Gyre. J. Phys. Oceanogr., 37, 162–173. doi: http://dx.doi.org/10.1175/JPO3004.1, 2007.

Roemmich, D. and Gilson, J.: The 2004–2008 mean and annual cycle of temperature, salinity, and steric height in the global ocean from the Argo Program, Progress in Oceanography, Volume 82, Issue 2,Pages 81-100, ISSN 0079-6611, http://dx.doi.org/10.1016/j.pocean.2009.03.004, 2009.

Roemmich, D., and the Argo Steering Team: Argo: the challenge of continuing 10 years of progress. Oceanography, 22(3), 46-55, doi:10.5670/oceanog.2009.65., 2009.

Vallis, G. K.: Atmospheric and Oceanic Fluid Dynamics. Cambridge University Press, 745 pp., 2006.

von Schuckmann, K. and Le Traon, P. Y.: How well can we derive Global Ocean Indicators from Argo data ? Ocean Science, 7(6), 783-791. Publisher's official version: http://dx.doi.org/10.5194/os-7-783-2011, 2011.

Table 1: Simulation of the impact of Argo sampling on the estimation of correlation functions.

Simulation type	Estimated L	Associated error (1 σ)
L=100 km – 2005 sampling	70 km	20 km
L=100 km – 2013 sampling	89 km	11 km
L=400 km – 2005 sampling	372 km	30 km
L=400 km – 2013 sampling	418 km	20 km

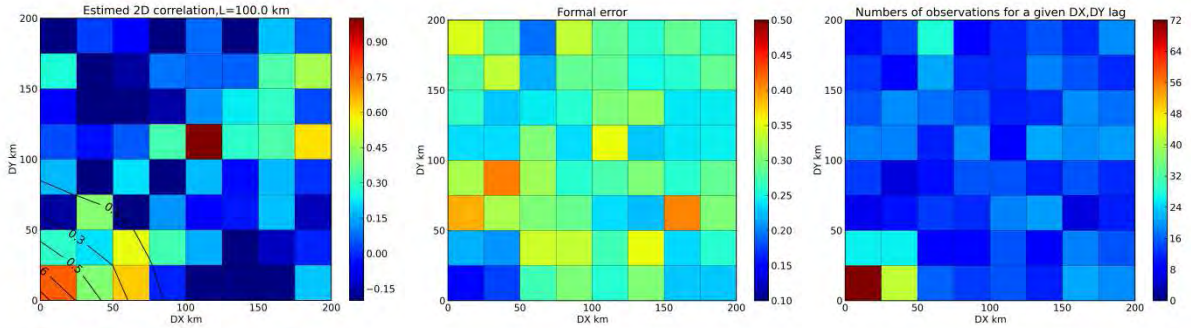


Figure 1a: Estimated 2D covariance fields for the L=100 km simulation for the 2005 Argo sampling (left), associated formal errors (middle) and number of observation pairs (right).

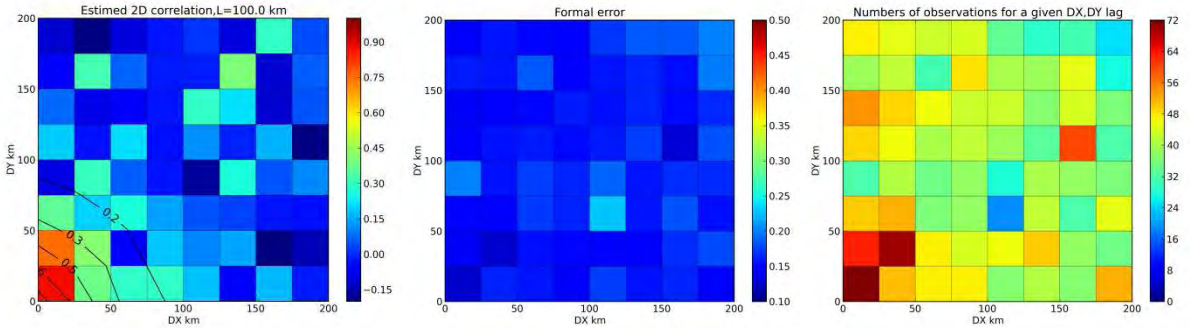


Figure 1b: Estimated 2D covariance fields for the L=100 km simulation for the 2013 Argo sampling (left), associated formal errors (middle) and number of observation pairs (right).

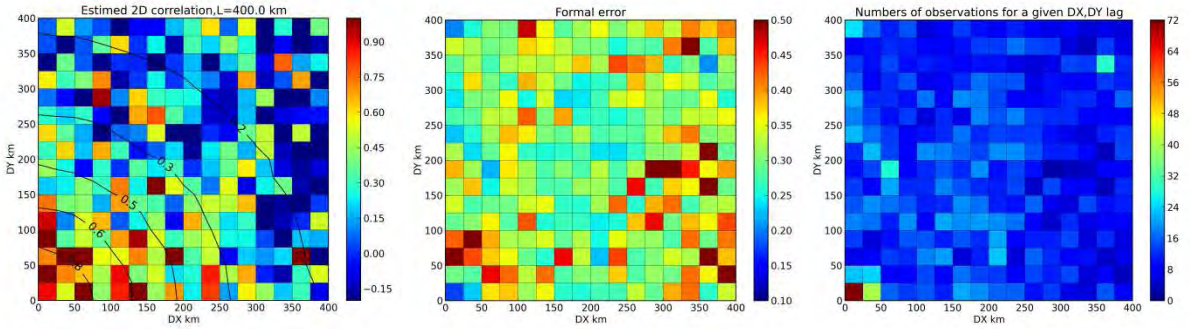


Figure 2a: Estimated 2D covariance fields for the L=400 km simulation for the 2005 Argo sampling (left), associated formal errors (middle) and number of observation pairs (right).

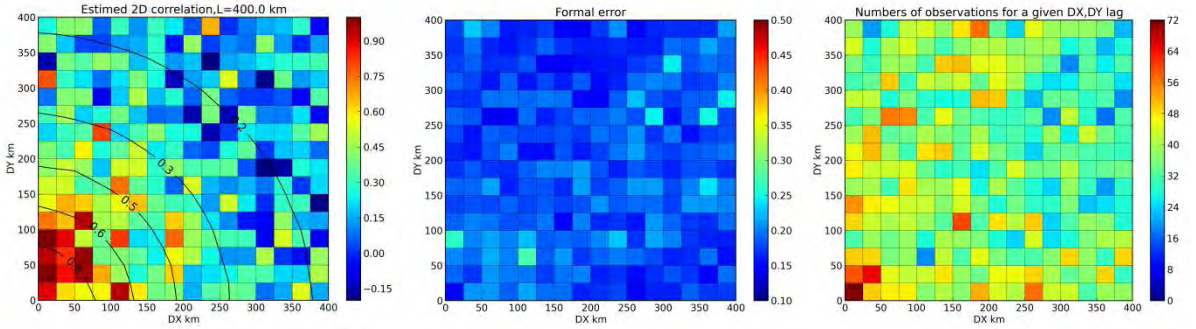


Figure 2b: Estimated 2D covariance fields for the L=400 km simulation for the 2013 Argo sampling (left), associated formal errors (middle) and number of observation pairs (right). Black isolines (left) correspond to the adjusted covariance model.

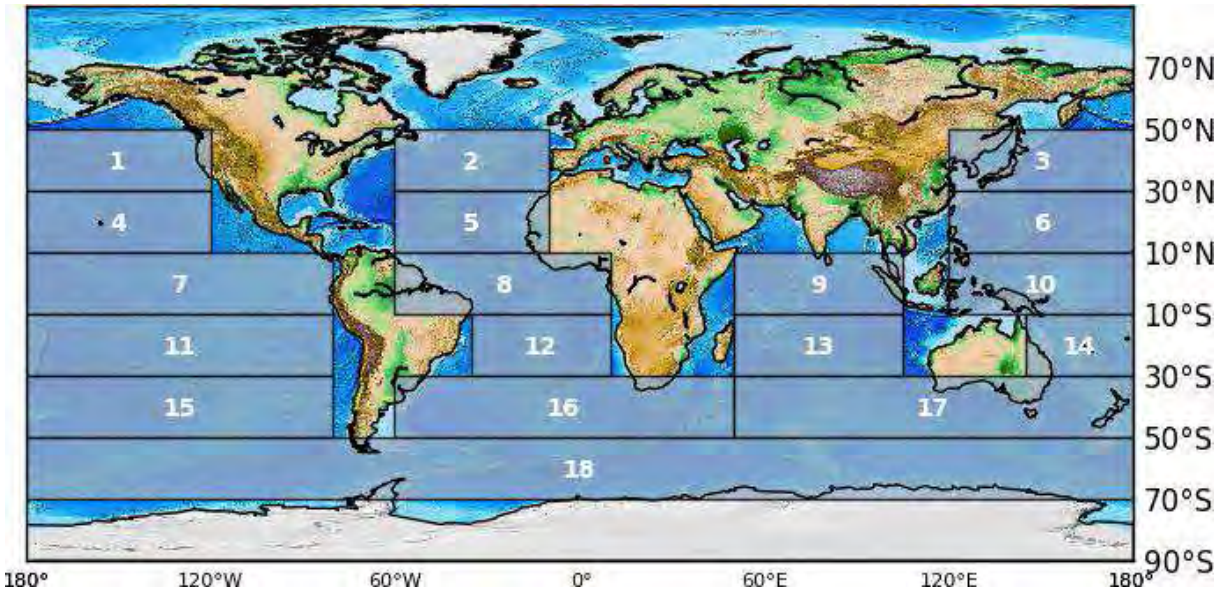


Figure 3: Large scale areas where temperature and salinity spatial correlations were calculated.

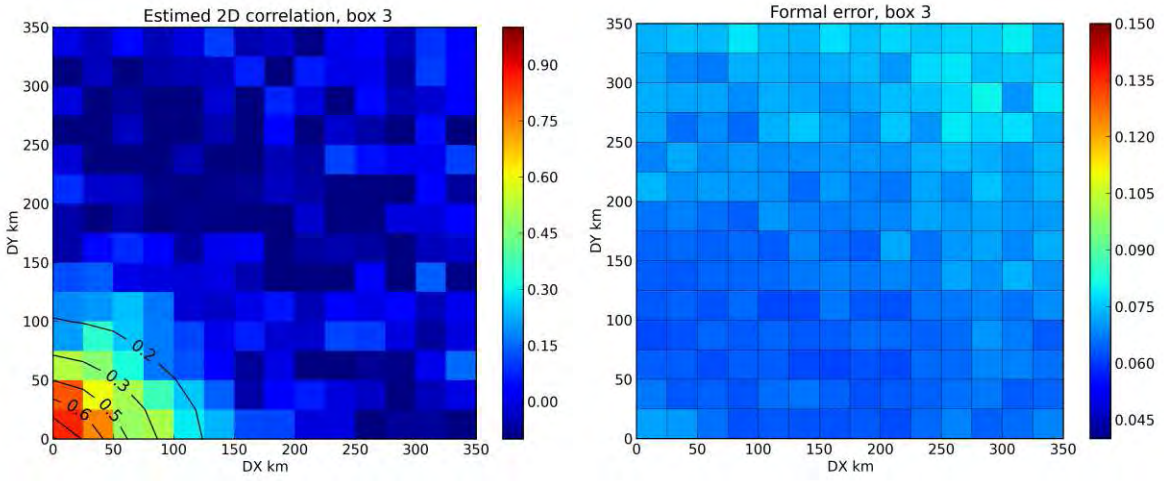


Figure 4: 2D covariance calculated in a 20°x60° area (box number 3) in the North Pacific Ocean (left) and associated formal error (right) for temperature at 200 m. Black isolines (left) correspond to the adjusted covariance model.

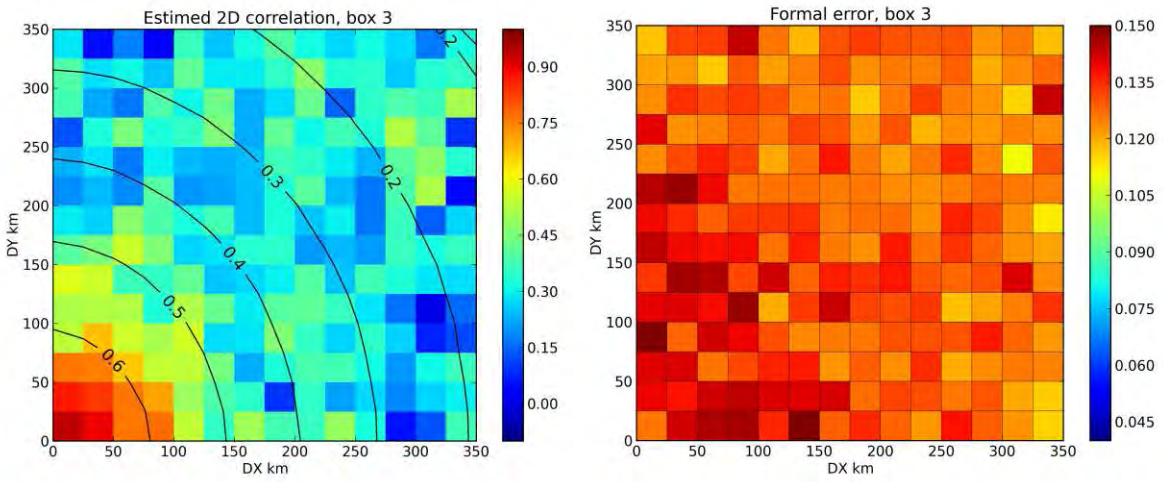


Figure 5: 2D covariance calculated in a 20°x60° area (box number 3) in the North Pacific Ocean (left) and associated formal error (right) for temperature at 1000 m. Black isolines (left) correspond to the adjusted covariance model.

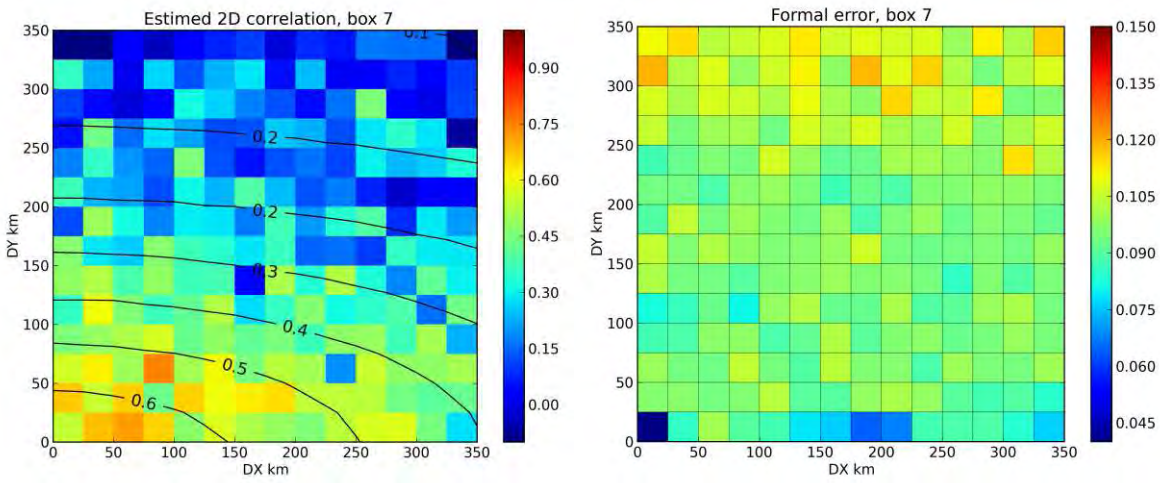


Figure 6: 2D covariance calculated in a 100°x20° area (box number 7) in the Equatorial Pacific Ocean (left) and associated formal error (right) for temperature at 200 m. Black isolines correspond to the adjusted covariance model.

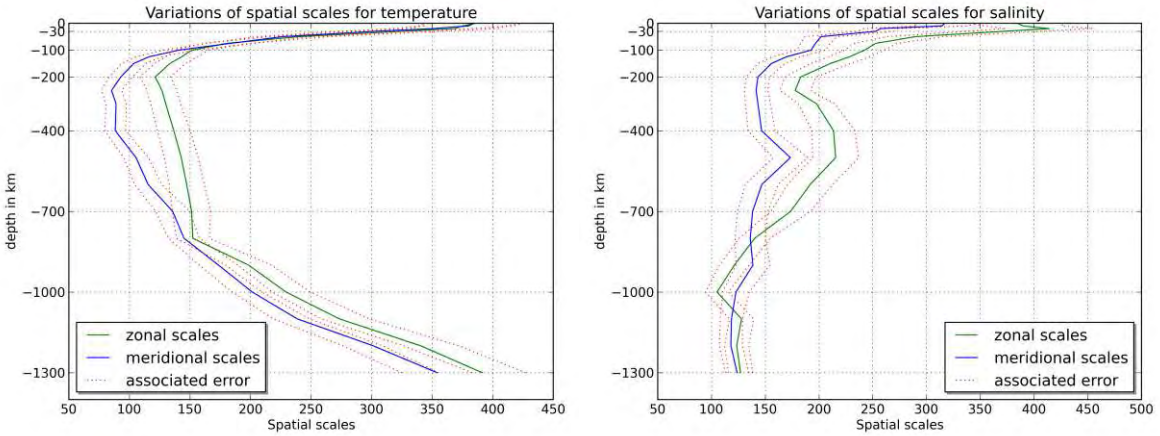


Figure 7: Variations of zonal and meridional spatial scales for temperature (left) and salinity (right) according to depth for box 18 (High latitude Southern Hemisphere). Dotted lines represent standard fitting errors.

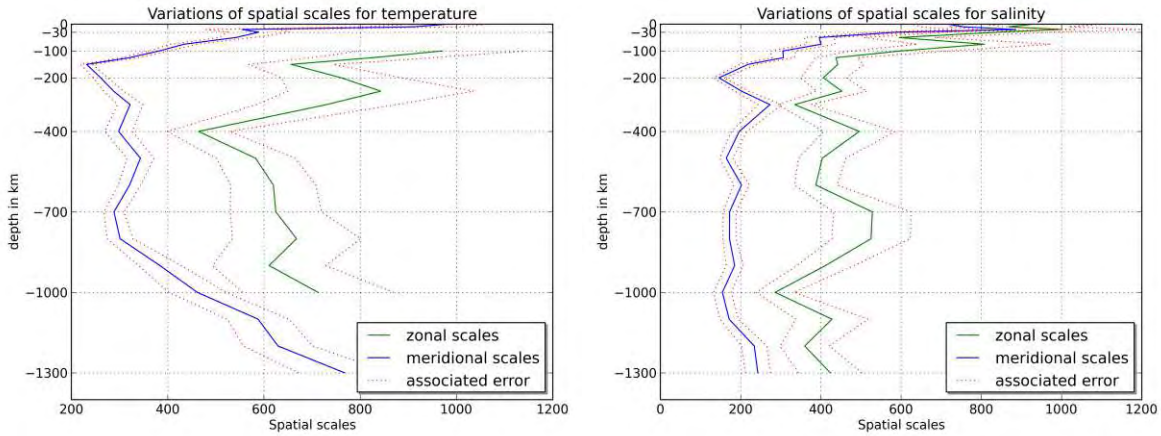


Figure 8: Variations of zonal and meridional spatial scales for temperature (left) and salinity (right) according to depth for box 9 (Equatorial Indian Pacific). Dotted lines represent standard fitting errors.

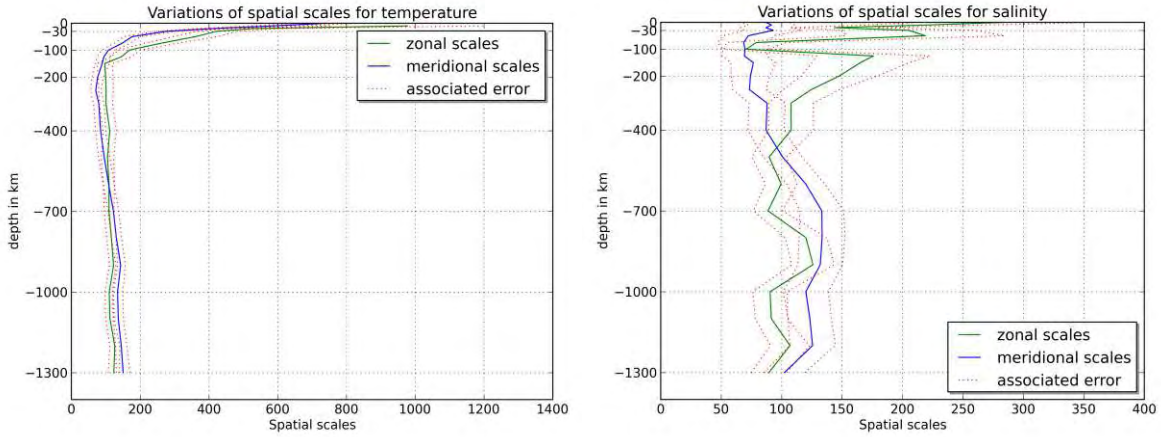


Figure 9: Variations of zonal and meridional spatial scales for temperature (left) and salinity (right) according to depth for box 2 (mid latitude North Atlantic). Dotted lines represent standard fitting errors.

4.4.2 Erreur formelle sur l'estimation de la covariance

En complément de la démarche abordée dans l'article (section 4.4.1), nous allons ici détailler les étapes permettant de définir l'erreur formelle sur l'estimation de la covariance.

Soit par définition l'estimation de la covariance notée $\mathbf{cov_{est}}$ et pour une variable z centrée ($\mathbf{E}(z)= 0$), on a :

$$\begin{aligned} \mathbf{cov}_{est}(\triangle x, \triangle y) &= \mathbf{E}(zz') \\ &= \frac{1}{N}\sum_{i=1}^{N} z_i z_i' \quad tel \quad que \quad z, z' \in \mathbf{C}_{(\triangle x, \triangle y)} \end{aligned} \tag{4.11}$$

On peut exprimer la variance de l'erreur faite sur l'estimation telle que :

$$\mathbf{Var}_{err} = \mathbf{E}[\mathbf{cov}_{est}(\triangle x, \triangle y) - \mathbf{cov}(\triangle x, \triangle y)]^2 \tag{4.12}$$

$$\tag{4.13}$$

On a exprimé la covariance connue **cov**.
Exprimons la quantité :

$$\begin{aligned} &[\mathbf{cov}_{est} - \mathbf{cov}]^2 \\ &\qquad = [\frac{1}{N}\sum_{i=1}^{N} z_i z_i' - \mathbf{cov}]^2 \\ &\qquad = \frac{1}{N^2}\sum_{i=1}^{N}\sum_{r=1}^{N}(z_i z_i' z_r z_r' - z_i z_i' \mathbf{cov} - z_r z_r' \mathbf{cov} + \mathbf{cov}^2) \end{aligned} \tag{4.14}$$

On applique l'opérateur linéaire $\mathbf{E}[\bullet] = \overline{\bullet}$

$$\begin{aligned} \mathbf{Var}_{err} &= \frac{1}{N^2}\sum_{i=1}^{N}\sum_{r=1}^{N}(\overline{z_i z_i' z_r z_r'} - 2\mathbf{cov}^2 + \mathbf{cov}^2) \\ &= \frac{1}{N^2}\sum_{i=1}^{N}\sum_{r=1}^{N}(\overline{z_i z_i' z_r z_r'} - \mathbf{cov}^2) \end{aligned} \tag{4.15}$$

D'après *Bendat and Piersol* [1986] :

On considère N variables aléatoires, x_i pour $i \in [0, N]$ qui sont éventuellement correlés. On note respectivement pour moyenne, variance et covariance les quantités suivantes :

$$\begin{aligned} \mu_i &= \mathbf{E}[x_i] \\ \sigma_i^2 &= \mathbf{E}[(x_i - \mu_i)^2] \\ C_{ij} &= \mathbf{E}[(x_i - \mu_i)(x_j - \mu_j)] \\ &= \mathbf{E}[(x_i x_j)] - \mathbf{E}[x_i].\mathbf{E}[x_j] \\ \text{et} \quad C_{ii} &= \sigma_i^2 \end{aligned} \tag{4.16}$$

D'après le théorème d'Isserlis [4] (*Bendat and Piersol* [1986]), on exprime le moment du 4^{eme} ordre comme la somme de différentes paires de moments du 2^{eme} ordre .

$$\mathbf{E}[x_1 x_2 x_3 x_4] = C_{12} C_{34} + C_{13} C_{24} + C_{14} C_{23} \tag{4.17}$$

Alors si on considère

$$\begin{aligned} \overline{z_i z_i' z_r z_r'} &= \mathbf{E}[z_i z_i' z_r z_r'] \\ &= \overline{z_i z_i'}.\overline{z_r z_r'} + \overline{z_i z_r}.\overline{z_i' z_r'} + \overline{z_i z_r'}.\overline{z_i' z_r} \\ &= \mathbf{cov}^2(\triangle x, \triangle y) + \mathbf{cov}^2(\delta x, \delta y) + \mathbf{cov}(\triangle x + \delta x, \triangle y + \delta y). \\ &\quad \mathbf{cov}(\delta x - \triangle x, \delta y - \triangle y) \end{aligned} \tag{4.18}$$

L'erreur sur l'estimation repose sur les distances effectives $(\delta x, \delta y)$ des points présents dans une même classe, en l'occurence $\mathbf{C}_{(\triangle x, \triangle y)}$.
Par conséquent :

$$\begin{aligned} \mathbf{Var}_{err} &= \frac{1}{N^2} \sum_{i=1}^{N} \sum_{r=1}^{N} (\overline{z_i z_i' z_r z_r'} - \mathbf{cov}^2(\triangle x, \triangle y)) \\ &= \frac{1}{N^2} \sum_{i,r} [\mathbf{cov}^2(\delta x, \delta y) + \\ &\quad \mathbf{cov}(\triangle x + \delta x, \triangle y + \delta y).\mathbf{cov}(\triangle x - \delta x, \triangle y - \delta y)] \end{aligned} \tag{4.19}$$

Pour un cas simple, on retrouve une expression courante de la variance de l'erreur. Si $\triangle x' = 0, \triangle y' = 0$ et si $i = r$

$$\begin{aligned} \mathbf{Var}_{err} &= \frac{1}{N^2} N[2\mathbf{cov}^2] \\ &= \frac{2\sigma^4}{N} \end{aligned} \tag{4.20}$$

4. ce théorème permet d'exprimer les moments d'ordre supérieur de distributions normales multivariées en fonction des matrices de covariance.

4.5 Conclusion

Dans ce chapitre nous avons estimé les fonctions de corrélations puis les échelles spatiales zonales et méridionales de la température et de la salinité à différentes profondeurs des erreurs modèles ([observations Argo-run libre du modèle au 1/4°]) et du signal océanique ([observations Argo-climatologie mensuelle Levitus]) sur plusieurs zones grande échelle recouvrant l'ensemble du globe. Les échelles spatiales estimées en température et salinité révélent une tendance décroissante vers les pôles et croissante en fonction de la profondeur. Elles varient de plus de 500 km aux latitudes tropicales jusqu'à 100 km aux hautes latitudes. Les échelles zonales et méridionales sont similaires à l'exception des latitudes tropicales. On observe un facteur de 2 à 3 entre les échelles zonales et méridionales. Nous avons aussi constaté des différences significatives des échelles spatiales en température et en salinité pouvant refléter la nature dynamique différente de ces deux paramètres fondamentaux caractérisant l'océan. Enfin nous avons pu, grâçe aux observations Argo, approfondir le comportement de la simulation $T335$ dans l'océan profond, ainsi que sa capacité à reproduire les échelles de la variabilité océanique. Nous avons constaté des différences notables entre les échelles spatiales des erreurs modèles et du signal (issues des anomalies par rapport à la climatologie). Notamment en surface dans le cas des échelles spatiales en température, le modèle représente mieux la variabilité à grande échelle. Enfin en profondeur les échelles spatiales des erreurs modèles différent également significativement de celles du signal océanique.

Chapitre 5

Analyse des méthodes basées sur le contenu en information et étude de sensibilité

Sommaire

5.1 Introduction 91
5.2 Mesure d'impact des observations sur l'analyse 92
5.2.1 Définition analytique 92
5.2.2 Définition dans le cadre du BLUE 92
5.2.3 Definition dans le cadre de l'analyse objective 93
5.2.4 Méthodes alternatives pour l'approximation du DFS . . 94
5.3 Etude de sensibilité du contenu en information 96
5.3.1 Simulation de distributions gaussiennes et estimateur de Gauss-Markov 97
5.3.2 Tests d'influence sur le DFS 98
5.4 Comparaison des estimations du contenu en information 105
5.4.1 Illustration à partir d'une application au système d'assimilation de Mercator Océan 105
5.4.2 Mise en pratique dans le cadre d'analyse objective . . . 106
5.5 Conclusion 109

5.1 Introduction

Au cœur d'un système d'analyse et de prévision océaniques se trouve un modèle mathématique décrivant l'océan, basé sur des équations complexes tenant compte des phénomènes physiques et de leur évolution dans le temps. Afin que la simulation numérique de l'océan soit le plus réaliste possible, les modèles numériques océaniques doivent recevoir des données indispensables dont la plupart sont issues de sources satellitaires et in-situ (température, salinité, niveau de surface et vitesses de courants...). Ces mesures sont souvent de qualité variable, et disposées de manière irrégulière dans l'espace et/ou le temps. La représentation spatio-temporelle des processus océaniques faite par le modèle est aussi imparfaite. Le modèle souffre de plusieurs sources d'erreur notamment celle liée aux équations du modèle (discrétisation, paramétrisations des équations de la physique), aux forcages et aux conditions aux limites. Il est donc crucial de traiter chacune des informations provenant du modèle ou des observations au sein d'un système d'assimilation en tenant compte de l'erreur dont l'information est entachée. Pour cela l'assimilation de données reposant sur un processus mathématique permet d'obtenir la représentation la plus probable de l'état d'un système à partir de toutes les sources d'informations disponibles et des contraintes connues (*Bouttier* [2004]). En effet l'analyse résultant d'un système d'assimilation est influencée à la fois par les observations récemment assimilées et par l'information portée par l'ébauche. Il existe différentes approches pour mesurer l'impact des observations assimilées dont l'une d'elles repose sur le contenu en information, noté par la suite DFS pour Degrees of Freedom of the System. Des méthodes d'estimation du DFS ont déjà été développé pour des applications liées à l'atmosphère (*Chapnik et al.* [2006], *Cardinali et al.* [2004]) puis plus récemment lors d'applications océanographiques (*Oke et al.* [2009], *Dibarboure et al.* [2011]). Le DFS permet d'évaluer la contribution relative des observations dans n'importe quel système d'assimilation de données. Les méthodes d'estimation du contenu en information dans un système d'analyse et de prévisions réalistes s'articulent autour de l'expression de matrices de grandes dimensions. Les calculs matriciels peuvent s'avérer longs et coûteux. Pour éviter l'explicitation directe de matrices si denses, *Girard* [1987], *Desroziers and Ivanov* [2001] puis *Chapnik et al.* [2006] ont proposé de nouveaux schémas d'implémentation afin d'estimer la trace de ces matrices lors des calculs pratiques du DFS. Enfin *Lupu and Gauthier* [2010] décrit une méthode pratique afin d'estimer le DFS de manière plus directe. Nous allons comparer ces différentes approches d'estimation du contenu en information afin de déterminer laquelle conviendrait le mieux à une implémentation dans le système d'assimilation de Mercator Océan. Ainsi dans ce chapitre nous présentons la définition analytique du DFS et ses formulations dans le cadre du BLUE et de l'analyse objective en section 5.2, puis en sections 5.3 et 5.4 par le biais d'applications nous mènerons une étude de sensibilité sur le DFS,

et la mise en pratique de plusieurs méthodes alternatives (définies en section 5.2.4) afin d'évaluer approximativement le DFS dans un cadre d'analyse objective.

5.2 Mesure d'impact des observations sur l'analyse

5.2.1 Définition analytique

On considère un schéma d'assimilation de donnée, soit x_a le vecteur d'analyse optimale résultant, alors on définit le DFS comme étant la trace de la jacobienne du vecteur d'analyse dans l'espace des observations par rapport aux observations :

$$\mathsf{DFS} = \mathsf{Tr}\{\frac{\partial(\mathbf{H}x_a)}{\partial \mathbf{y}}\} \tag{5.1}$$

Où $\mathbf{H}$ est l'opérateur d'observations linéarisé au voisinage de l'ébauche, et $\mathbf{y}$ le vecteur d'observations. Le DFS ou degrès de liberté du signal, permet de quantifier l'information utile contenue dans les observations en ce sens qu'il caractérise comment le système d'assimilation utilise les observations pour construire le signal à partir de l'ébauche (*Rabier* [2005]).

5.2.2 Définition dans le cadre du BLUE

Rappellons brièvement l'approche du BLUE qui permet d'obtenir le meilleur estimateur non biaisé du problème linéaire de l'assimilation. En pratique, il s'agit de trouver la meilleure combinaison entre un état à priori du système que l'on appelle l'ébauche, et les observations (*Gelb* [1974], *Bouttier and Courtier* [1999]). Le critère d'optimalité pour obtenir le meilleur estimateur statistique de cette combinaison est la détermination du minimum de variance d'erreur d'estimation. Les informations issues de l'ébauche et des observations sont entachées d'erreur. La combinaison des deux par le biais de divers algorithmes d'assimilation (BLUE, filtre de Kalman notamment) permet de se rapprocher de l'état réel du système. Ainsi, l'algorithme d'assimilation identifie à un instant t donné, un état analysé $\mathbf{x}_a$ de la variable aléatoire $\mathbf{x}$ dépendant linéairement de l'ébauche $\mathbf{x}_b$ et des observations $\mathbf{y}^o$ (les étapes de calculs sont détaillées en annexe C). On écrit l'état analysé $\mathbf{x}_a$ tel que :

$$\mathbf{x}_a = \mathbf{x}_b + \mathbf{K}\left(\mathbf{y}^o - \mathbf{H}\mathbf{x}_b\right) = \mathbf{x}_b + \mathbf{B}\mathbf{H}^T(\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R})^{-1}\left(\mathbf{y}^o - \mathbf{H}\mathbf{x}_b\right) \tag{5.2}$$

La matrice $\mathbf{K}$, appelée matrice de gain, permet de définir l'état assimilé comme la somme de l'ébauche et de l'incrément d'analyse (terme correctif)

$$\delta \mathbf{x}_a = \mathbf{K}\left(\mathbf{y}^o - \mathbf{H}\mathbf{x}_b\right) = \mathbf{K}d \tag{5.3}$$

avec d le vecteur d'innovation qui représente la mesure résiduelle dans l'espace des observations.

Ainsi la définition analytique du DFS s'obtient en dérivant (5.2) dans l'espace des observations, et (5.1) devient :

$$\frac{\partial(\mathbf{H}x_a)}{\partial \mathbf{y}} = \frac{\partial\left(\mathbf{H}x_b + \mathbf{H}\mathbf{K}\left(\mathbf{y}^o - \mathbf{H}x_b\right)\right)}{\partial \mathbf{y}} \tag{5.4}$$

$$\mathsf{DFS} = \mathsf{Tr}\{(\mathbf{HK})\} \tag{5.5}$$

L'expression (5.5) relie le DFS à la trace de la matrice de gain (Kalman).

5.2.3 Definition dans le cadre de l'analyse objective

Dans le cadre simple d'une analyse objective (*Bretherton et al.* [1976]; *Gandin* [1963]), le vecteur d'ébauche peut ne pas être utilisé et être remplacé par les valeurs d'observations issues de climatologies. On cherche le meilleur estimateur linéaire au sens des moindres carrés non biaisé de $\mathbf{x}_a$, noté $\widehat{x_a}$. Le nouvel état estimé est basé seulement sur les observations de telle sorte que

$$\widehat{x_a} = \sum \alpha_{ij}\mathbf{y} \tag{5.6}$$

D'après le théorème de Gauss-Markov on obtient :

$$\widehat{x_a} = \mathbf{CdgCdd^{-1}y}$$

Enfin on applique la définition analytique du DFS dans le cadre d'une analyse objective.

$$\mathsf{DFS} = \mathsf{Tr}\{\mathbf{HCdgCdd^{-1}}\} \tag{5.7}$$

On note **Cdg** [1] la matrice d'erreurs de covariance champs-observations, **Cdd** [2] la matrice d'erreurs de covariance au sein des observations, **H** est l'opérateur d'observations.

1. data-grid
2. data-data

5.2.4 Méthodes alternatives pour l'approximation du DFS

L'information contenue dans les différents types de jeux d'observations ne possède pas le même impact sur l'analyse. Dans la littérature, *Cardinali et al.* [2004] propose une méthode permettant de distinguer l'apport respectif des différentes observations sur l'analyse. Cette approche permet également de mesurer l'influence partielle d'un sous-ensemble d'observations restreint à une zone géographique donnée. De plus *Chapnik et al.* [2006] aborde une approche similaire basée sur l'estimation précise de la trace de la matrice d'influence [3]. Cette approche évoque l'ajout de perturbations aléatoires sur les observations. L'objectif est, connaissant l'incertitude sur les observations, de mesurer la sensibilité de l'analyse par rapport aux observations perturbées (méthode de randomisation de *Girard* [1987]). Enfin une approche réalisée plus récemment repose sur l'utilisation des vecteurs d'incréments et d'innovations (notés respectivement $\mathbf{d}_b^o$ et $\mathbf{d}_b^a$). A partir de ces quantités il est possible de diagnostiquer les erreurs d'observations, d'ébauche et d'analyse à posteriori, et d'obtenir une expression du DFS (*Desroziers et al.* [2005]; *Lupu and Gauthier* [2010]). En se basant sur la formulation analytique du DFS (notre référence) nous allons confronter deux des approches citées ci-dessus : la méthode reposant sur la randomisation de Girard développée dans *Girard* [1987] et *Chapnik et al.* [2006], et la méthode dite aposteriori abordée dans *Lupu and Gauthier* [2010]. L'expression analytique du DFS ainsi que les deux approches d'estimation comparées sont détaillés ci-dessous et mises en pratique dans le cadre d'une analyse objective à la section 5.4.

Approximation du DFS selon la méthode de perturbations de Girard

En utilisant l'approche introduite dans *Desroziers and Chapnik* [2001], puis utilisée dans *Chapnik et al.* [2006], on montre alors qu'une approximation du DFS peut s'exprimer comme suit :

$$\mathsf{DFS}_{\mathsf{Girard}} = (\mathbf{y}^* - \mathbf{y})^t \mathbf{R}^{-1} (\mathbf{H}x_a^* - \mathbf{H}x_a) \tag{5.8}$$

Les observations sont obtenues en ajoutant des petites perturbations telles que

$$\mathbf{y}^* = \mathbf{y} + \mathbf{R}^{1/2}\varsigma \tag{5.9}$$

où $\varsigma \sim \mathsf{N}(0, I_p)$. Enfin $\mathbf{H}x_a^*$, $\mathbf{H}x_a$ sont les analyses dans l'espace des observations obtenues respectivement à partir des observations perturbées et non perturbées.

formulation dans le cadre de l'analyse objective De la même manière on se place dans le cadre de l'analyse objective, alors l'approximation du DFS s'écrit comme suit,

3. ou aussi matrice de gain définie à la section 5.2

$$\mathsf{DFS}_{\mathsf{Girard}} = (\mathbf{y}^* - \mathbf{y})^t \mathbf{R}^{-1} (\mathbf{H}\widehat{x_a}^* - \mathbf{H}\widehat{x_a}) \tag{5.10}$$

Approximation du DFS selon la méthode dite aposteriori

Un ensemble de diagnostics dans l'espace des observations basé sur les vecteurs d'innovation, d'incrément a été proposé par *Desroziers et al.* [2005]. A partir de ces quantités, on montre alors qu'il est possible de diagnostiquer les erreurs statistiques d'observations, d'ébauche, et d'analyse a posteriori. Par conséquent, on note $\mathbf{B}$ et $\mathbf{R}$ les matrices de covariances d'erreurs diagnostiquées respectivement d'ébauche et d'observations telles que :

$$\mathbf{E}[\mathbf{d}_b^a \mathbf{d}_b^o] = \mathbf{H}\mathbf{B}\mathbf{H}^t \tag{5.11}$$

$$\mathbf{E}[\mathbf{d}_b^o \mathbf{d}_b^o] = \mathbf{R} \tag{5.12}$$

De plus *Lupu and Gauthier* [2010] souligne aussi qu'une expression du DFS peut être obtenue via les vecteurs :

$$\mathbf{d}_b^a = \mathbf{H}(x_a) - \mathbf{H}(x_b) \approx \mathbf{H}\delta x_a = \mathbf{H}\mathbf{K}\mathbf{d}_o^b,$$
$$\mathbf{d}_a^o = \mathbf{y} - \mathbf{H}(x_b + \delta x_a) \approx (\mathbf{I} - \mathbf{H}\mathbf{K})\mathbf{d}_o^b = \mathbf{R}\mathbf{D}^{-1}\mathbf{d}_o^b$$

Où $\mathbf{K}$ est la matrice de gain de Kalman, et $\mathbf{D} = \mathbf{H}\mathbf{B}\mathbf{H}^t + \mathbf{R}$ est la matrice de covariance d'innovation. De la démonstration faite par *Lupu and Gauthier* [2010], il vient

$$\mathbf{E}[\mathbf{d}_b^{at} \mathbf{R}^{-1} \mathbf{d}_a^o] = \mathbf{E}[\mathsf{Tr}\{\mathbf{D}^{-1}\mathbf{D}^e\mathbf{K}^t\mathbf{H}^t\}] \tag{5.13}$$

Il se présente alors deux cas de figures :

– La matrice de covariance d'innovation $\mathbf{D}$ prescrite dans l'assimilation coïncide avec celle estimée $\mathbf{D}^e$, dans ce cas le DFS s'écrit :

$$\mathsf{DFS}_{\mathsf{Apost}} = \mathsf{Tr}(\mathbf{H}\mathbf{K}) = \mathbf{E}[(\mathbf{d}_b^a)^t \mathbf{R}^{-1} \mathbf{d}_a^o] \tag{5.14}$$

L'expression (5.14) donne un moyen efficace d'estimer le DFS pour un schéma d'assimilation où seuls les sous-produits de l'assimilation sont connus.

– Les matrices connues à priori $\mathbf{D}$ et à posteriori $\mathbf{D}^e$ ne sont pas cohérentes. Par conséquent ces deux termes matriciels dans (5.13) ne se compensent plus. *Desroziers et al.* [2005] propose alors de considérer ces matrices de covariance estimées comme des matrices de covariance ajustées. Les équations (5.11) et (5.12) peuvent être écrites au sens de matrices de covariances ajustées, et notées $\mathbf{B}^e$, $\mathbf{R}^e$ comme suit :

$$\mathbf{E}[\mathbf{d}_b^a \mathbf{d}_b^o] = \mathbf{H}\mathbf{B}^e\mathbf{H}^t = \mathbf{H}\mathbf{B}\mathbf{H}^t(\mathbf{D}^{-1}\mathbf{D}^e) \tag{5.15}$$

$$\mathbf{E}[\mathbf{d}_b^o \mathbf{d}_b^o] = \mathbf{R}^e = \mathbf{R}(\mathbf{D}^{-1}\mathbf{D}^e) \tag{5.16}$$

Lupu and Gauthier [2010] montre ensuite que la définition du $\mathsf{DFS}_{\mathsf{Apost}}$ peut être ré-écrite avec la matrice de covariance d'erreurs d'observations $\mathbf{R}^e$

$$\mathsf{DFS}_{\mathsf{Apost}} = \mathsf{Tr}(\mathbf{H}\mathbf{K}^e) = \mathbf{E}[(\mathbf{d}_b^a)^t(\mathbf{R}^e)^{-1}\mathbf{d}_a^o] \tag{5.17}$$

formulation dans le cadre de l'analyse objective Exprimons enfin le cas de l'analyse objective

$$\mathsf{DFS}_{\mathsf{Apost}} = \mathbf{E}[\widehat{\mathbf{d}_b^{at}}\mathbf{R}^{-1}\widehat{\mathbf{d}_a^o}] = \mathbf{E}[\mathbf{H}\widehat{x_a}\mathbf{R}^{-1}(\mathbf{y} - \mathbf{H}\widehat{x_a})] \tag{5.18}$$

On précise les notations du vecteur dit de résidu et l'image du vecteur d'incrément d'analyse dans l'espace des observations en (5.19)

$$\widehat{\mathbf{d}_a^o} = \mathbf{y} - \mathbf{H}\widehat{x_a} \tag{5.19}$$

$$\widehat{\mathbf{d}_b^a} = \mathbf{H}\widehat{x_a} - \mathbf{H}\widehat{x_b}, \quad ici \quad \mathbf{H}\widehat{x_b} = 0 \tag{5.20}$$

Enfin, supposer la matrice de covariance d'erreurs d'observations diagonale implique que l'erreur entre deux observations distinctes ne soit pas corrélées. Pour des types d'observations comme les instruments in-situ, cette hypothèse est raisonnable.

5.3 Etude de sensibilité du contenu en information

Le contenu en information est très utile pour évaluer la contribution relative des observations assimilées dans un schéma d'assmilation de données. On peut alors se demander comment évolue le contenu en information si les observations sont bruitées, et/ou corrélées. Le but de cette étude est d'illustrer les facteurs influençant le DFS. Pour cela, nous utilisons un schéma d'analyse objective permettant de reconstituer un champ observé à partir de mesures issues d'un champ $\mathcal{Z}$ dont on a une connaissance statistique à priori (moyenne et covariance). Via la définition analytique du contenu en information décrite en section 5.2 nous disposons des valeurs du DFS global associé au champ reconstruit. Nous allons souligner dans la section 5.3.2, l'influence des paramètres tels que l'échelle de corrélation de la matrice de covariance du champ et la variance du bruit sur le DFS d'une part, et sur les champs vrai et estimé d'autre part.

5.3.1 Simulation de distributions gaussiennes et estimateur de Gauss-Markov

On se place sur une grille de points régulière variant de 30° à 35° en longitude et 40° à 45° en latitude. Le pas spatial est de 15km [4]. On génère un champ 2D suivant la loi de probabilité gaussienne telle que $\mathcal{Z} \sim \mathcal{N}(0, \mathbf{B})$. Soit donc B la matrice de covariance du signal, symétrique semi-définie positive de type

$$B(i,j) = \exp(\frac{-d(i,j)^2}{L^2}). \tag{5.21}$$

où **d** est la distance entre deux points de grille i et j en km, et **L** représente l'échelle de corrélation (en km aussi). Elle joue le rôle du diamètre de la bulle d'influence dans laquelle se trouvent les observations. L'analyse objective effectue une moyenne pondérée dans cette bulle d'influence afin de cibler un certain nombre d'observations. La détermination des cœfficients de pondération est liée à l'erreur des observations (ici synthétiques) et la distance au point que l'on cherche à estimer. Plus le point observé est éloigné du point à estimer, plus le coefficient poids est faible. De la même manière, le poids est d'autant plus faible que l'observation est entachée d'erreurs. Enfin l'estimation du champ 2D connaissant un nombre fini d'observations est possible grâce au théorème de Gauss-Markov. On suppose que les observations sont des mesures du champ à estimer. On cherche le meilleur estimateur linéaire au sens des moindres carrés non biaisé de $\mathcal{Z}$ (on donnera l'-expression matricielle de l'estimateur) . Le théorème de Gauss-Markov assure que parmi, tous les estimateurs linéaires non biaisés d'une combinaison linéaire des observations, l'estimateur par moindres carrés présente une variance minimale. Le réseau d'observations est régulier, et chaque observation est espacée de 80km. On obtient l'ensemble des mesures via les valeurs du champ aux points d'observations, et celles entachées d'erreurs notée $\mathbf{Y}_{obs}$ via un bruit aléatoire gaussien de moyenne nulle et de variance Br :

$$\mathbf{Y}_{obs} = \mathcal{Z}(i,j) + \varsigma_{i,j} \tag{5.22}$$

où $\varsigma_{i,j} \sim \mathcal{N}(0, Br)$ et $\mathcal{Z}$ est le champ à estimer.

En effet, soit $\theta_{est}(i,j)$ un estimateur linéaire de $\mathcal{Z}$ au point (i,j),

$$\theta_{est}(i,j) = \sum \alpha_{ij} \mathbf{Y}_{obs} \tag{5.23}$$

tel que :

- $\mathbf{E}(\varsigma_{ij}) = 0$
- $cov(\varsigma_{ii}, \varsigma_{jj}) = Br$
- $var(\varsigma_{ii}) < \infty$

4. La grille possède $25 * 25$ nœuds, sans niveau vertical.

On cherche la meilleure combinaison linéaire des observations $\mathbf{Y}_{obs}$ telle que

$$\mathbf{E}^2 =< [\theta - \theta_{est}]^2 > \tag{5.24}$$

soit minimale.

Alors sous une formulation matricielle on a

$$\theta_{est} = \mathbf{CdgCdd^{-1}Y}_{obs} \tag{5.25a}$$

$$\mathbf{Err} = diag(1 - \mathbf{CdgCdd^{-1}Cdg}^t \tag{5.25b}$$

On note encore **Cdg** la matrice d'erreurs de covariance champs-observations et $\mathbf{Cdd^{-1}}$ la matrice d'erreurs de covariance au sein des observations (t est l'opérateur de transposition).

5.3.2 Tests d'influence sur le DFS

La première expérience a eu pour but de vérifier l'influence de la longueur de corrélation sur les valeurs du DFS global et à chaque point d'observation notés DFS_{obs}, la seconde celle de la variance du bruit de mesure. Le DFS global est obtenu via $\mathsf{Tr}\{\frac{\partial(\mathbf{H}x_a)}{\partial \mathbf{y}}\}$ où x_a est équivalent, ici à θ_{est} (détaillée dans la section 5.2). Le DFS global représente pour une valeur d'échelle de correlation **L** donnée, la somme des dfs_{obs}. En d'autres termes, les éléments diagonaux de la matrice $\mathbf{HCdgCdd^{-1}}$ fournissent l'apport d'information en chaque point d'observation. On rappelle aussi que le DFS est non normalisé par le nombre total d'observations $\mathsf{N_{obs}}$ et que le pas spatial entre deux observations est de 80 km. Par exemple, dans le tableau 5.1 on a :

- pour $\mathbf{L} = 40$ km , la valeur totale sur l'ensemble des 16 observations à cette échelle de corrélation est donnée par le $\mathbf{DFS} = 11.42$
- de la même manière, $\mathbf{L} = 160$ km, $\mathbf{DFS} = 7.56$

De plus, les quantitées notées $dfs_{obs_{min}}$ et $dfs_{obs_{max}}$, représentent à un point d'observation donné, l'apport d'information minimal respectivement maximal.

Variation du DFS en fonction de la longueur de corrélation pour un bruit de mesure fixé

Les variations du DFS en fonction de la longueur de corrélation sont représentées sur la figure 5.1 pour $\mathbf{L}_{min} = 40km$, à $\mathbf{L}_{max} = 240km$ ($\Delta\mathbf{L} = 20$ km). Le DFS décroit lorsque la longueur de corrélation augmente (5.1), et le champ reconstruit devient moins sensible à chacune des observations. En effet lorsque la longueur de corrélation augmente, le nombre d'observations exploitées lors la reconstruction fait de même (et pourrait fournir des observations redondantes). Observons alors les écarts entre les différents dfs_{obs} au point d'observation via le tableau 5.1.

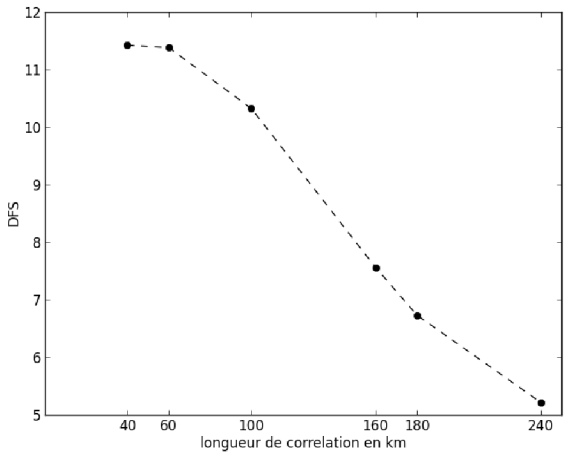


Figure 5.1 – *Variation du DFS en fonction de la longueur de corrélation pour un bruit de mesure de 40% pour les* 16 *observations*

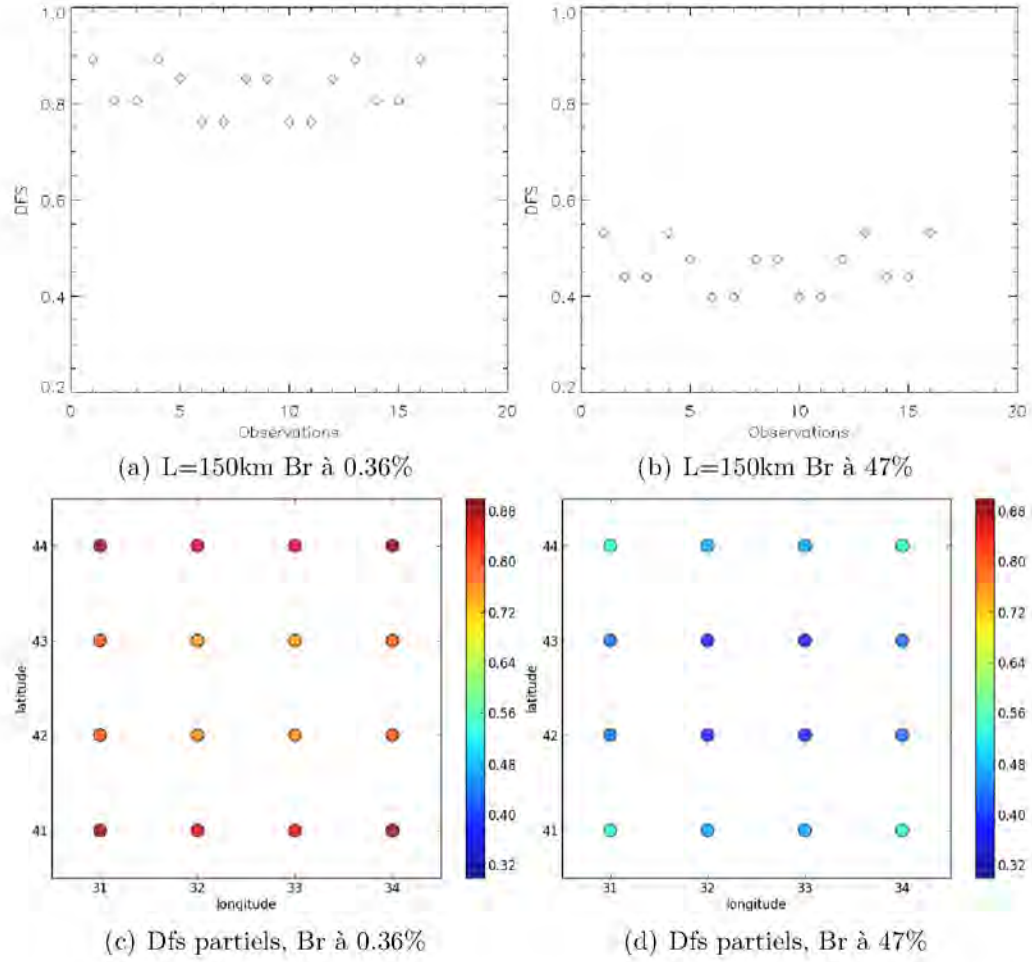
.

L(km)	DFS	$DFS_{obs_{min}}$	$DFS_{obs_{max}}$
40	11.42	0.71	0.71
60	11.38	0.71	0.72
100	10,32	0.62	0.66
160	7.56	0.39	0.54
180	6.73	0.33	0.51
240	5.21	0.23	0.43

Table 5.1 – *Variation du DFS en fonction de la longueur de la corrélation minimale* 40 *km, maximale* 240 *km, pour un bruit de mesure de 40% pour les* 16 *observations*

Comme on pourrait s'y attendre la bulle d'influence délimitée par la longueur de corrélation grandit et a pour conséquence de modifier le poids de chaque observation présente. Selon la localisation du point d'observations sur la grille, lorsque la longueur de corrélation est supérieure au pas spatial d'observations (ici $dx_{obs} = 80$ km) il est alors possible de tenir compte de 3 à 8 observations voisines (si le point est au bord ou au centre). Cela entraine une différence de symétrie entre les dfs_{obs}. Pour exemple les différentes valeurs des dfs_{obs} pour 16 observations peuvent êtres observées sur les figures 5.2. Les figures 5.2 a et b soulignent 4 groupes de 4 observations comportant des valeurs de dfs_{obs} identiques. Celles-ci dépendent bien de la localisation spatiale des observations, les figures 5.2 c et d permettent de représenter la décroissance du dfs_{obs} lorsque les observations sont au cœur de la grille.

En effet les valeurs des points intérieurs sont relativement plus petites que celles des points dits extérieurs, conséquence de l'apport d'information des observations environnantes.



(a) L=150km Br à 0.36%

(b) L=150km Br à 47%

(c) Dfs partiels, Br à 0.36%

(d) Dfs partiels, Br à 47%

Figure 5.2 – *Dfs partiels pour chaque observation pour une longueur de corrélation commune $L = 150$ km et un bruit à 0.36% (gauche) et à 47% (droite).*

Variation du DFS en fonction du bruit de mesure pour une longueur de corrélation fixée

Dans un deuxième temps, nous observons l'influence du bruit de mesure sur le DFS. On a choisi deux configurations spatiales $L_1 < dx_{obs}$ et $L_2 > dx_{obs}$. Comme le montre l'ensemble des résultats présents dans les tableaux 5.2 et 5.3, lorsque le bruit de mesure augmente, le DFS diminue. En effet les observations sont moins précises et par conséquent apportent moins à la reconstruction. Notons aussi que

la sensibilité au bruit de mesure du DFS est différente selon les deux configurations :

Lorsque $L_1 < dx_{obs}$, les dfs_{obs} sont identiques, l'information aux points d'observations est utilisée de manière homogène et ce indépendamment de la valeur de bruit de mesure.

Par ailleurs, comme on peut le voir sur les figures 5.2, lorsque $L_2 \geq dx_{obs}$ les dfs_{obs} ne sont pas uniformes sur la grille et diminuent lorsque l'observation se rapproche du centre et inversement. Notons que l'interdépendance entre les dfs_{obs} en fonction des positions des observations est indépendante du bruit de mesure. En effet les figures 5.2 c et d) soulignent la décroissance des dfs_{obs} lorsque le bruit de mesure augmente, mais aussi la conservation de la même répartition du contenu en information selon la position des observations sur la grille. Les figures 5.2 a et b) le confirment.

Enfin à partir de la figure 5.3, on observe que la décroissance du DFS est moindre lorsque la longueur de corrélation est plus petite. Toujours lié à la définition du DFS, l'information utilisée est plus importante lorsque les observations sont moins corrélées.

expérience	**Br**	**DFS**	**dfs min obs**	**dfs max obs**
1	6%	15.96	0.99	0.99
2	40%	13.96	0.87	0.87
3	70%	10.74	0.67	0.67
4	100%	7.87	0.49	0.49

Table 5.2 – *Influence du bruit de mesure pour* $L_1 = 60km$

expérience	**Br**	**DFS**	**dfs min obs**	**dfs max obs**
1	6%	15.8	0.98	0.99
2	40%	12.14	0.71	0.8
3	70%	8.74	0.5	0.58
4	100%	6.39	0.37	0.43

Table 5.3 – *Influence du bruit de mesure pour* $L_2 = 120km$

Comparaison entre le champ vrai et estimé, et erreurs théoriques

Les figures 5.4 représentent les champs vrais, estimés pour différents bruits de mesure et longueurs de corrélation. Les champs estimés pour le cas $L_1 < dx_{obs}$, quelque soit le bruit de mesure imposé restent en deçà des champs vrais (figures

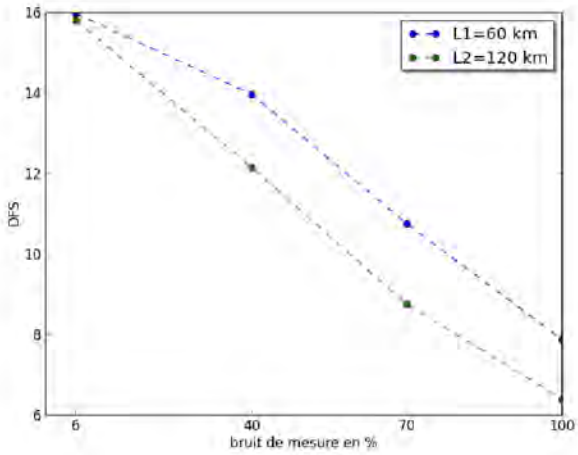
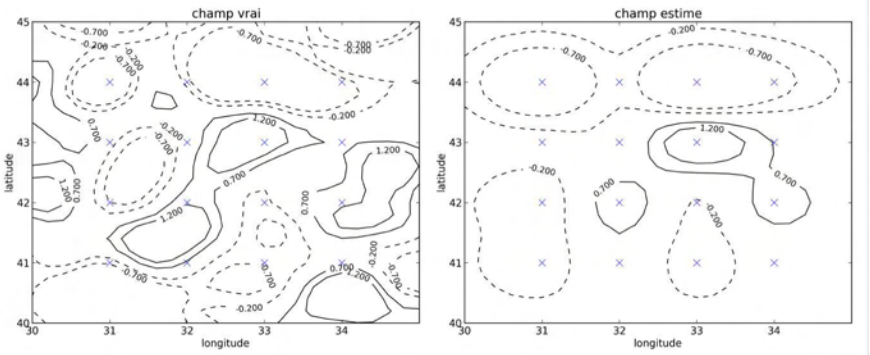


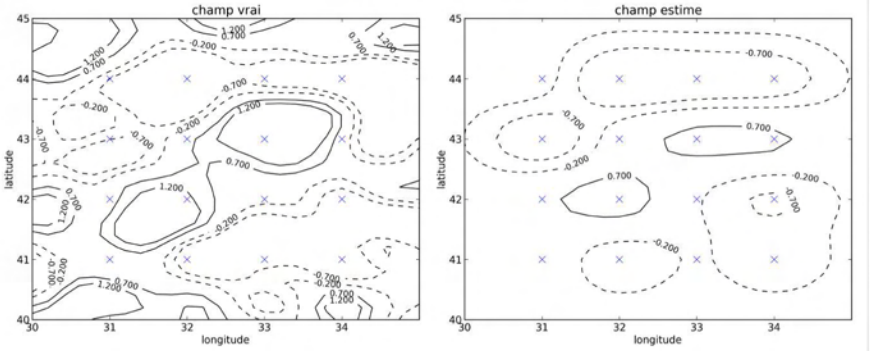
Figure 5.3 – *Variation du DFS en fonction des différentes valeurs de bruits testés (expériences 1 à 4).*

5.4 a et b). Ce constat semble être en accord avec les résultats précédents basés sur le contenu en information : la reconstruction nécessite de davantage d'information que si $L_2 > dx_{obs}$, et celle-ci sera d'autant plus dégradée si les observations sont peu corrélées et entachées d'erreurs. Aussi pour le cas $L_2 > dx_{obs}$, les champs estimés sont moins dégradés à minima de bruit de mesure (figures 5.4 c et d).

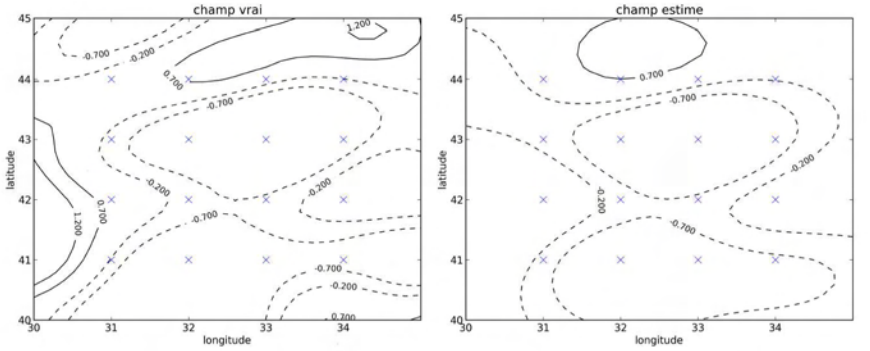
Les figures 5.5 soulignent l'influence du bruit de mesure sur les erreurs théoriques : la zone d'incertitude grandit avec le bruit de mesure. Ces quatre dernières figures reflètent aussi l'impact de la longueur de corrélation sur la zone d'incertitude et donc sur les erreurs théoriques. Lorsque la longueur de corrélation augmente, ces dernières diminuent. La reconstruction nécessite de moins d'information lorsque les observations sont corrélées et engendrerait moins d'erreurs. Ce résultat est cohérent avec la décroissance observée du DFS lorsque L augmente.



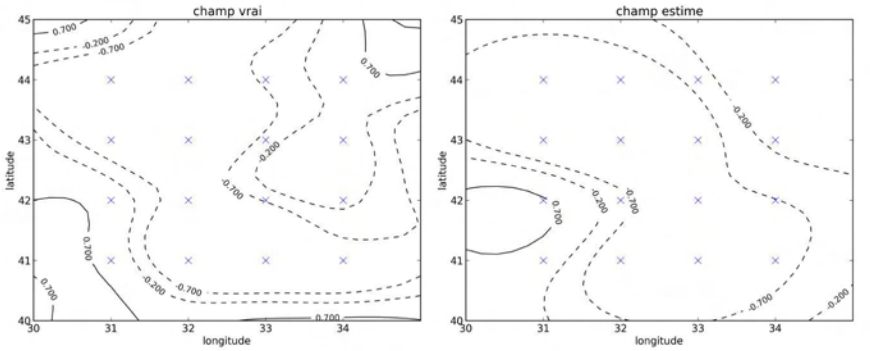
(a) $L_1 = 60$ km et Br à 10%, DFS = 15.25



(b) $L_1 = 60$ km et Br à 50%, DFS = 12.72

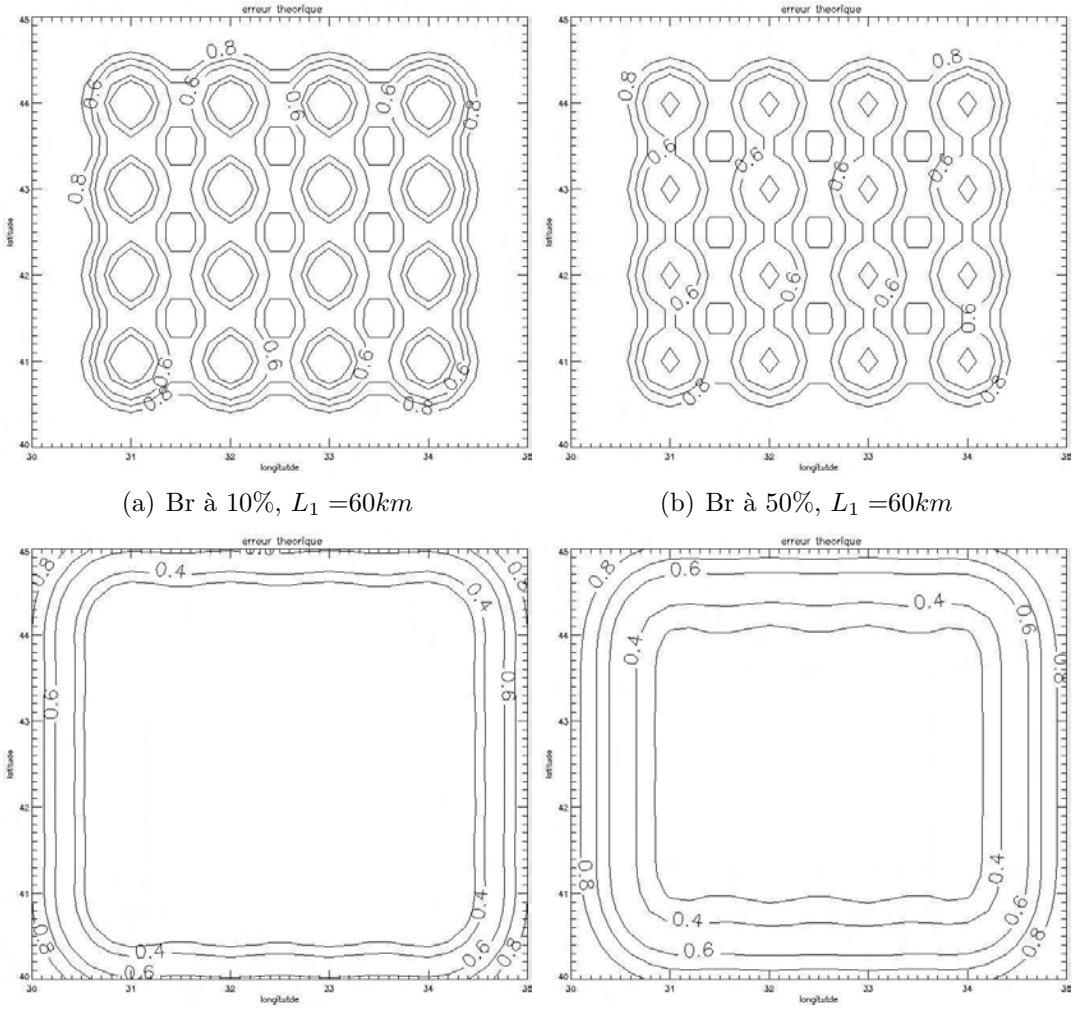


(c) $L_2 = 120$ km et Br à 10%, DFS = 15.40



(d) $L_2 = 120$ km et Br à 50%, DFS = 11.66

Figure 5.4 – *Influence du bruit sur les champs vrai et estimé pour $L_1 < dx_{obs}$ et $L_2 > dx_{obs}$. Les contours en pointillés sont pour des valeurs négatives. Les observations (croix bleue) sont espacées régulièrement de $dx_{obs} = 80$ km.*



(a) Br à 10%, L_1 =60km

(b) Br à 50%, L_1 =60km

(c) Br à 10%, L_2 =120km

(d) Br à 50%, L_2 =120km

Figure 5.5 – *Influence du bruit de mesure sur les erreurs théoriques, pour* $L_1 < dx_{obs}$ *et* $L_2 > dx_{obs}$*. Les contours non nommés sont à* 0.3 *ou* 0.7*. Les observations sont espacées régulièrement de* $dx_{obs} = 80$ *km.*

5.4 Comparaison des estimations du contenu en information

5.4.1 Illustration à partir d'une application au système d'assimilation de Mercator Océan

Le DFS est difficilement estimable de manière exacte dans les systèmes de grande dimension et avec inversion locale tels que ceux opérés par Mercator Océan. C'est pourquoi des méthodes alternatives sont recherchées. Deux approches présentées et définies en section 5.2 ont été testées dans le système global au 1/4° : celle utilisant les différences aux points d'observation entre deux analyses obtenues avec et sans bruitage des observations et celle proposée par Lupu et al [2010]. Alors que la première méthode requiert la réalisation de deux expériences et la perturbation des observations suivant les variances d'erreurs prescrites sur les observations, la seconde méthode est plus attractive car elle ne requiert aucune simulation supplémentaire. Elle est donc beaucoup plus facile à mettre en œuvre dans le cadre d'un système opérationnel : les variables nécessaires à son calcul sont accessibles "offline". Cependant lorsque les deux approches ont été testées pour estimer le DFS de la température de surface (SST) dans une analyse du système global au 1/4°, les résultats sont apparus très différents. La figure (5.6) montre les DFS moyens par bin de 5° x 5° divisés par le nombre d'observations estimés par la méthode dite ici de Girard sur une semaine de août 2010. Alors que les résultats obtenus via la première méthode semblent pouvoir être expliqués physiquement, via la profondeur de la couche de mélange, cela n'est pas le cas pour l'estimation proposée par Lupu (carte non montrée). L'absence d'une estimation exacte du DFS nous empêche de valider ces résultats, et a motivé les tests réalisés dans ce chapitre et présentés en section 5.4.2.

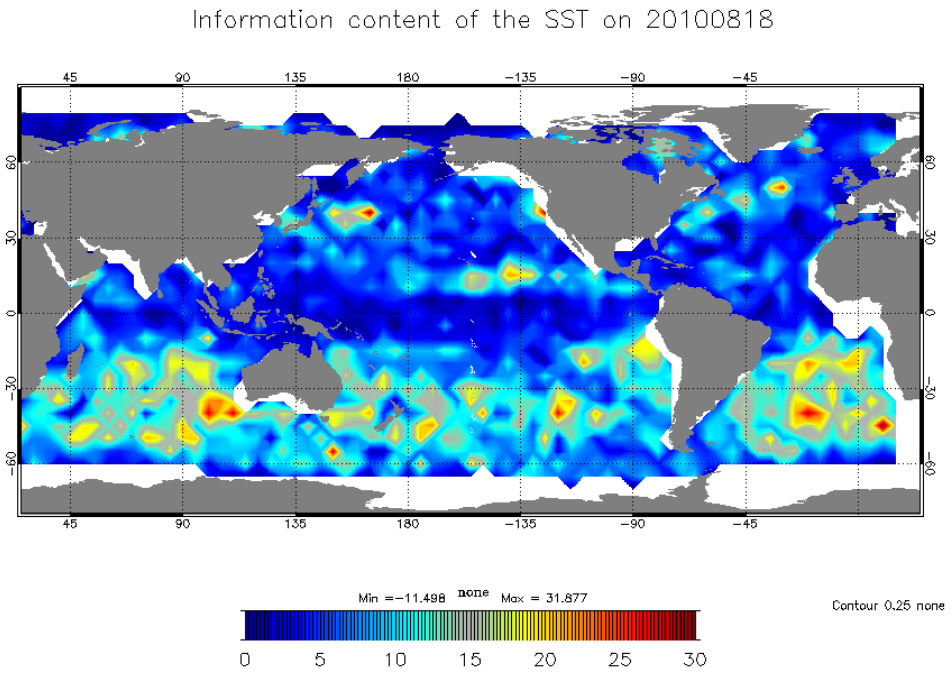


Figure 5.6 – *Contenu en information des observations de température de surface (SST) en pourcentage pour une analyse du système global au* $1/4°$ *en août* 2010*, figure réalisée par E. Remy.*

5.4.2 Mise en pratique dans le cadre d'analyse objective

Ainsi dans cette section on cherche à confronter différentes méthodes de calculs du contenu en information : celle issue de la méthode de randomisation de Girard, celle obtenue via la matrice de covariance d'erreurs des observations supposée diagonale et estimée a posteriori et celle issue de la formulation analytique (définitions section 5.2).

On se place dans le même domaine $2D$ décrit en section 5.3 (grille régulière, pas spatial de 15 km). On note Δx le pas spatial entre deux observations, égal à $1°$ (i.e 80 km). Le vecteur d'état vrai noté x_t est de dimension $ns = 26 > p = 16$, où p est la dimension du vecteur d'observations. Le nombre de degrès de liberté du système ns est supérieur, comme souvent dans la réalité, au nombre d'observations, ce qui en fait un système sous-déterminé. Le système d'observations est donc fixé à 16 observations réparties régulièrement tous les degrès et centrées loin des bords du domaine (même configuration que pour la section 5.3). Les observations sont simulées en ajoutant un bruit normal au vecteur d'état vrai (notation 5.26). On rapelle brièvement que le vecteur d'état suit une loi normale centrée. La matrice

B de covariance du vecteur d'état a été définie à la section 5.3.1.

$$x_t \sim \mathcal{N}(0, \mathbf{B})$$

$$\mathbf{Y}^0 = \mathbf{H}(x_t) + \epsilon^0 \tag{5.26}$$

$$\epsilon^0 \sim \mathcal{N}(0, \sigma_{0t})$$

$$\mathbf{R} = \mathbf{E}[\epsilon^0(\epsilon^0)^t] = \sigma_0^2 \mathbf{I}_p$$

Où σ_{0t} est la variance d'erreurs d'observations spécifiée (true) et σ_0 celle que l'on cherche à ajuster via la matrice de covariance d'erreurs d'observations **R**. Les erreurs d'observations sont décorrélées spatialement ce qui justifie la forme diagonale de la matrice de covariance d'erreurs d'observations.

Lors des deux expériences réalisées, on utilise une liste croissante de longueur de corrélation du signal L_t prescite dans la définition de la matrice de covariance des erreurs statistiques utiles à l'analyse objective (aussi définies dans la section 5.2.3). La première expérience repose sur l'hypothèse que la variance d'erreurs d'observations est parfaitement connue ($\sigma_{0t} = \sigma_0$). La deuxième expérience repose sur l'hypothèse que la variance d'erreur d'observations est sous-estimée ($\sigma_{0t} > \sigma_0$). Pour l'ensemble de ces expériences la variance d'erreurs d'observations se doit d'être faible par rapport au vecteur d'état afin de permettre une bonne estimation, car dans un cas d'analyse objective nous ne disposons pas d'ébauche pour améliorer le vecteur d'état. Par conséquent les observations seront faiblement bruitées (proches du vecteur vrai) afin de permettre une reconstruction de bonne qualité : $\sigma_0 t = 0.2$.

1^e **expérience : variance d'erreurs d'observations parfaitement connue et observations faiblement bruitées** Lorsque la variance d'erreurs d'observations est connue parfaitement, les deux formulations estimées (5.18) et (5.10) du DFS coïncident avec la formulation objective théorique. De plus l'expression du DFS issue des statistiques d'erreurs à posteriori et celle issue de la méthode de Girard présentent des résultats similaires, tableau (5.4) : leur différence relative pour l'ensemble des longueurs de corrélation testées est en moyenne inférieure à 2%. Enfin lorsque la longueur de corrélation augmente le contenu en information tends à diminuer : cette tendance est en accord avec les tests des facteurs d'influence sur le DFS réalisés à la section (5.2.3).

variance d'erreur des observations parfaitement connue : $\sigma_{0t} = \sigma_0 = 0.2$			
L(km)	$\mathsf{DFS}_{\mathsf{Analytique}}$	$\mathsf{DFS}_{\mathsf{Girard}}$	$\mathsf{DFS}_{\mathsf{Apost}}$
$60 < \Delta x$	15,36	15,26	15,92
100	14,84	14,94	15,04
180	10,98	10,76	10,43
240	8,60	8,76	8,81
300	7,09	6,81	7,01

Table 5.4 – 1^e *expérience : variance d'erreurs d'observations parfaitement connue et observations faiblement bruitées*

2^e **expérience : variance d'erreurs d'observations sous-estimée et observations faiblement bruitées** Dans cette série de résultats (tableau 5.5), la variance d'erreur d'observations est choisie inférieure à celle connue . Par comparaison numérique avec l'expérience précédente, à moindre bruit le DFS est supérieur. Cette tendance est encore en accord avec celle observée dans la section 5.2.3. De plus, les résultats issus des expressions du $\mathsf{DFS}_{\mathsf{Girard}}$ et du $\mathsf{DFS}_{\mathsf{Apost}}$ sont similaires, avec une différence relative de 1% à 3% pour une longueur de corrélation, $L_t = 60$ km inférieure à la résolution spatiale des observations. Par ailleurs dès que L_t est $> \Delta x$ les résultats issus de l'expression $\mathsf{DFS}_{\mathsf{Apost}}$ augmentent plus fortement que pour le cas issu du $\mathsf{DFS}_{\mathsf{Girard}}$, avec une différence relative atteignant les 35%. De plus lorsque la longueur de corrélation augmente, les valeurs du $\mathsf{DFS}_{\mathsf{Apost}}$ ne diminuent pas, contrairement à ce qui serait attendu (tendance soulignée en section 5.2.3). La tendance ainsi que les valeurs numériques issues de l'expression $\mathsf{DFS}_{\mathsf{Apost}}$ ne sont pas en accord avec ceux de la formulation analytique.

variance d'erreur des observations sous-estimée : σ_{0t}=0.2 et σ_0=0.1			
L(km)	$\mathsf{DFS}_{\mathsf{Analytique}}$	$\mathsf{DFS}_{\mathsf{Girard}}$	$\mathsf{DFS}_{\mathsf{Apost}}$
60	15,83	15,43	15.93
100	15,67	15,93	16,74
180	13,15	12,99	17,14
240	10,71	10,65	15,41
300	8,92	9,00	12,42

Table 5.5 – 2^e *expérience : variance d'erreurs d'observations sous-estimées et observations faiblement bruitées*

L'hypothèse de variance d'erreurs d'observations sous-estimée faite en expérience 2, permet de remarquer que le DFS est dans ce cas sur-estimé. Lors de ces deux expériences nous nous sommes placés dans un cadre d'analyse objective afin d'avoir accès à la formulation analytique du DFS. Par comparaison elle permet

donc de conclure quant à la robustesse de des formulations présentées en section 5.2. La méthode d'approximation du DFS basée sur la randomisation de Girard présente des résultats cohérents avec la formulation analytique du DFS, ce qui n'est pas le cas pour la seconde expérience réalisée avec la méthode dite aposteriori. En effet lorsque l'hypothèse de variance d'erreurs d'observations sous-estimée est faite, on se place alors dans le deuxième cas de figure décrit en section 5.2. Lorsque les matrices de covariance d'innovation connues à priori et a posteriori ne sont plus cohérentes, il semble que la formulation $\mathsf{DFS}_{\mathsf{Apost}}$ ne soit plus valable. Enfin la méthode d'approximation du DFS basée sur la randomisation de Girard, et utilisant deux jeux d'observations perturbées (suivant les variances d'erreurs prescrites sur les observations) semble plus robuste pour estimer le DFS.

5.5 Conclusion

Au sein de ce chapitre, nous avons mis en évidence les relations inhérentes entre le contenu en information, les échelles de corrélation et le bruit de mesure. Il a été mit en avant que le DFS permet de quantifier l'apport d'information et d'analyser en quelle proportion les observations sont utilisées (proportion quantitative et localisée selon la grille). Il paraît donc essentiel d'évaluer précisément et de manière optimale le contenu en information. De ce fait, plusieurs formulations ont été évaluées dans un cas simple d'analyse objective dans le but d'élaborer une première approche d'estimation du DFS applicable par la suite dans le système d'assimilation de données de Mercator Océan. Le DFS peut être calculé de manière exacte et comparé à ces différentes estimations. Les résultats obtenus semblent indiquer que seule la première appoche (celle dite de Girard) se montre robuste. La seconde alternative sera donc écartée pour le calcul des DFS dans les systèmes mis en œuvre par Mercator Océan. Des travaux et tests complémentaires sont actuellement effectués par les équipes *R&D* d'assimilation de Mercator Océan.

Conclusion et perspectives

Synthèse des travaux

Les travaux de thèse ont montré l'apport des données Argo pour valider les modèles océaniques et caractériser les erreurs modèles (biais moyens, biais saisonniers et écarts-types). Les données Argo permettent aussi maintenant de caractériser plus finement les erreur modèles via l'estimation des échelles associées. Nous avons aussi montré que les échelles spatiales peuvent être calculées précisément à la fois pour les erreurs du modèle et pour le signal océanique. Il est d'ailleurs intéressant de relier la structure spatiale des erreurs modèles à celles du signal. Les données Argo permettent aussi et surtout de contraindre les modèles par le biais de l'assimilation de données. Comme montré récemment par *Turpin et al.* [2015] l'impact des données Argo pour l'analyse et la prévision océanique est majeur. Cet impact est quantifié via des méthodes coûteuses de type OSE (Observing System Evaluations) ; des techniques alternatives et complémentaires ont été proposé dans la littérature (*Oke et al.* [2009], *Cardinali et al.* [2004]). Nous avons testé ces différentes méthodes dans des cas simplifiés afin de préparer leur utilisation dans les systèmes plus complexes d'assimilation de données mis en oeuvre par Mercator Océan.

Principaux résultats

Au cours de cette thèse, nous avons mis en œuvre des méhodes pour l'utilisation des observations Argo afin de caractériser les erreurs modèles et de contraindre les systèmes d'assimilation.

Caractérisation des erreurs modèles

Lors de cette étude nous avons fait le choix d'utiliser une simulation numérique issue d'un modèle de circulation générale au $1/4°$ *eddy-permitting* couvrant la période $2006 - 2009$ n'ayant pas assimilé les données Argo (et aucune autre). Caractériser les erreurs modèles a permis de déterminer dans quelle mesure la simu-

lation numérique, étudiée dans cette thèse, diffère de la réalité observée par les données Argo. L'évaluation des erreurs modèles passe par une étape de colocalisation nécessaire à l'extraction de l'information simulée pour la comparer dans le même référentiel que celui des observations. Les erreurs modèles sont ensuite issues d'un traitement statistique fournissant une vision globale sur la période 2006-2009 et par bins de 4°x4° de la surface à 2000 m. Ainsi nous avons pu souligner la nature grande échelle des erreurs modèles. Globalement la simulation numérique présente une thermocline trop diffuse et surestime les champs de salinité de la surface à 2000 m. Les biais en température les plus importants se trouvent dans l'Atlantique Nord, les régions du Kuroshio, le courant des Aiguilles et du courant Antarctique Circumpolaire. Les biais en salinité ont une répartition géographique différente et plus homogène verticalement.

Facteurs impactant l'estimation des échelles spatiales de corrélation

L'algorithme statistique d'estimation des échelles spatiales de corrélation repose sur la sélection d'un nombre suffisant de profils d'observations Argo selon leurs position et date. Le champ de covariance (ou de corrélation) résultant est alors uniquement dépendant des distances zonales et méridionales séparant les points d'observations et non plus leur localisation exacte. Les résultats présentés au chapitres 4 ont montré que l'estimation des fonctions de corrélation dépend de plusieurs facteurs liés à l'échantillonnage. Une estimation statistique fiable des fonctions de corrélation nécessite un nombre important de paires d'observations pour chaque classes de distance. Dans les conditions numériques étudiées, de 50 à 100 paires d'observations sont nécessaires. Notons que pour atteindre ce nombre moyen il est nécessaire d'exploiter plusieurs années de données sur des zones d'au moins 20°x20°. Enfin les estimations des fonctions de corrélation réalisées à partir de deux années isolées 2005 et 2013 ont permis de souligner que l'échantillonnage du réseau Argo fournit en 2013 des résultats plus précis, avec une diminution à la fois des erreurs formelles sur les estimations et des erreurs d'ajustement sur les échelles spatiales déduites.

Caractérisation des échelles spatiales des erreurs modèles et des anomalies par rapport à la climatologie

Nous avons analysé la nature et des variations, en fonction des latitudes et de la profondeur, des échelles spatiales des anomalies de température et de la salinité par rapport à la climatologie Levitus 2009 (*Levitus et al.* [2009]). Notons que cette caractérisation menée de manière globale n'avait encore jamais été faite. L'ensemble des résultats est cohérent avec les études précédentes reposant sur des données altimétriques ou issues de réseau d'observations in-situ sporadiques. Les échelles

spatiales de température zonales et méridionales varient de 900 km aux latitudes tropicales à 350 km aux latitudes moyennes. De plus on observe une diminution vers les pôles, jusqu'à 100 km aux hautes latitudes. Les fonctions de corrélation estimées sont anisotropes aux latitudes équatoriales/tropicales. Ainsi on a pu noter un facteur entre les échelles zonales et méridionales allant de 2 à 3. Il est intéressant de souligner les variations des échelles selon la profondeur, notamment en température. En effet à l'exception des premiers 200 m les échelles spatiales de température augmentent avec la profondeur. Nous avons montré également que les échelles spatiales en température et salinité ne possèdent pas les mêmes tendances. Particulièrement, la croissance des échelles en fonction de la profondeur est absente dans le cas de la salinité.

Nous avons comparé les échelles du signal aux échelles associées aux erreurs modèles. Les échelles spatiales estimées à partir des erreurs modèles en température et salinité ont montré des variations similaires au cas climatologique en fonction des latitudes et de la profondeur mais aussi des différences significatives en surface et en profondeur, notamment en température. Le modèle parvient de manière plus réaliste à représenter la variabilité grande échelle. Enfin en profondeur les variations des échelles issues des deux calculs ont aussi présenté des différences notables. Il en résulte qu'approximer la structure des erreurs modèles par celles du signal n'est sans doute pas une solution suffisante pour approximer les erreurs sur l'ébauche modèle lors de l'assimilation de données.

Analyse des méthodes basées sur le contenu en information

Estimer le DFS (ou le contenu en information) de manière exacte s'avère difficile dans les systèmes de grande dimension et avec une inversion locale. Il existe des formulations alternatives permettant d'aproximer le DFS mais en l'absence d'estimation exacte il n'est pas possible de conclure sur la robustesse des formulations testées. C'est pourquoi dans un cadre simple d'analyse objective nous avons pu calculer de manière exacte le DFS et confronter deux formulations basées sur le contenu en information. La première est extraite de la méthode proposée par *Girard* [1987] et la seconde par *Lupu and Gauthier* [2010]. Chacune d'elles a pour but, connaissant l'incertitude sur les observations, de mesurer la sensibilité de l'analyse par rapport aux observations. Cette comparaison nous a permis de souligner certaines recommandations nécessaires à l'exploitation de la formulation proposée par *Lupu and Gauthier* [2010] dans un schéma plus complexe d'assimilation de données. Les résultats obtenus avec la formulation proposée par *Lupu and Gauthier* [2010] ne sont pas systématiquement cohérents avec l'expression analytique ni avec la formulation proposée par *Girard* [1987]. Nous privilégierons alors la formulation proposée par *Girard* [1987] pour la mise en place future d'un calcul des DFS dans les systèmes mis en œuvre par Mercator Océan. Ces travaux sont

actuellement menés par les équipes de *R&D* d'assimilation de Mercator Océan.

Perspectives

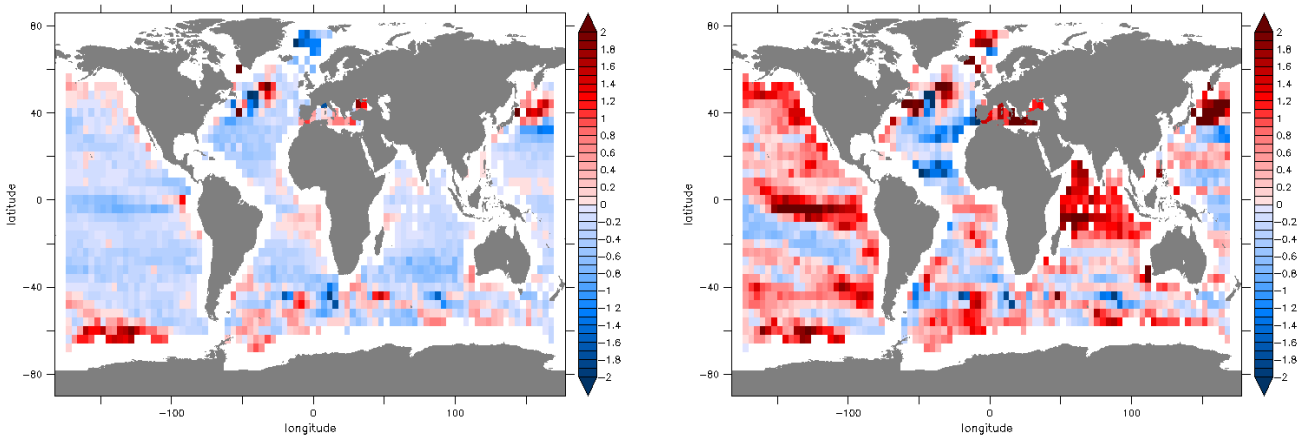
Les réseaux d'observations de l'océan global, notamment Argo, sont relativement récents et couvrent des périodes encore assez courtes. Le réseau Argo n'a atteint son objectif initial (3000 flotteurs en opération) que fin 2007 seulement ; la serie temporelle Argo continue de s'allonger permettant des estimations statistiques des erreurs modèles et du signal océanique ainsi que des échelles spatiales associées de plus en plus fiables. Il est donc important de poursuivre ces estimations sur une période plus longue, permettant ainsi d'analyser les erreurs modèles à basse fréquence et aussi d'affiner les résolutions des calculs d'échelle (par exemple sur des zones de 5°x5° via des fenêtres glissantes afin de mieux caractériser les variations géographiques). Il serait utile d'étendre ces estimations au calcul d'échelles temporelles et spatio-temporelles. Grâçe à ces futures estimations, nous pourrions mieux définir les évolutions souhaitables des caractéristiques d'échantillonnage du réseau Argo ; celles-ci devraient, en effet, être en théorie reliées aux échelles spatio-temporelles des erreurs modèles et/ou du signal. On pourrait ainsi, par exemple, recommander d'affiner la résolution spatiale du réseau dans les régions où les erreurs modèles et/ou de la variabilité sont associées à de plus petites échelles spatiales. Il faudrait aussi analyser comment ces informations sur les erreurs modèles et leurs échelles pourraient être prises en compte pour améliorer les schémas d'assimilation de données mis en œuvre à Mercator Ocean. Par ailleurs il serait également intéressant d'estimer les échelles spatiales de corrélation des erreurs [analyse - observations] à partir de la réanalyse Glorys de Mercator Océan afin de connaître l'impact de l'assimilation de données sur ces échelles. Les caractérisations des erreurs modèles pourront aussi être utilement menées sur de nouvelles simulations notamment à plus haute résolution (par exemple modèle global au 1/12°. Enfin, dans la dernière partie, nous avons souligné certains recommendations quant à l'exploitation de formulations basées sur le contenu en information. Il sera très intéressant de mettre en œuvre ces techniques dans les systèmes d'assimilation Mercator Océan et de confronter les résultats obtenus à ceux des OSEs récemment menées à Mercator Océan pour les données Argo (*Turpin et al.* [2015]).

Annexe A

Catalogue des biais annuels et saisonniers modèle-observations sur la période 2006-2009

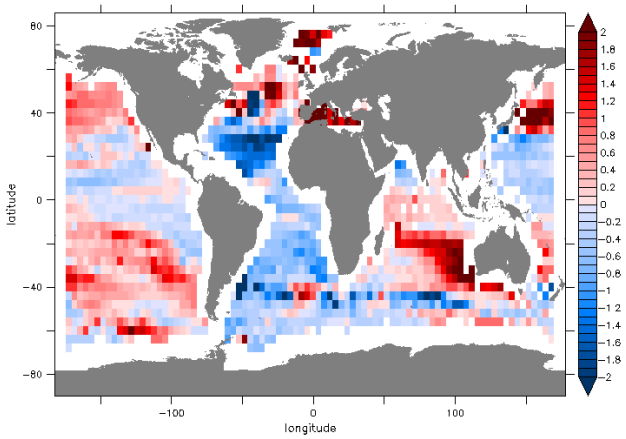
Biais annuels en température 2006-2009

Répartition verticale faite selon les niveaux de la climatologie Lévitus 2009

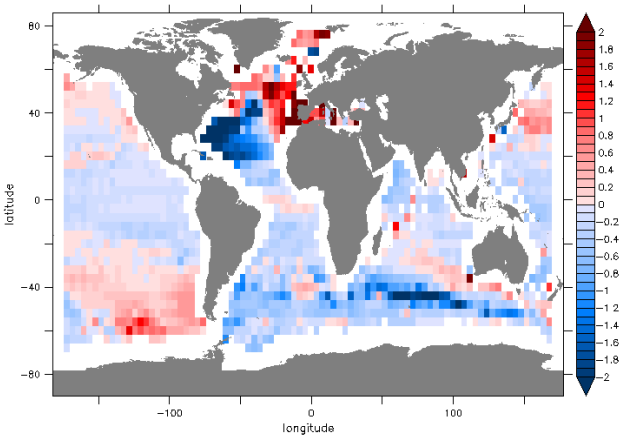


(a) 0-10 m

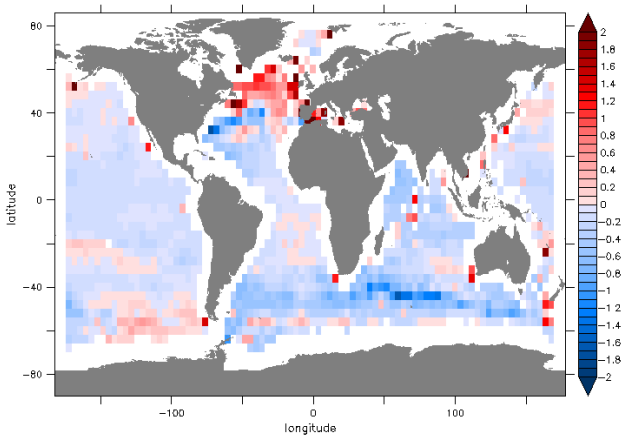
(b) 125-150 m



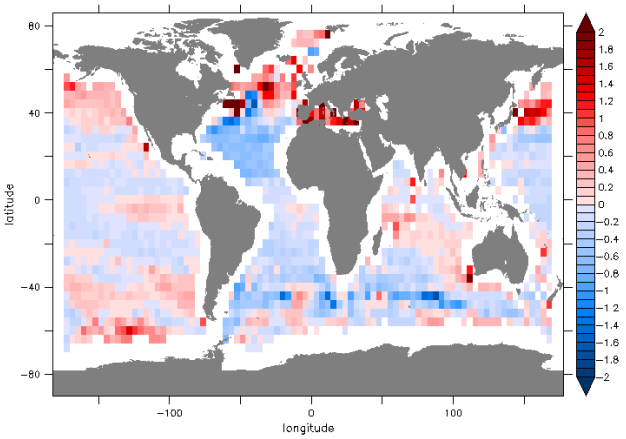
(c) 300-400 m



(d) 600-700 m



(e) 1000-1100 m

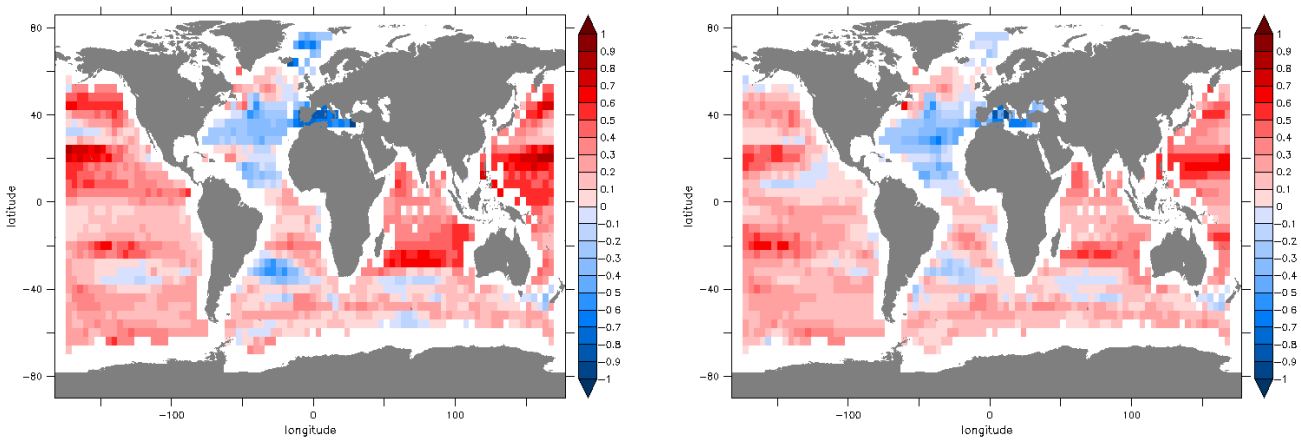


(f) 0-2000 m

Figure A.1 – *Erreurs modèle moyennes en température (en °C) sur la période 2006-2009 pour différents niveaux standards Levitus.*

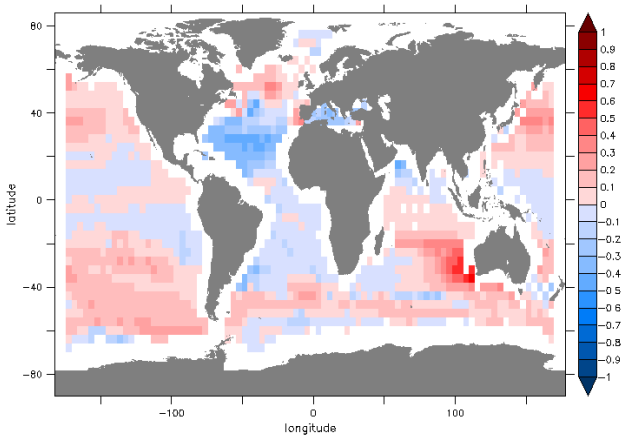
Biais annuels en salinité 2006-2009

Répartition verticale faite selon les niveaux de la climatoligie Lévitus.

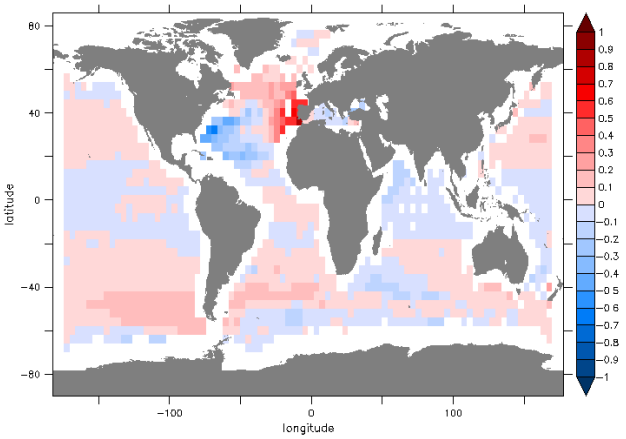


(a) 0-10 m

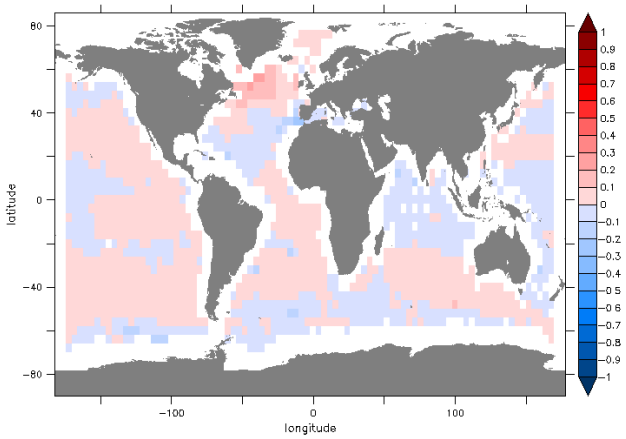
(b) 125-150 m



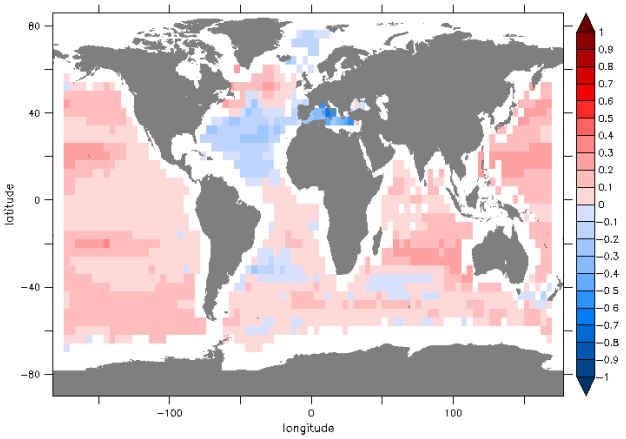
(c) 300-400 m



(d) 600-700 m



(e) 1000-1100 m

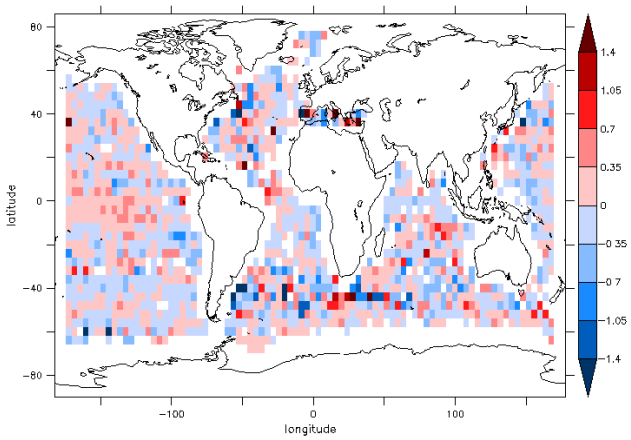


(f) 0-2000 m

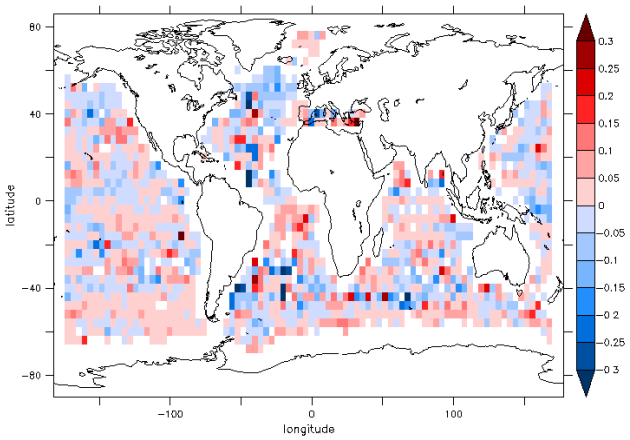
Figure A.2 – *Erreurs modèle moyennes en salinité (en psu) sur la période 2006-2009 pour différents niveaux standards Levitus.*

Biais saisonniers en température et salinité 2006-2009

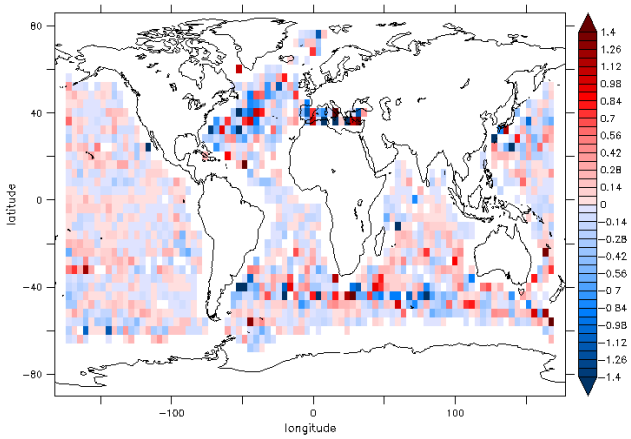
Printemps 2006-2009



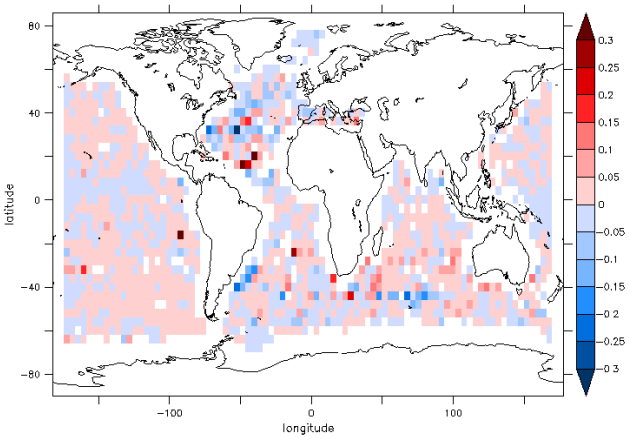
(a) écarts en température, $0 - 300$m



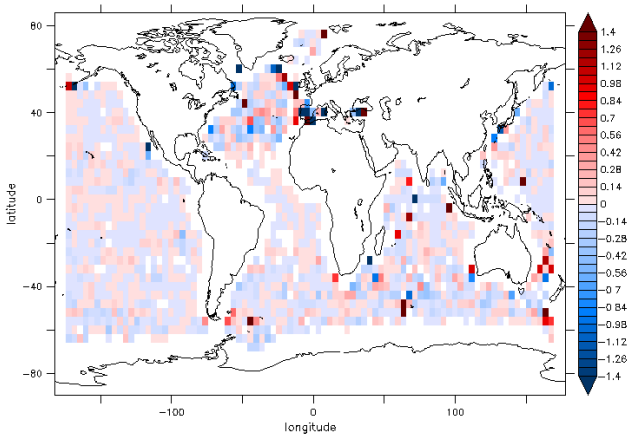
(b) écarts en salinité, $0 - 300$m



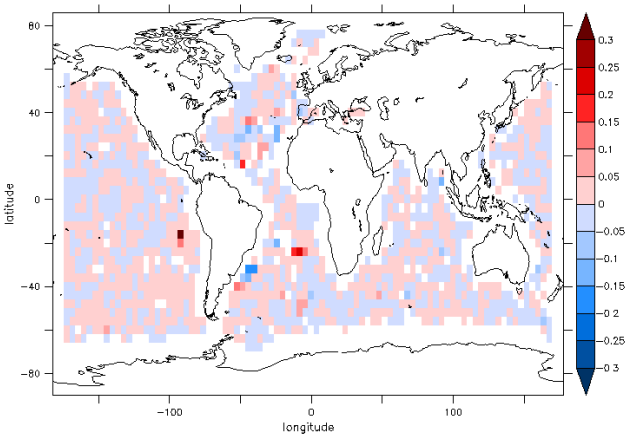
(c) écarts en température, $300 - 700$m



(d) écarts en salinité, $300 - 700$m



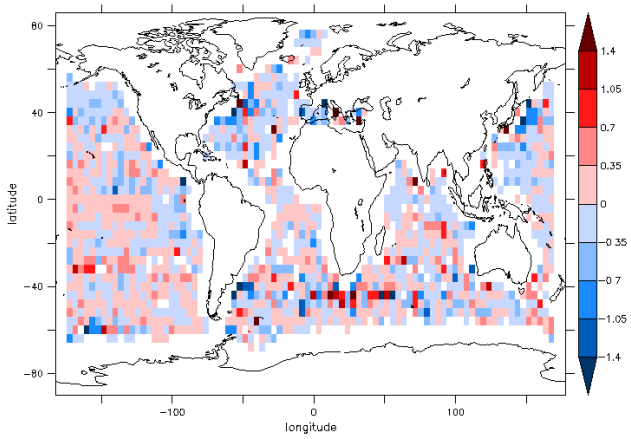
(e) écarts en température, $700 - 2000$m



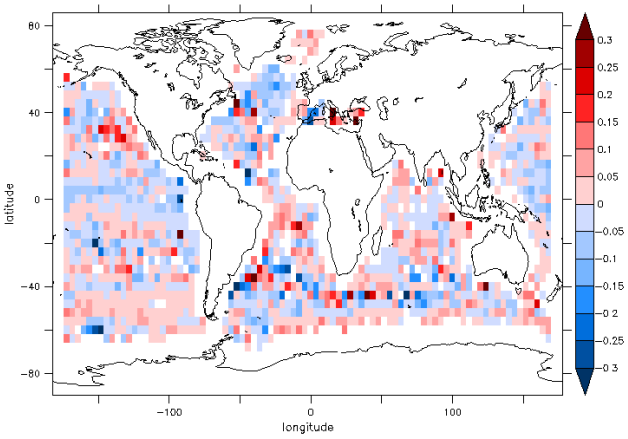
(f) écarts en salinité, $700 - 2000$m

Figure A.3 – *Ecarts moyens en température et salinité entre le modèle et les données Argo sur les couches* $dp1 = [0 - 300]m$, $dp2 = [300 - 700]m$ *et* $dp3 = [700 - 2000]m$ *pour les printemps* $2006 - 2009$.

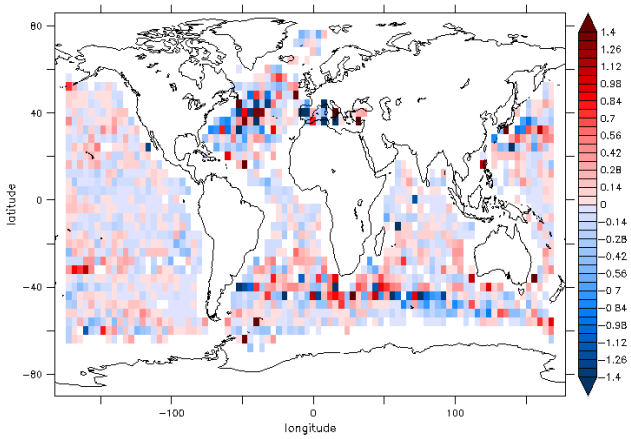
Eté 2006-2009



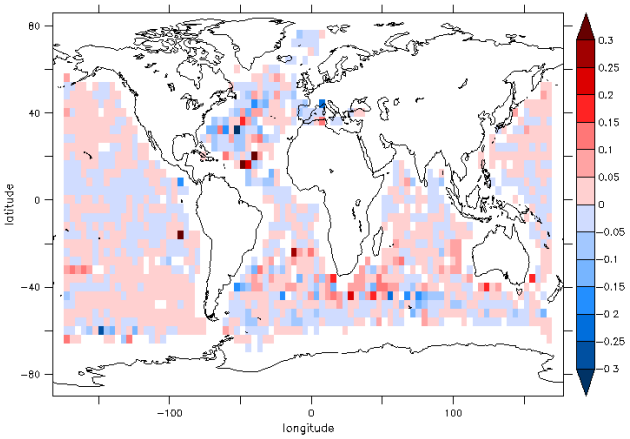
(a) écarts en température, $0 - 300$ m



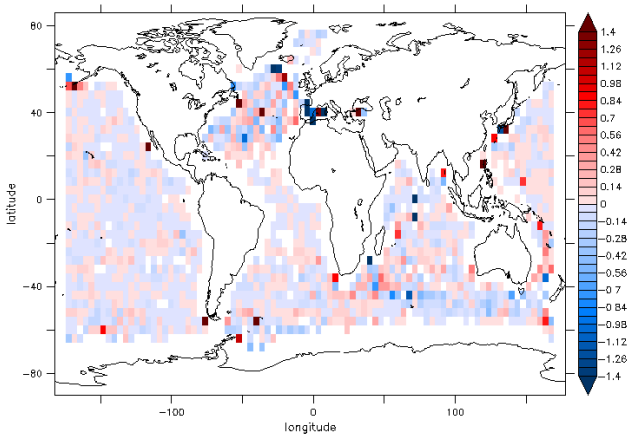
(b) écarts en salinité, $0 - 300$ m



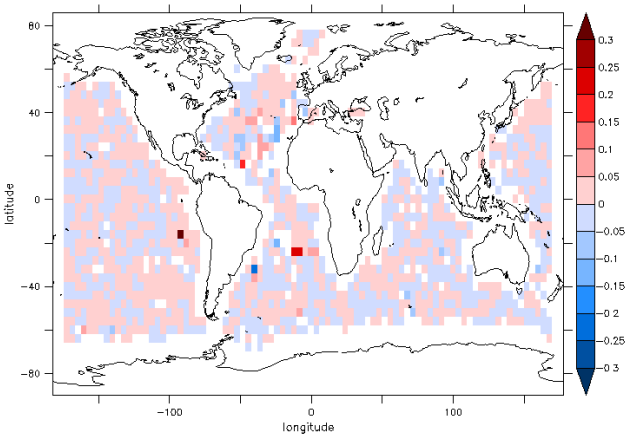
(c) écarts en température, $300 - 700$ m



(d) écarts en salinité, $300 - 700$ m



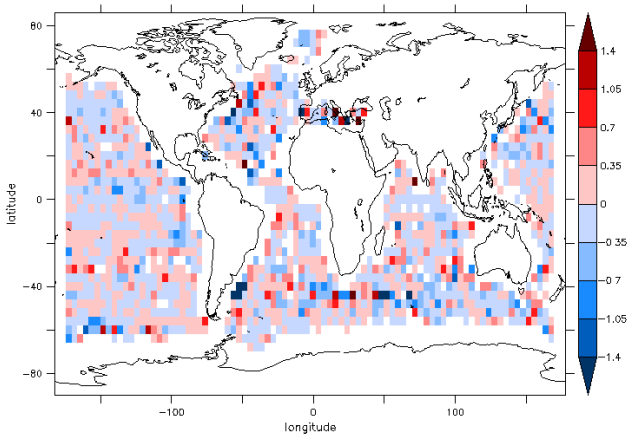
(e) écarts en température, $700 - 2000$ m



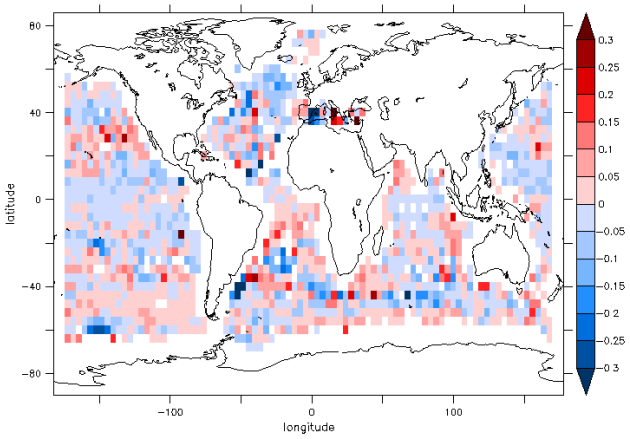
(f) écarts en salinité, $700 - 2000$ m

Figure A.4 – *Ecarts moyens en température et salinité entre le modèle et les données Argo sur les couches* $dp1 = [0 - 300]m$, $dp2 = [300 - 700]m$ *et* $dp3 = [700 - 2000]m$ *pour les étés 2006-2009.*

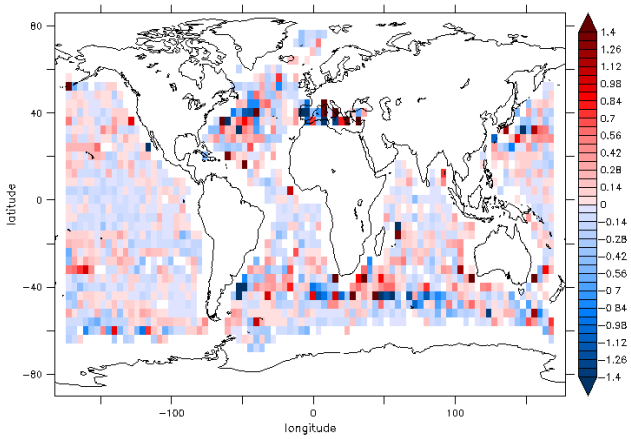
Automne 2006-2009



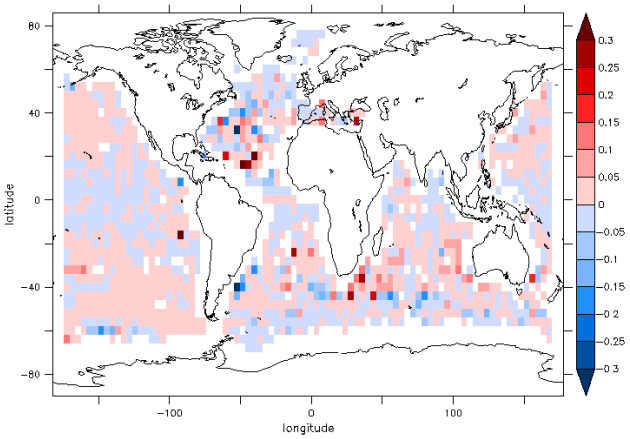
(a) écarts en température, $0 - 300$ m



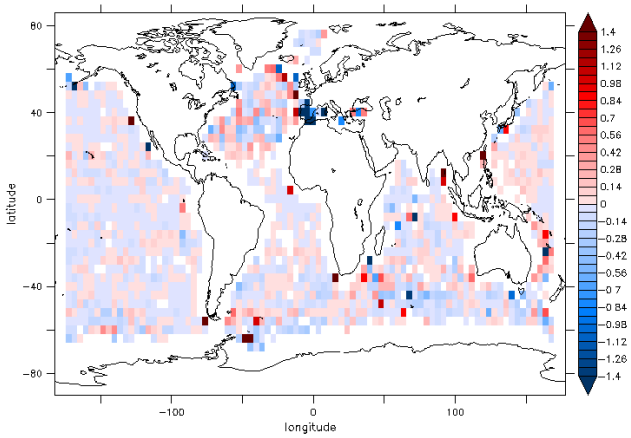
(b) écarts en salinité, $0 - 300$ m



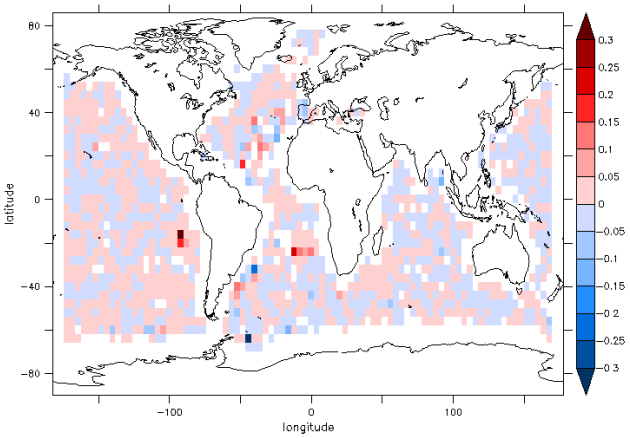
(c) écarts en température, $300 - 700$ m



(d) écarts en salinité, $300 - 700$ m



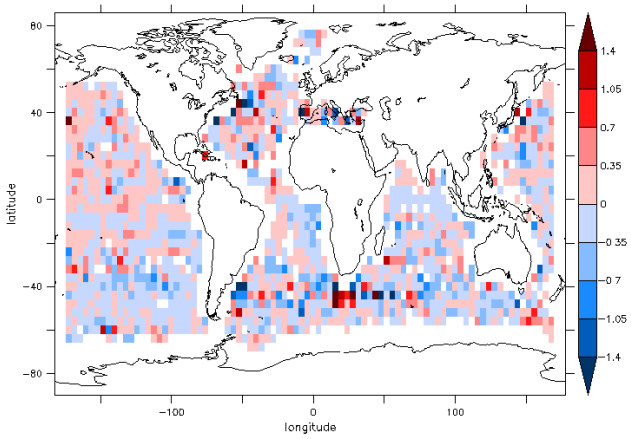
(e) écarts en température, $700 - 2000$ m



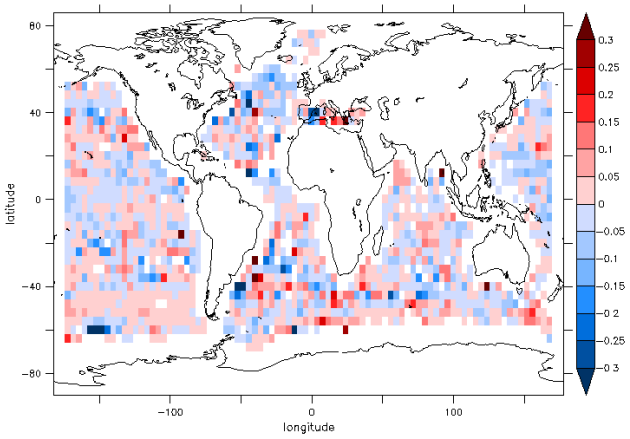
(f) écarts en salinité, $700 - 2000$ m

Figure A.5 – *Ecarts moyens en température et salinité entre le modèle et les données Argo sur les couches* $dp1 = [0 - 300]m$, $dp2 = [300 - 700]m$ *et* $dp3 = [700 - 2000]m$ *pour les autumnes 2006-2009.*

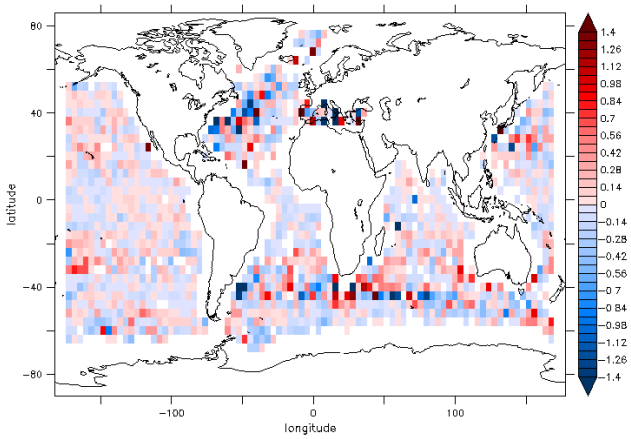
Hiver 2006-2009



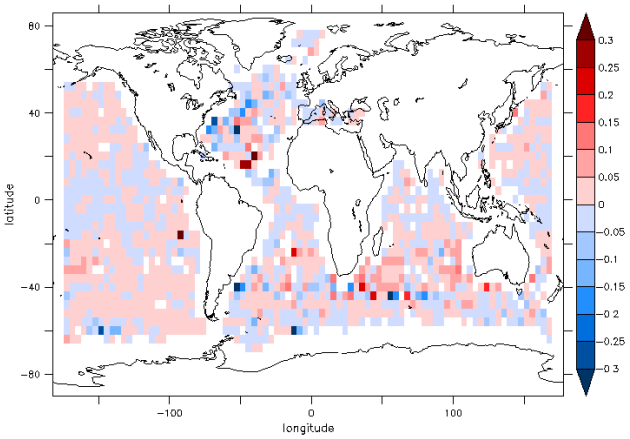
(a) écarts en température, $0 - 300$ m



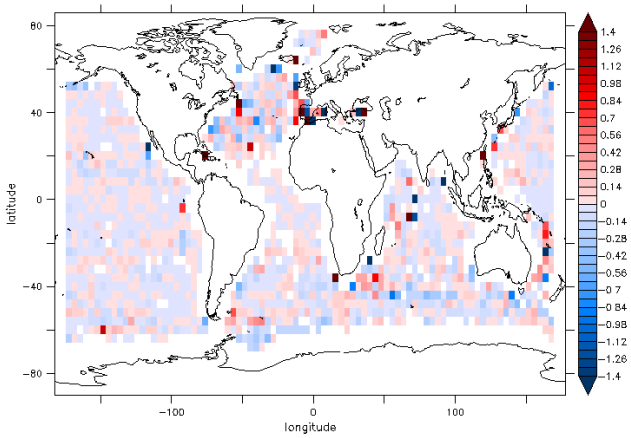
(b) écarts en salinité, $0 - 300$ m



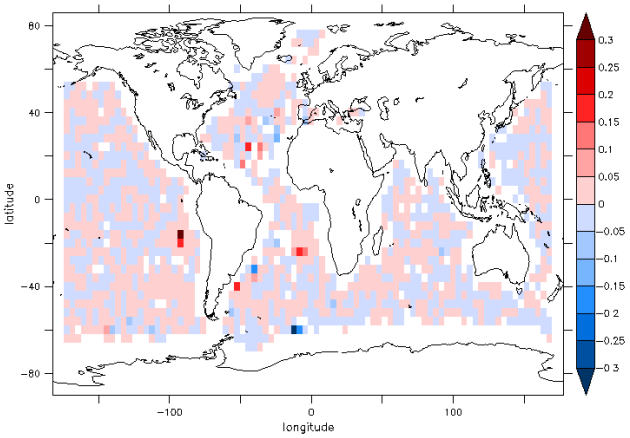
(c) écarts en température, $300 - 700$ m



(d) écart en salinité, $300 - 700$ m



(e) écarts en température, $700 - 2000$ m



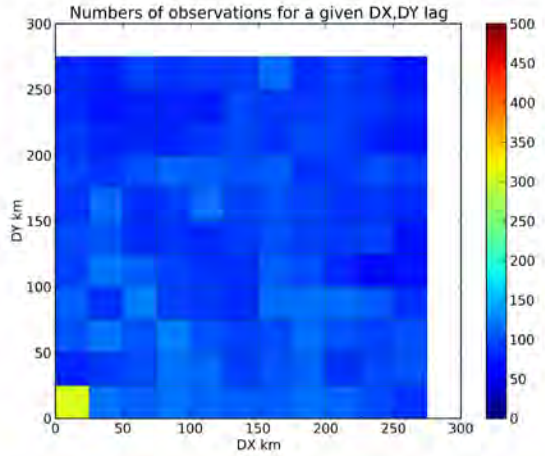
(f) écarts en salinité, $700 - 2000$ m

Figure A.6 – *Ecarts moyens en température et salinité entre le modèle et les données Argo sur les couches* $dp1 = [0 - 300]m$, $dp2 = [300 - 700]m$ *et* $dp3 = [700 - 2000]m$ *pour les Hivers 2006-2009.*

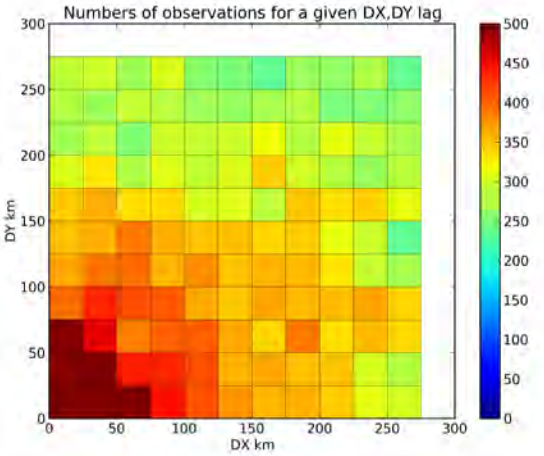
Annexe B

Catalogue des échelles spatiales de corrélation des erreurs modèles et de la circulation océanique

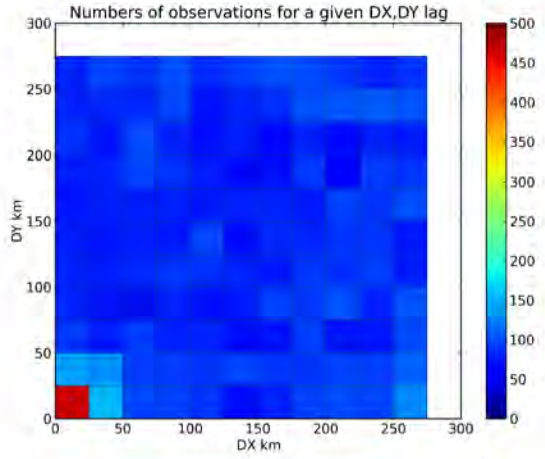
Répartition du nombres de paires de profils en fonction des zones à 200m sur la période 2006 – 2009



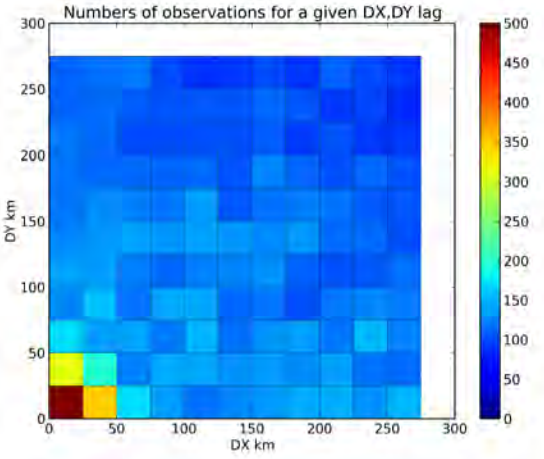
zone 1, Pacifique Est



zone 3, Pacifique Ouest

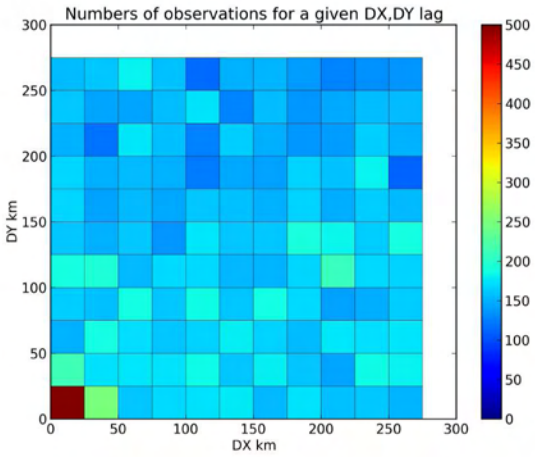


zone 4, Pacifique Est

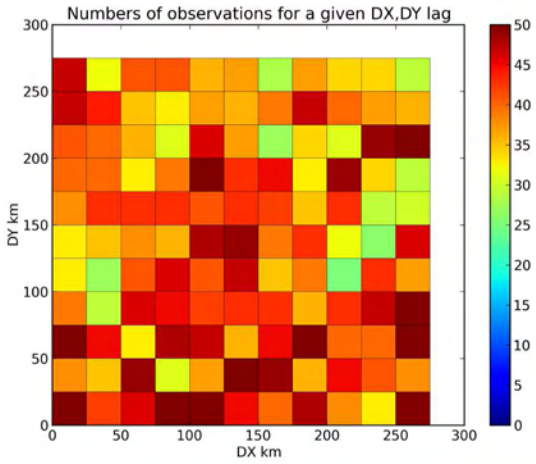


zone 5, Pacifique Ouest

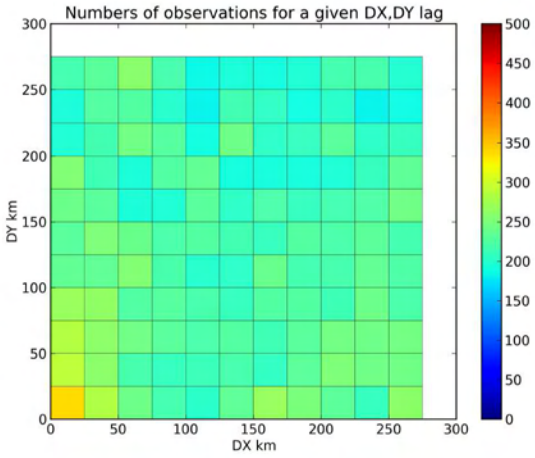
Figure B.1 – *Répartition des paires d'observations selon les classes de distances pour l'océan Pacifique, à une profondeur de* 200*m*



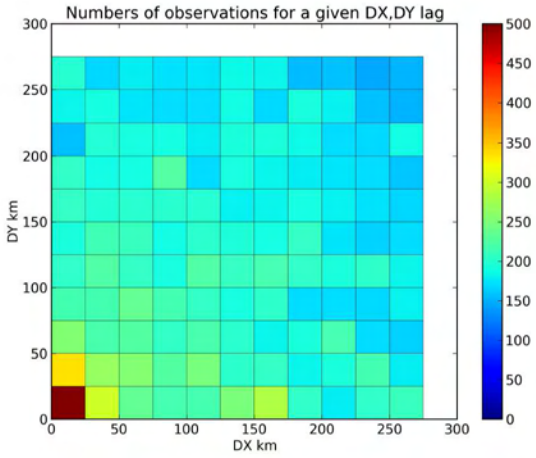
zone 7 Pacifique Est



zone 10 Pacifique Ouest

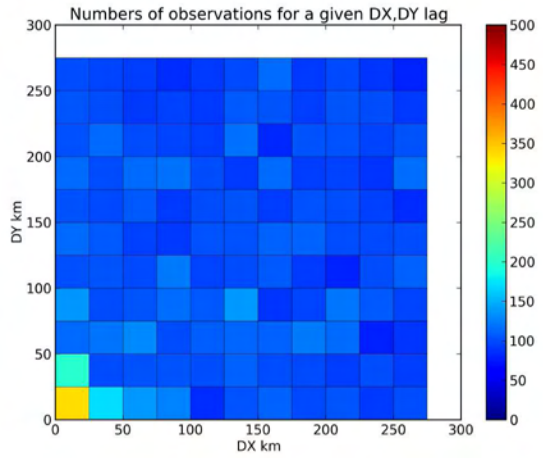


zone 11, Pacifique Est

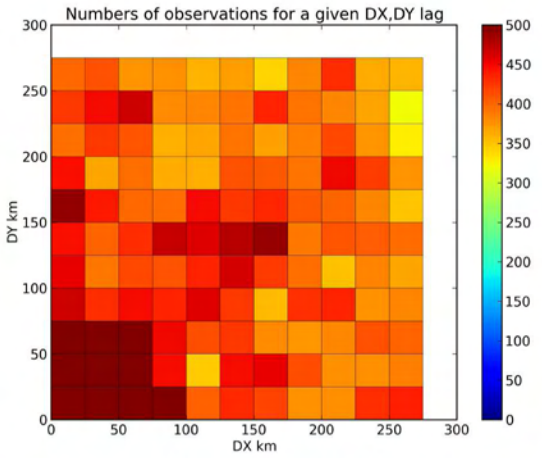


zone 14, Antarctique

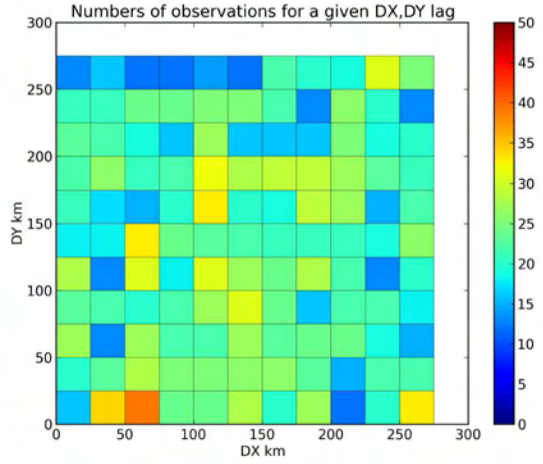
Figure B.2 – *Répartition des paires d'observations selon les classes de distances pour les océans Pacifique Sud et Antarctique*



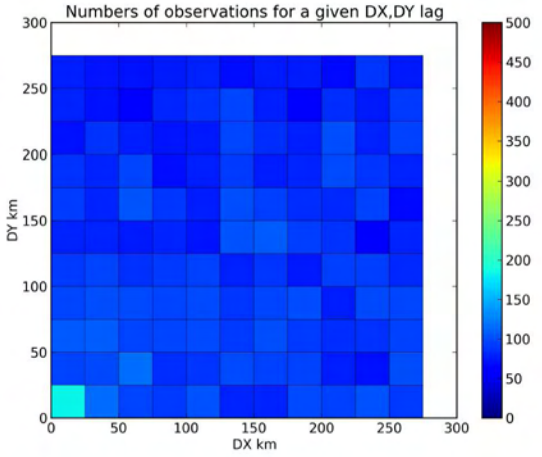
zone 2, Atlantique Nord



zone 6, Equateur

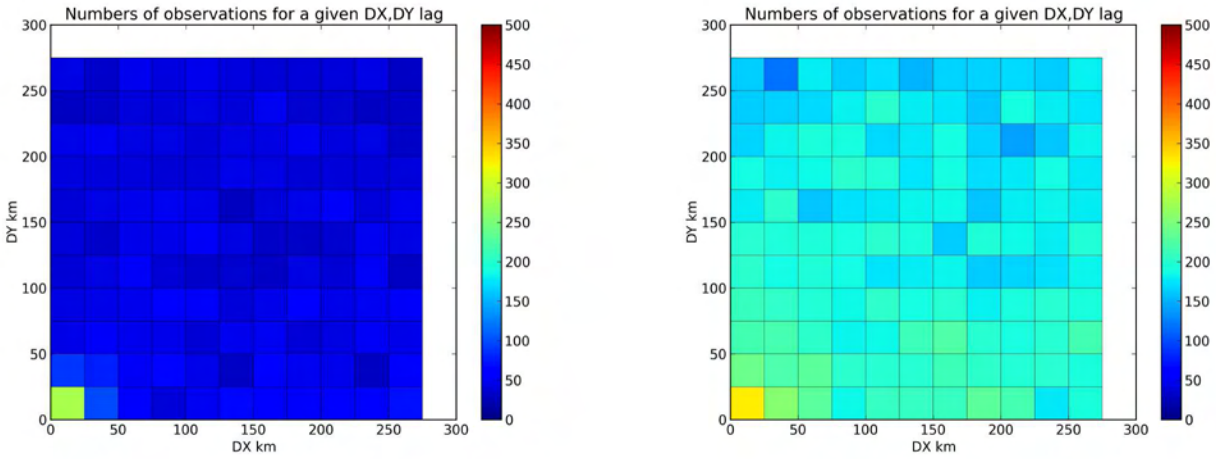


zone 8, Atlantique Sud



zone 12 Atlantique Sud

Figure B.3 – *Répartition des paires d'observations selon les classes de distances pour les régions situées dans l'océan Atlantique Sud*



zone 9, Indien *zone 13, Indien*

Figure B.4 – *Répartition des paires d'observations selon les classes de distances pour les régions situées dans l'océan Indien*

Structure verticale des échelles spatiales en température et salinité

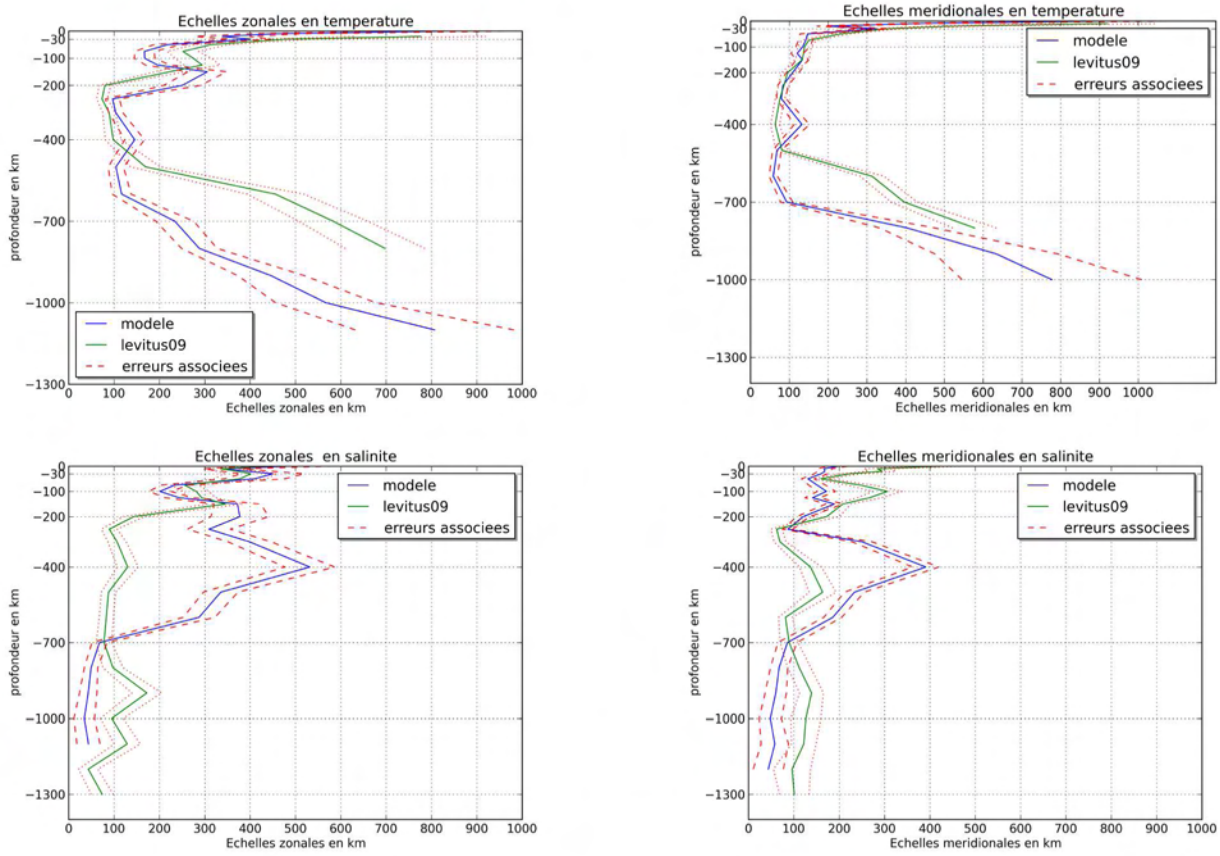


Figure B.5 – *Corrélation estimée en température (haut) et en salinité (bas) à partir des erreurs modèles (bleu) et pour le signal [climatologie − observations] (vert) dans l'océan Nord-Est Pacifique (zone 1). Les incertitudes liées à l'ajustement des échelles sont aussi représentées (pointillés rouge)*

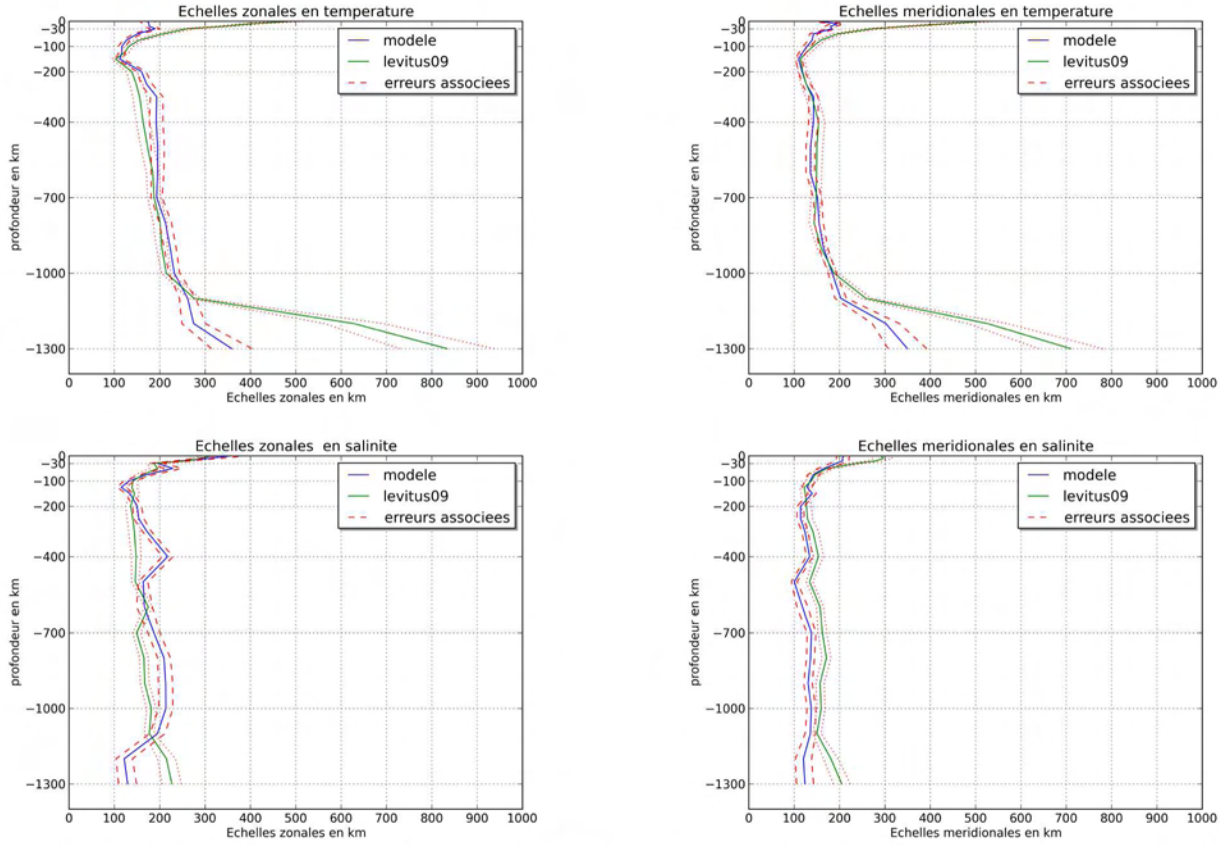


Figure B.6 – *Corrélation estimée en température (haut) et en salinité (bas) à partir des erreurs modèles (bleu) et pour le signal [climatologie − observations] (vert) dans l'océan Nord-Ouest Pacifique (zone 3). Les incertitudes liées à l'ajustement des échelles sont aussi représentées (pointillés rouge)*

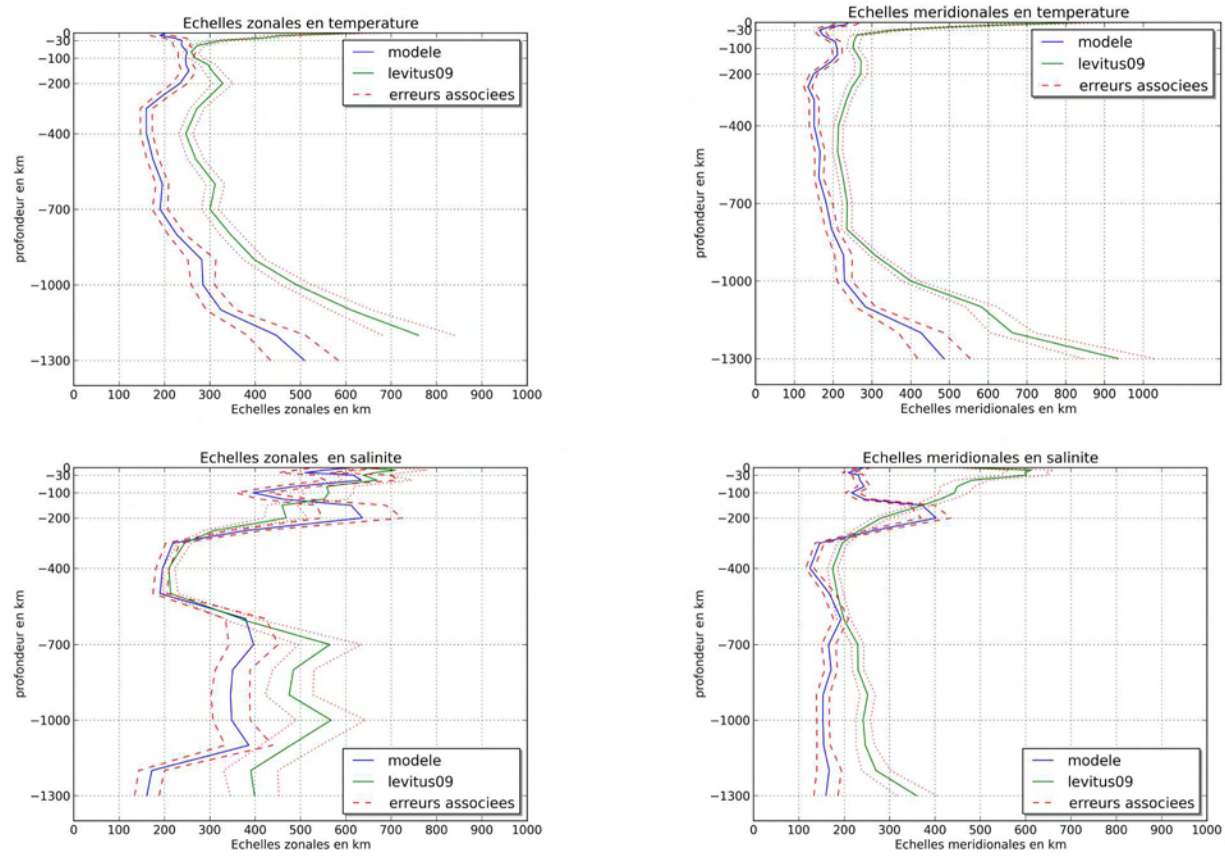


Figure B.7 – *Corrélation estimée en température (haut) et en salinité (bas) à partir des erreurs modèles (bleu) et pour le signal [climatologie − observations] (vert) dans l'océan Nord-Ouest Pacifique (zone 5). Les incertitudes liées à l'ajustement des échelles sont aussi représentées (pointillés rouge)*

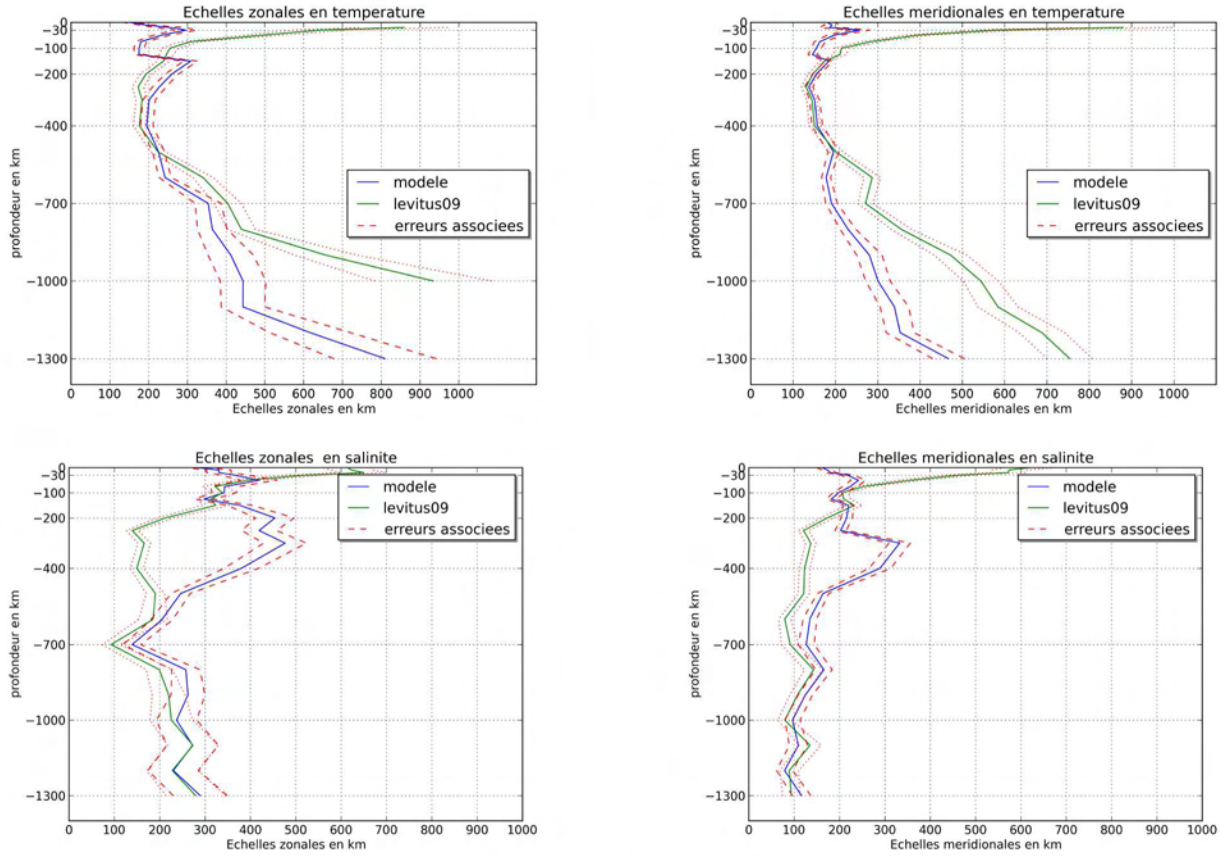


Figure B.8 – *Corrélation estimée en température (haut) et en salinité (bas) à partir des erreurs modèles (bleu) et pour le signal [climatologie – observations] (vert) dans l'océan Sud-Est Pacifique (zone 7). Les incertitudes liées à l'ajustement des échelles sont aussi représentées (pointillés rouge)*

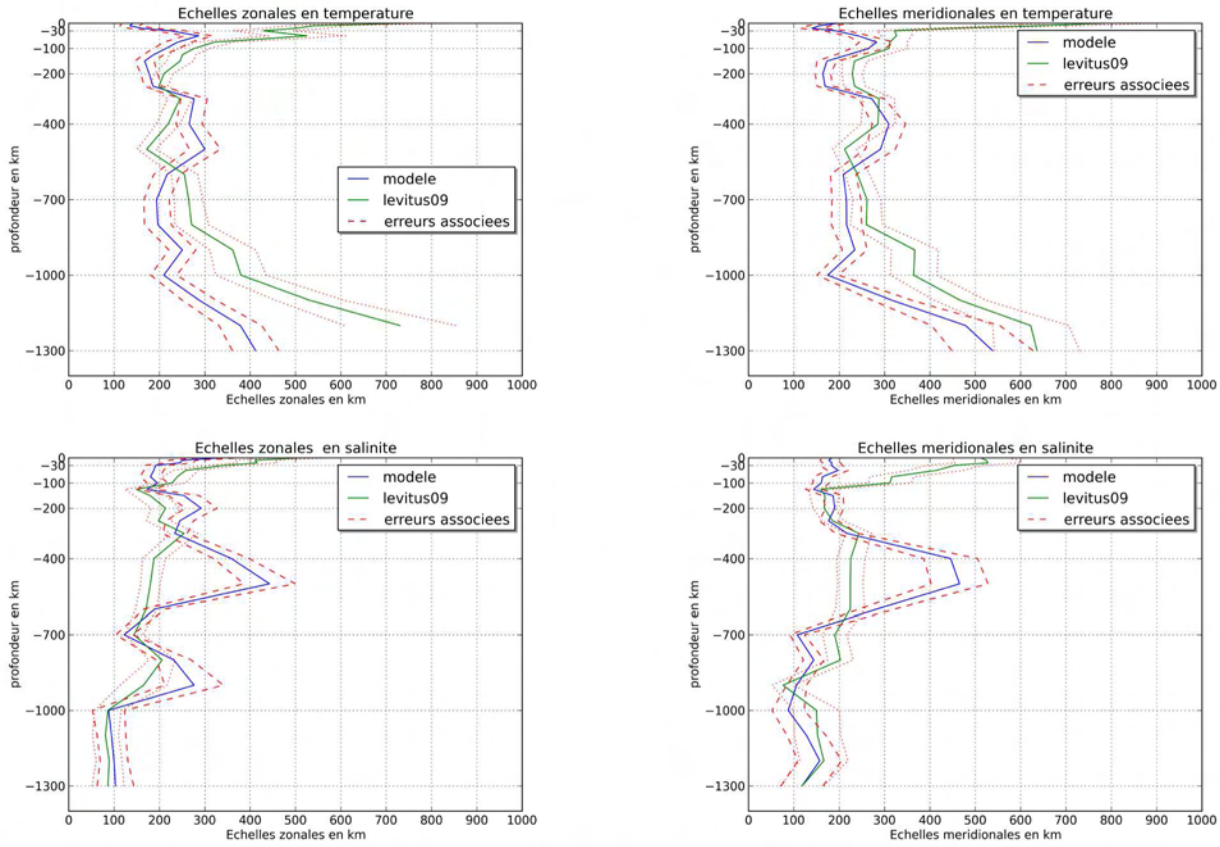


Figure B.9 – *Corrélation estimée en température (haut) et en salinité (bas) à partir des erreurs modèles (bleu) et pour le signal [climatologie – observations] (vert) dans l'océan Sud indien (zone 9). Les incertitudes liées à l'ajustement des échelles sont aussi représentées (pointillés rouge)*

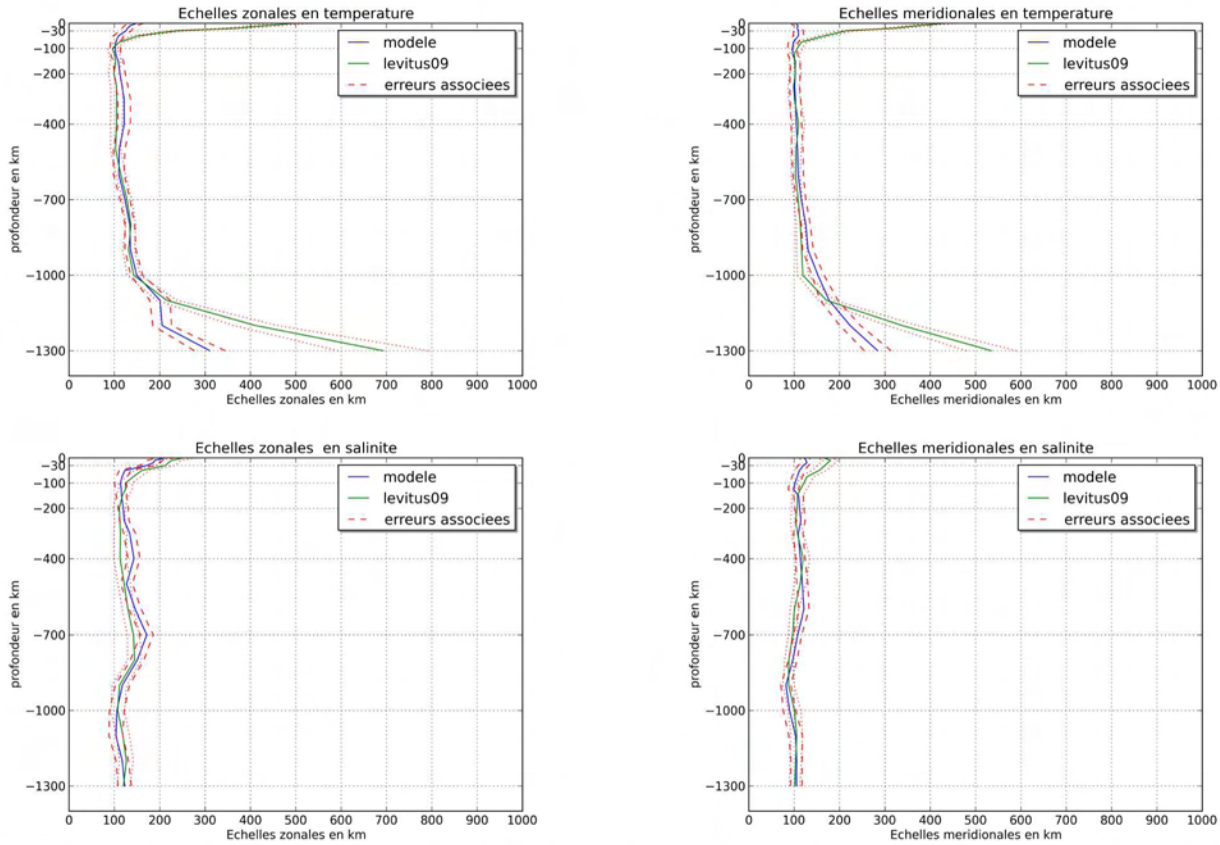


Figure B.10 – *Corrélation estimée en température (haut) et en salinité (bas) à partir des erreurs modèles (bleu) et pour le signal [climatologie – observations] (vert) dans l'océan Austral (zone* 13*). Les incertitudes liées à l'ajustement des échelles sont aussi représentées (pointillés rouge)*

Annexe C

Méthode d'assimilation BLUE

Approche du BLUE

Rappellons brièvement l'approche du BLUE qui permet d'obtenir le meilleur estimateur non biaisé du problème linéaire de l'assimilation. En pratique, il s'agit de trouver la meilleure combinaison entre un état à priori du système que l'on appelle l'ébauche, et les observations (*Gelb* [1974], *Bouttier and Courtier* [1999]). Le critère d'optimalité pour obtenir le meilleur estimateur statistique de cette combinaison est la détermination du minimum de variance d'erreur d'estimation. Les informations issues de l'ébauche et des observations sont entachées d'erreur. La combinaison des deux par le biais de divers algorithmes d'assimilation (BLUE, filtre de Kalman notamment) permet de se rapprocher de l'état réel du système. Ainsi, l'algorithme d'assimilation identifie à un instant t donné, un état analysé $\mathbf{x}^a$ de la variable aléatoire $\mathbf{x}$ dépendant linéairement de l'ébauche $\mathbf{x}^b$ et des observations $\mathbf{y}^o$. Cherchons à déterminer les matrices $\mathbf{L}$ et $\mathbf{K}$.

$$\mathbf{x}^a = \mathbf{L}\mathbf{x}^b + \mathbf{K}\mathbf{y}^o \tag{C.1}$$

L'état vrai $\mathbf{x}^t$ est approximé via l'état analysé $\mathbf{x}^a$. Il vient

$$\mathbf{x}^b = \mathbf{x}^t + \epsilon^b$$

$$\mathbf{y}^o = \mathbf{H}\mathbf{x}^t + \epsilon^o$$

$$\mathbf{x}^a = \mathbf{x}^t + \epsilon^a$$

Où ϵ^b, ϵ^o et ϵ^a sont les erreurs liées à l'ébauche, aux observations et à l'analyse. $\mathbf{H}$ est l'opérateur d'observation permettant de passer de l'espace du modèle à celui des observations. Enfin, l'estimateur linéaire recherché est non biaisé, cela introduit des contraintes sur $\mathbf{L}$ et $\mathbf{K}$:

$$\mathbb{E}[\epsilon^a] = \mathbb{E}[(\mathbf{x}^a - \mathbf{x}^t)] = \mathbb{E}[\mathbf{L}\mathbf{x}^t + \mathbf{L}\epsilon^b + \mathbf{K}\mathbf{H}\mathbf{x}^t + \mathbf{K}\epsilon^o - \mathbf{x}^t] = 0 \quad \text{(C.2)}$$

De plus les erreurs d'observation et d'ébauche sont supposées non biaisées, l'équation C.2 devient alors

$$(\mathbf{L} - \mathbf{I})\,\mathbb{E}[\mathbf{x}^t] + \mathbf{K}\mathbf{H}\mathbb{E}[\mathbf{x}^t] = (\mathbf{L} - \mathbf{I} + \mathbf{K}\mathbf{H})\,\mathbb{E}[\mathbf{x}^t] = 0 \quad \text{(C.3)}$$

La condition pour que l'estimateur $\mathbf{x}^a$ soit non biaisé devient donc :

$$\mathbf{L} + \mathbf{K}\mathbf{H} = \mathbf{I}\,^1 \quad \text{(C.4)}$$

On souhaite déterminer $\mathbf{x}^a$ tel qu'il soit le plus près de l'état vrai $\mathbf{x}^t$ au sens optimal. En d'autres termes l'estimateur optimal minimise la trace de la matrice de covariance d'erreurs d'analyse $\mathbf{A} = \mathbb{E}[\epsilon^a \epsilon^{aT}]$. En développant cette dernière on exprime

$$\mathbf{A} = (\mathbf{I} - \mathbf{K}\mathbf{H})\,\mathbf{B}\,((\mathbf{I} - \mathbf{K}\mathbf{H}))^T + \mathbf{K}\mathbf{R}\mathbf{K}^T$$

La minimisation de la trace de $\mathbf{A}$ permet d'établir une expression de la matrice $\mathbf{K}$

$$min(Tr(\mathbf{A})) \Rightarrow \mathbf{K} = \mathbf{B}\mathbf{H}^T(\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R})^{-1} \quad \text{(C.5)}$$

Ce qui nous mène enfin à l'expression

$$\mathbf{A} = (\mathbf{I} - \mathbf{K}\mathbf{H})\,\mathbf{B} \quad \text{(C.6)}$$

En utilisant l'équation (C.4) dans (C.1), on écrit l'état analysé tel que

$$\mathbf{x}^a = \mathbf{x}^b + \mathbf{K}\left(\mathbf{y}^o - \mathbf{H}\mathbf{x}^b\right) = \mathbf{x}^b + \mathbf{B}\mathbf{H}^T(\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R})^{-1}\left(\mathbf{y}^o - \mathbf{H}\mathbf{x}^b\right) \quad \text{(C.7)}$$

La minimisation de la variance de l'écart entre l'analyse $\mathbf{x}^a$ et l'état vrai permet d'éxiber l'estimateur du BLUE sous la forme d'une correction de l'ébauche par un terme proportionnel à l'écart entre les observations. La matrice $\mathbf{K}$, appelée matrice de gain, permet de définir l'état assimilé comme la somme de l'ébauche et de l'incrément d'analyse (terme correctif)

$$\delta\mathbf{x}^a = \mathbf{K}\left(\mathbf{y}^o - \mathbf{H}\mathbf{x}^b\right) = \mathbf{K}d \quad \text{(C.8)}$$

avec d le vecteur d'innovation qui représente la mesure résiduelle dans l'espace des observations.

1. $\mathbf{I}$ est la matrice identitée

Bibliographie

Allard, D., Statistiques spatiales : introduction à la géostatistique, 2012.

Antonov, J., D. Seidov, R. Locarnini, A. V. Mishonov, H. E. Garcia, O. K. Baranova, M. M. Zweng, and D. R. Johnson, World ocean atlas 2009, volume 2 : Salinity., *NOAA Atlas*, *NESDIS 68*, 184pp, 2010.

Arhan, M., and A. C. D. Verdière, Dynamics of eddy motions in the eastern north atlantic, *Journal Phys. Oceanogr.*, *15*, 1985.

Bell, M., P. LeTraon, E. Dombrowsky, A. Schiller, and K. Wilmer-Becker, Godae oceanview : From an experiment towards a long term international ocean analysis and forecasting program, 2010.

Bendat, J., and A. Piersol, *Random Data, Anlysis and measurment procedures*, Wiley-Interscience Publication, 1986.

Beuvier, J., Modélisation de la variabilité climatique de la circulation et des masses d'eau en mer méditerranée : impact des échanges océan-atmosphère, Ph.D. thesis, Ecole polytechnique (France), 2011.

Bouttier, F., Assimilation de données, 2004.

Bouttier, F., and P. Courtier, Data assimilation concepts and methods, 1999.

Bretherton, F., R. Davis, and C. B. Fandry, A technique for objective analysis and design of oceanographic experiments applied to mode-73, *Deep-Sea Research*, *23*, 559–582, 1976.

Cabanes, C., et al., The cora dataset : validation and diagnostics of in situ ocean temperature and salinity measurements, *Ocean Science*, *9*, 1–18, 2013.

Cardinali, C., S. Pezzulli, and E. Andersson, Influence-matrix diagnostic of a data assimilation system, *Quart. J. Roy. Meteorol. Soc*, *130*, 2767–2786, 2004.

Chapnik, B., G. Desroziers, F. Rabier, and O. Talagrand, Diagnosis and tuning of observational error in a quasi-operational data assimilation setting, *Quart. J. Roy. Meteorol. Soc*, *132*, 543–565, 2006.

Claustre, H., et al., Guidelines towards an integrated ocean observation system for ecosystems and biogeochemical cycles., in *Proceedings of the OceanObs'09 :Sustained Ocean Observations and Information for Society Conference (Vol. 1), Venice, Italy,21-25 September 2009, Hall, J., Harrison D. E. and Stammer, D., Eds., ESA Publication WPP-306*, 2010.

Dee, D., et al., The era-interim reanalysis : configuration and performance of the data assimilation system, *Quarterly Journal of the Royal Meteorological Society*, *137*, 553–597, 2011.

Desroziers, G., and B. Chapnik, Diagnosis and adaptive tuning of information error parameters in a variational assimilation, *Quart. J. Roy. Meteorol. Soc*, *127*, 1433–1452, 2001.

Desroziers, G., and S. Ivanov, Diagnosis and adaptive tuning of observation-error parameters in a variational assimilation, *Quart. J. Roy. Meteorol. Soc*, *127*, 1433–1452, 2001.

Desroziers, G., L. Berre, B. Chapnik, and P. Poli, Diagnosis of observation, background and analysis-error statistics in obeservation space, *Quart. J. Roy. Meteorol. Soc*, *131*, 3385–3396, 2005.

Dexter, P., and C. Summerhayes, *Troubled waters : Ocean Science and Governance*, vol. 11, david pugh and coeff holland ed., 2010.

Dibarboure, G., M. Pujol, F. Briol, P. L. Traon, G. Larnicol, N. Picot, F. Mert, and M. Ablain, Jason-2 in duacs : first tandem results and impact on processing and products, *Marine Geodesy*, *34*, 214–241, 2011.

Ducet, N., P.-Y. L. Traon, and G. Reverdin, Global high resolution mapping of ocean circulation from topex/poseidon an ers-1 and -2, *J. Geophys. Res.*, *105*, 19 447–19 498, 2000.

Durack, P., and S. Wijffels, Fifty-year trends in global ocean salinities and their relationship to broad-scale warming, *J. Climate*, *23*, 434,342–43,622, 2010.

Ferry, N., et al., Glorys2v1 global ocean reanalysis of the altimetric era (1992-2009) at meso scale, *Mercator Newsletter.*, *44*, 29–39, 2012.

Fichefet, T., and M. Maqueda, Sensitivity of a global sea ice model to the treatment of ice thermodynamics and dynamics, *Journal of Geophysical Research*, *102*, 12,609–12,646, 1997.

Freeland, H., et al., Argo - a decade of progress., *OceanObs'09 : Sustained Ocean Observations and Information for Society*, *2*, 2010.

Gaillard, F., E. Autret, V. Thierry, P. Gallaup, C. Coatanoan, and T. Loubrieu, Quality control of a large argo datasets, *J. Atmos Oceanic Technol*, *26*, 337–351, 2009.

Gandin, L., The objective analysis of meteorological fields, *Hydrometeorological Publ. House.*, p. 242 pp, 1963.

Garet, O., and A. Kurtzmann, *De l'intégration aux probabilités*, vol. 13, ellipses ed., 2011.

Gavin, P., The levenberg-marquart method for nonlinear least squares curve-fitting problems, 2013.

Gelb, A., Applied optimal estimation, 1974.

Girard, A., A fast monte carlo cross-validation procedure for large least problems with noisy data, *Tech. rep.*, IMAG,Grenoble,France, 1987.

groupe Tourbillon, L., The tourbillon experiment : a study of a mesoscale eddy in the eastern north atlantic, *Deep-Sea Res.*, *30*, 475–511, 1978.

Guinehut, S., A.-L. Dhomps, G. Larnicol, and P.-Y. L. Traon, High resolution 3-d temperature and salinity fieds derived from in situ and satellite observations, *Ocean Science*, *8*, 2012.

Huffman, G., R. Adler, D. Bolvin, and G. Gu, Improving the global precipiation record : Pcp version 2.1, *Geophys. Res. Lett.*, *36*, 2009.

Jacobs, G., C. N. Barron, and R. C. Rhodes, Mesoscale characteristics, *Geophys. Res*, *106*, 19,581–19,595, 2001.

Juza, M., Modélisation numérique et observations de l'océan global : développement des interfaces, évaluation de simulations et de réseaux d'observations, investigations dynamiques, Ph.D. thesis, Université de Grenoble, 2011.

Juza, M., T. Penduff, and B. Barnier, How should the argo array be extended to better monitor the global ocean heat content variability ?, *Newsletter Mercator, special issue with Coriolis*, *41*, 2011a.

Juza, M., T. Penduff, J. Brankart, and B. Barnier, Estimating the distortion of mixed layer property distributions induced by the argo sampling, *Journal of Operational Oceanography*, *15*, 45–58, 2011b.

Kallberg, P., Forecast drift in era-interim, *ECMWF ERA Report*, 2011.

Kamenkovich, I., W. Cheng, E. S. Sarachik, and D. Harrison, Simulation of the argo observing system in an ocean general circulation model, *J. Geophys Res.*, *114*, C09,021, 2009.

Kamenkovich, I., W. Cheng, C. Schmid, and D. E. Harrison, Effects of eddies on ocean observing system with profiling floats : Idealized simulations of the argo array, *J. of Geo. Research*, *116*, C06,003, 2011.

Kessler, W., M. C. Spillane, M. J. McPhaden, and D. E. Harrison, Scales of variability in the equatorial inferred from tropical atmosphere-ocean buoy array, *J. of Climate*, *9*, 2999–3024, 1996.

Large, W., and S. Yeager, The global climatology of an inter-annually varying air-sea flux data set, climate dynamics, *33*, 341–364, 2009.

Lellouche, J., et al., Evaluation of global monitoring anf forecasting systems at mercator ocean, *Ocean Sci.*, *9*, 1–25, 2013.

LeTraon, P., Y. Faugere, F. Hernandez, J. Dorandeu, F. Mertz, and M. Ablain, Can we merge geosat follow-on with topex/poseidon and ers-2 for an improved description of the ocean circulation ?, *American Meteor. Society*, *20*, 889–895, 2003.

Levenberg, K., A method for the solution of certain non-linear problems in least squares, *The Quartely of Applied Mathematics*, *2*, 1944.

Levitus, S., J. Antonov, T. Boyer, M. Conkright, T. O'Brien, C. Stephens, L. Stathoplos, D. Johnson, and R. Gelfeld, World ocean data base, *NOAA Atlas*, *NESDIS 18*, 1998.

Levitus, S., J. Antonov, and T. Boyer, Warming the world ocean, 1955-2003, *Geophysical Research Letters*, *32*, LO2604, 2005.

Levitus, S., et al., World ocean heat content and thermosteric sea level change (0–2000 m), 1955–2010, *Geophysical Research Letters*, *39*, 2009.

Levitus, S., et al., World ocean database, *Data Science Journal*, *12*, 2013.

Locarnini, R., A. V. Mishonov, J. Antonov, T. P. Boyer, H. E. Garcia, O. K. Baranova, M. M. Zweng, and D. R. Johnson, World ocean atlas 2009, volume 1 : Temperature., *NOAA Atlas*, *NESDIS 68*, 184pp, 2010.

Lupu, C., and P. Gauthier, Evaluation of the impact of observations on analyses in 3d- and 4d-var based on information content, *Monthly Weather Review*, *139*, 2010.

Madec, G., Nemo reference manual, ocean dynamics component : Nemo-opa. preliminary version, *Note du Pôle de modélisation, Institut Pierre-Simon Laplace (IPSL)*, pp. ISSN No 1288–1619, 2008.

Madec, G., P. Delecluse, M. Imbard, and C. Levy, Opa 8.1 ocean general circulation model reference manual, *Notes de l'IPSL, Université P. et M. Curie*, *11*, 91, 1998.

Martin, M., D. Lea, and P. R. Oke, Demonstrating the complementary of observations in an operational ocean forecasting system, *Quaterly Journal of the Royal Meteorological Society*, 2014.

Oke, P., and Schiller, Impact of argo, sst, and altimeter data on an eddy-resolving ocean reanalysis, *Geophysical Research Letters*, *34(19)*, L19,601, 2007.

Oke, P., et al., Observing system evaluations using godae systems, *Oceanography (The Oceanography Society)*, *22*, 144–153, 2009.

Oke, P., et al., Assessing the impact of observations on ocean forecasts and reanalyses : Part 1, global studies, 2014.

Pham, D., J. Verron, and M. C. Roubaud, A singular evolutive extended kalman filter for data assimilation in oceanography, *J. Mar. Syst*, *16*, 323–340, 1998.

Rabier, F., Overview of a global data assimilation developments in numerical weather-prediction centres, *Quart. J. Roy. Meteorol. Soc*, *131*, 3215–3233, 2005.

Roemmich, D., and J. Gilson, The 2004-2008 mean and annual cycle of temperature, salinity, and steric height in the global ocean from the argo program, *Progress in Oceanography*, *82*, 81–100, 2010.

Roemmich, D., and the Argo Steering Team, Argo, the challenge of continuing 10 years of progress, *Oceanography (The Oceanography Society)*, *22*, 46–55, 2009.

Smith, G., and K. Haines, Evaluation of the s(t) assimilation method with the argo dataset., *Quarterly Journal of the Royal Meteorological Society*, *135*, 739–756, 2009.

Stackhouse, J., W. Paul, K. S. Gupta, J. C. Stephen, J. C. Mikovitz, T. Thang, and L. Hinkelman, The nasa/gewex surface radiation budget release 3.0 : 24.5-year, *GEWEX News*, *21, No1 February*, 10–12, 2011.

Turpin, V., E. Remy, and P. L. Traon, Argo impact studies with the mercator 1/4° global ocean data assimilation system, *Ocean Science (submitted)*, 2015.

Vecchi, G., and M. J. Harrison, An observing system simulation experiment for the indian ocean, *J. of Climate*, *20*, 3300–3319, 2005.

Wong, A., R. Keely, T. Carval, and the Argo Data Management Team, Argo quality control manual version 2.6, *http ://dx.doi.org/10.13155/33951*, 2014.

Table des figures

1.1 Système Global d'Observation des Océans (GOOS) et état de développement des différentes composantes in-situ par rapport aux objectifs initiaux définis par la communauté océanographique il y a 15 ans. . 3

2.1 Positions des flotteurs Argo pour le mois de juin 2015 12
2.2 Flux de données Argo et acteurs en temps réel des traitements de données. 13

3.1 Nombre total d'observations en température sur la période 2006 − 2009 dans des boîtes de 4°x4° : en surface (a) , entre 1000 et 2000 mètres (b). 22
3.2 Nombre total d'observations en salinité sur la période 2006 − 2009 dans des boîtes de 4°x4° : en surface (a) , entre 1000 et 2000 mètres (b). 23
3.3 Application de la méthode de corrections : série temporelle de la salinité de surface avec correction des flux de précipitations (orange) et sans (bleu), en comparaison à celle issue de la climatologie Levitus (pointillés rouges). Figure réalisée par G. Garric. 25
3.4 Nombre moyen annuel de flotteurs Argo disponibles dans des boîtes de 4°x4° sur la période 2006-2009 27
3.5 Echantillonnage pour plusieurs grands bassins des profileurs Argo, Année 2009. 28
3.6 Série temporelle des écarts moyens globaux modèle-observations sur la classe de profondeur [0-2000]m en température (en °C) et en salinité (en psu) sur la période 2006-2009. 30
3.7 Série temporelle des écarts moyens globaux modèle-observations sur la classe de profondeur [0-600]m en température (en °C) et salinité (en psu) sur la période 2006-2009. 31
3.8 Ecarts moyens en température (en °C) et salinité (en psu) de surface sur la période 2006 − 2009 32

3.9 Erreurs modèle moyennes en température (en °C) et en salinité (psu) sur la période 2006-2009, sur les couches de profondeur $dp1 = 0 - 300$m et $dp2 = 300 - 700$ m 33
3.10 Erreurs modèle moyennes en température (en °C) et en salinité (psu) sur la période 2006-2009, sur la couche de profondeur $dp3 = 700 - 2000$ m 34
3.11 Variance des erreurs modèle en température (en $°C^2$) et salinité (en psu^2), pour les couches $dp1 = [0 - 300]$m, $dp2 = [300 - 700]$m . . . 36
3.12 Ecarts moyens en température (en °C) entre le modèle et les données Argo, en surface pour les étés et hivers en $2006 - 2009$ 37
3.13 Ecarts moyens en température entre le modèle et les données Argo sur les couches $dp1 = [0 - 300]$m, $dp2 = [300 - 700]$m et $dp3 = [700 - 2000]$m pour les étés (gauche) et hivers (droite) 2006-2009 39
3.14 Ecarts moyens en salinité entre le modèle et les données Argo sur les couches $dp1 = [0 - 300]$m, $dp2 = [300 - 700]$m et $dp3 = [700 - 2000]$m pour les étés (gauche) et hivers (droite) 2006-2009 40

4.1 Echantillonnage sur la zone test de calcul des profileurs Argo, année 2008. 48
4.2 a) et b) Exemples de simulation pour un membre z^t, c) la matrice de covariance associée 49
4.3 Champs de corrélation estimés et théoriques (contours en noir) sur la zone $\mathbf{R_{zone20}}$ pour les réseaux $Argo_{ideal}$ et $Argo_{0609}$ 51
4.4 Nombre de profileurs répartis par classe distance sur la zone $\mathbf{R_{zone20}}$, pour a) $\triangle x_m = 25$km b) $\triangle x_m = 55$km pour le réseau $Argo_{0609}$ . . 53
4.5 Champ de corrélation pour $\triangle x_m = 25$km, dans les cas $L_p = 105$km et $L_g = 400$km. 54
4.6 Ensemble des zones où les échelles spatiales de corrélation sont estimées. 56
4.7 Répartition des paires d'observations selon les classes de distances pour les régions situées dans l'océan Pacifique, à 200 m sur la période $2006 - 2009$ 57
4.8 Répartition des paires d'observations selon les classes de distances pour les régions situées dans l'océan Atlantique, aux tropiques et aux hautes latitudes de l'hémisphère sud, à 200 m, sur la période $2006 - 2009$ 58
4.9 Corrélation estimée en température à partir des erreurs modèles à 200m et 1000m dans l'Océan Nord Pacifique sur la période $2006 - 2009$ 61
4.10 Echelles spatiales zonales (bleu) et méridionales (rouge) des erreurs modèles en température en fonction des latitudes à 200 m sur la période $2006 - 2009$. 62

4.11 Corrélation estimée en température (haut) et en salinité (bas) à partir des erreurs modèles (bleu) et pour le signal ([*climatologie* − *observations*], en vert) dans l'océan Atlantique Nord (zone 2). Les incertitudes liées à l'ajustement des échelles sont aussi représentées (pointillés rouge) 65
4.12 Corrélation estimée en température (haut) et en salinité (bas) à partir des erreurs modèles (bleu) et pour le signal ([*climatologie* − *observations*], en vert) dans l'océan Atlantique Sud (zone 12). Les incertitudes liées à l'ajustement des échelles sont aussi représentées (pointillés rouge) 66
4.13 Corrélation estimée en température (haut) et en salinité (bas) à partir des erreurs modèles (bleu) et pour le signal ([*climatologie* − *observations*] en vert) dans la région tropicale (zone 6). Les incertitudes liées à l'ajustement des échelles sont aussi représentées (pointillés rouge) 67
4.14 Corrélation estimée en température (haut) et en salinité (bas) à partir des erreurs modèles (bleu) et pour le signal ([*climatologie* − *observations*] en vert) aux hautes latitudes de l'hémisphère Sud (zone 14). Les incertitudes liées à l'ajustement des échelles sont aussi représentées (pointillés rouge) 68
4.15 Corrélation estimée en température (haut) et en salinité (bas) à partir des erreurs modèles (bleu) et pour le signal [*climatologie* − *observations*] (vert) dans l'océan Nord-Est Pacifique (zone 4). Les incertitudes liées à l'ajustement des échelles sont aussi représentées (pointillés rouge) 69
4.16 Corrélation estimée en température (haut) et en salinité (bas) à partir des erreurs modèles (bleu) et pour le signal [*climatologie* − *observations*] (vert) dans l'océan Sud-Est Pacifique (zone 11). Les incertitudes liées à l'ajustement des échelles sont aussi représentées (pointillés rouge) 70

5.1 Variation du DFS en fonction de la longueur de corrélation pour un bruit de mesure de 40% pour les 16 observations 99
5.2 Dfs partiels pour chaque observation pour une longueur de corrélation commune L = 150 km et un bruit à 0.36% (gauche) et à 47% (droite). 100
5.3 Variation du DFS en fonction des différentes valeurs de bruits testés (expériences 1 à 4). 102
5.4 Influence du bruit sur les champs vrai et estimé pour $L_1 < dx_{obs}$ et $L_2 > dx_{obs}$. Les contours en pointillés sont pour des valeurs négatives. Les observations (croix bleue) sont espacées régulièrement de $dx_{obs} = 80$ km. 103

5.5 Influence du bruit de mesure sur les erreurs théoriques, pour $L_1 < dx_{obs}$ et $L_2 > dx_{obs}$. Les contours non nommés sont à 0.3 ou 0.7. Les observations sont espacées régulièrement de $dx_{obs} = 80$ km. 104

5.6 Contenu en information des observations de température de surface (SST) en pourcentage pour une analyse du système global au 1/4° en août 2010, figure réalisée par E. Remy. 106

A.1 Erreurs modèle moyennes en température (en °C) sur la période 2006-2009 pour différents niveaux standards Levitus. 115

A.2 Erreurs modèle moyennes en salinité (en psu) sur la période 2006-2009 pour différents niveaux standards Levitus. 117

A.3 Ecarts moyens en température et salinité entre le modèle et les données Argo sur les couches $dp1 = [0 - 300]$m, $dp2 = [300 - 700]$m et $dp3 = [700 - 2000]$m pour les printemps $2006 - 2009$. 119

A.4 Ecarts moyens en température et salinité entre le modèle et les données Argo sur les couches $dp1 = [0 - 300]$m, $dp2 = [300 - 700]$m et $dp3 = [700 - 2000]$m pour les étés 2006-2009. 121

A.5 Ecarts moyens en température et salinité entre le modèle et les données Argo sur les couches $dp1 = [0 - 300]$m, $dp2 = [300 - 700]$m et $dp3 = [700 - 2000]$m pour les autumnes 2006-2009. 123

A.6 Ecarts moyens en température et salinité entre le modèle et les données Argo sur les couches $dp1 = [0 - 300]$m, $dp2 = [300 - 700]$m et $dp3 = [700 - 2000]$m pour les Hivers 2006-2009. 125

B.1 Répartition des paires d'observations selon les classes de distances pour l'océan Pacifique, à une profondeur de 200m 127

B.2 Répartition des paires d'observations selon les classes de distances pour les océans Pacifique Sud et Antarctique 128

B.3 Répartition des paires d'observations selon les classes de distances pour les régions situées dans l'océan Atlantique Sud 129

B.4 Répartition des paires d'observations selon les classes de distances pour les régions situées dans l'océan Indien 130

B.5 Corrélation estimée en température (haut) et en salinité (bas) à partir des erreurs modèles (bleu) et pour le signal [$climatologie - observations$] (vert) dans l'océan Nord-Est Pacifique (zone 1). Les incertitudes liées à l'ajustement des échelles sont aussi représentées (pointillés rouge) 131

B.6 Corrélation estimée en température (haut) et en salinité (bas) à partir des erreurs modèles (bleu) et pour le signal [*climatologie* − *observations*] (vert) dans l'océan Nord-Ouest Pacifique (zone 3). Les incertitudes liées à l'ajustement des échelles sont aussi représentées (pointillés rouge) 132
B.7 Corrélation estimée en température (haut) et en salinité (bas) à partir des erreurs modèles (bleu) et pour le signal [*climatologie* − *observations*] (vert) dans l'océan Nord-Ouest Pacifique (zone 5). Les incertitudes liées à l'ajustement des échelles sont aussi représentées (pointillés rouge) 133
B.8 Corrélation estimée en température (haut) et en salinité (bas) à partir des erreurs modèles (bleu) et pour le signal [*climatologie* − *observations*] (vert) dans l'océan Sud-Est Pacifique (zone 7). Les incertitudes liées à l'ajustement des échelles sont aussi représentées (pointillés rouge) 134
B.9 Corrélation estimée en température (haut) et en salinité (bas) à partir des erreurs modèles (bleu) et pour le signal [*climatologie* − *observations*] (vert) dans l'océan Sud indien (zone 9). Les incertitudes liées à l'ajustement des échelles sont aussi représentées (pointillés rouge) 135
B.10 Corrélation estimée en température (haut) et en salinité (bas) à partir des erreurs modèles (bleu) et pour le signal [*climatologie* − *observations*] (vert) dans l'océan Austral (zone 13). Les incertitudes liées à l'ajustement des échelles sont aussi représentées (pointillés rouge) 136

Liste des tableaux

2.1 Caractéristiques du système d'assimilation de données au sein du système $PSY3V3R3$ de Mercator-Océan 16

3.1 Caractéristiques de la simulation numérique $T335$ réaliste réalisée à Mercator-Océan 26

4.1 Longueur de corrélation et niveau de bruit de mesure pour chaque réseau utilisé, avec comme valeurs théoriques : L_p=105km, L_g=400km et $\mathfrak{e}_{\text{th}} = 0.10$ 51
4.2 Ordre de grandeur des écarts entre les corrélations estimées et théoriques dans le cas des réseaux $Argo_{ideal}$ et $Argo_{0609}$ 52
4.3 Analyse de sensibilité réalisée selon les différentes résolutions spatiales ($\triangle x_m$) en fonction du type de structures spatiales observées $L_p = 105$ km, $L_g = 400$ km et $\mathfrak{e}_{\text{th}} = 0.10$ 53
4.4 Echelles spatiales zonales (L_x) et méridionales (L_y) pour plusieurs bassins (zones), et bruit de mesure ($\mathfrak{e}$) des observations estimés à 200 et 1000 m. 61

5.1 Variation du DFS en fonction de la longueur de la corrélation minimale 40 km, maximale 240 km, pour un bruit de mesure de 40% pour les 16 observations 99
5.2 Influence du bruit de mesure pour $L_1 = 60km$ 101
5.3 Influence du bruit de mesure pour $L_2 = 120km$ 101
5.4 1^e expérience : variance d'erreurs d'observations parfaitement connue et observations faiblement bruitées 108
5.5 2^e expérience : variance d'erreurs d'observations sous-estimées et observations faiblement bruitées 108