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

Sommaire

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. $\mathbf{E}$ 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 :

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

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

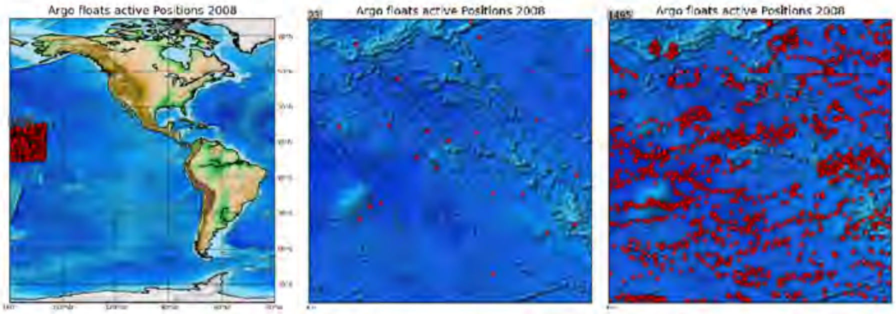
Mettre en pratique l'estimation statistique des échelles de corrélation spatiales a entrainé la réalisation de plusieurs tests afin de valider la méthodologie utilisée. Pour cela nous allons comparer l'estimation statistique de covariances spatiales à partir du réseau Argo des années 2006-2009, noté $Argo_{0609}$ 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 $Argo_{ideal}$

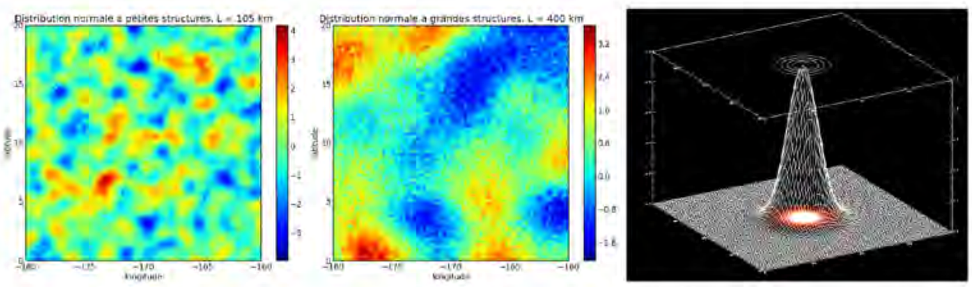
La construction d'un réseau d'observations idéal au 1/2° permet d'augmenter le nombre de profils appartenant à la zone $\mathbf{R_{zone}}$ par rapport à la réalité. Il est important de connaître la densité numérique moyenne par semaine car les statistiques sont réalisés sur cette période temporelle, 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 $\mathcal{Z} = \{\ z^t,\ t \in [0, N_{heb}]\}$, 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

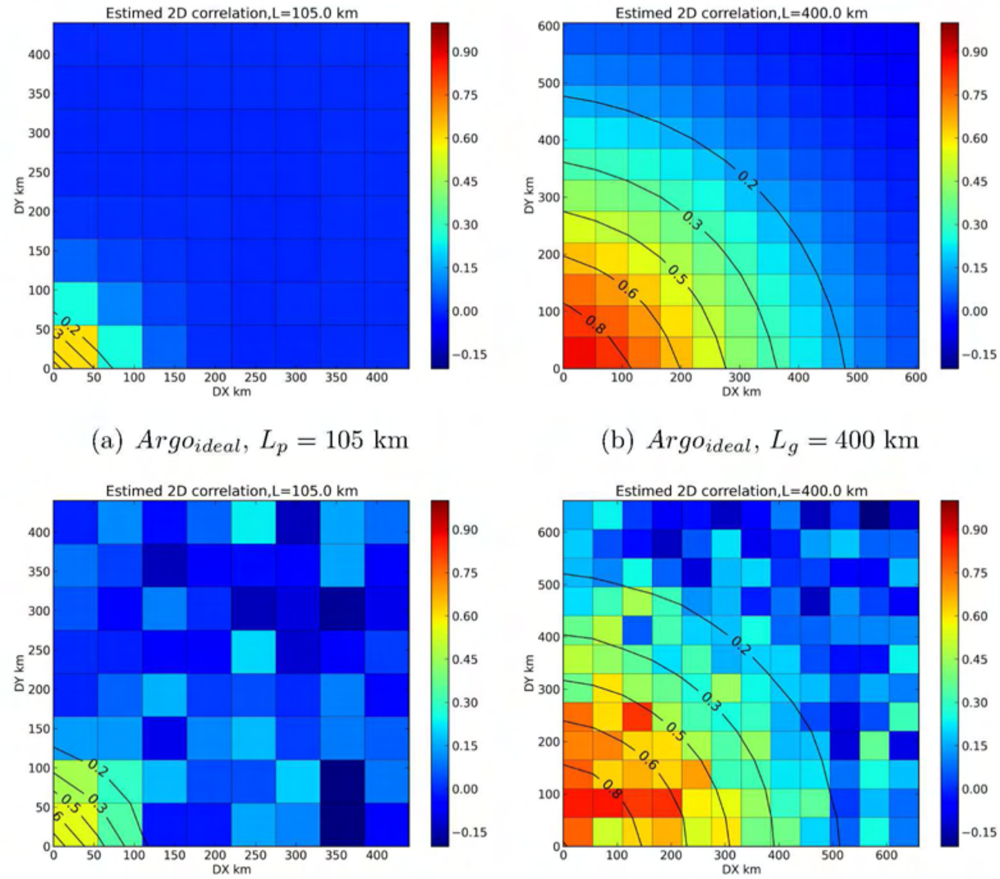
Figure 4.2 – *a) et b) Exemples de simulation pour un membre z^t, c) la matrice de covariance associée*

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* $\mathfrak{e}_{\mathfrak{th}} = 0.10$

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 $\mathscr{Z}$ 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.

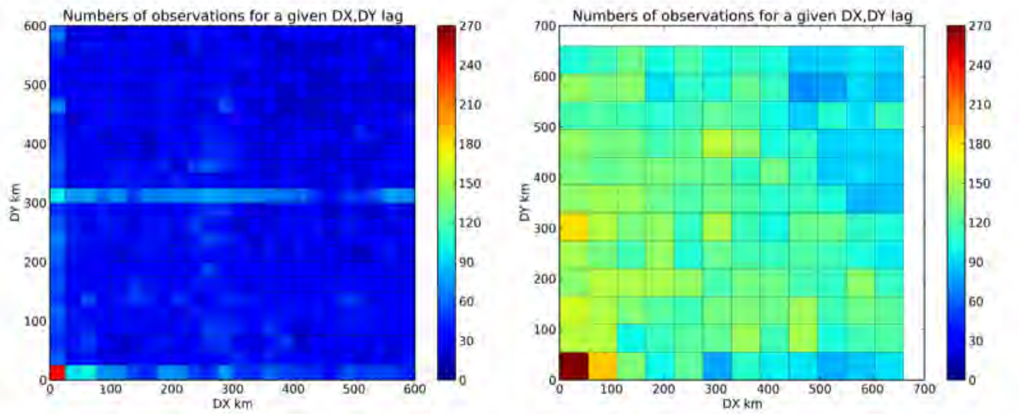
- La répartition des observations par classe pour les deux résolutions spatiales est homogène. Mais comme on peut s'y attendre elle est en moyenne environ 4 fois plus faible dans le cas $\triangle x_m = 25$ km (voir les figures 4.4, 4.5 b. et d.).
- Dans les deux cas $L_p = 105$ km et $L_g = 400$ km, 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 $\triangle x_m =$ 25km (et plus proche de celui prescrit en amont $\mathfrak{e}_{\mathfrak{th}} = 0.10$). 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.

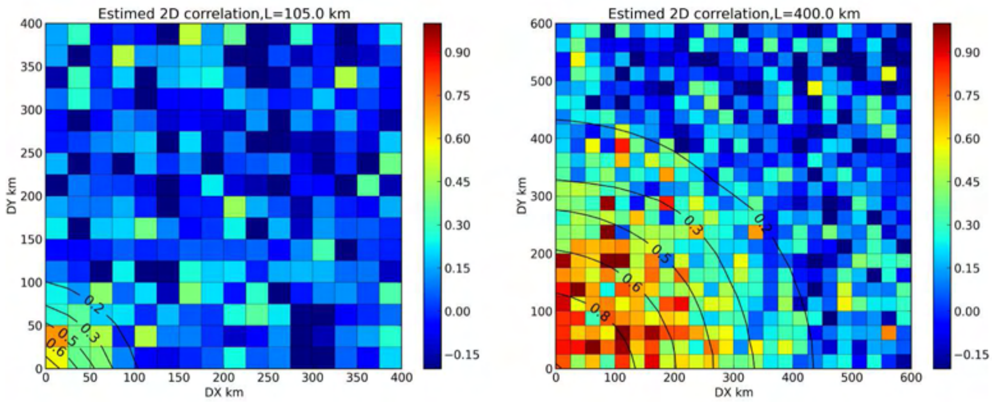
$\triangle x_m$	L_p	$\mathfrak{e}$
25	111.58 ± 13.47 km	0.56 ± 0.26
55	110.80 ± 13.20 km	0.61 ± 0.43
$\triangle x_m$	L_g	$\mathfrak{e}$
25	402.34 ± 11.24 km	0.14 ± 0.04
55	406.48 ± 11.62 km	0.16 ± 0.04
100	411.78 ± 11.61 km	0.19 ± 0.04

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* $\mathfrak{e}_{\mathfrak{th}} = 0.10$



(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.

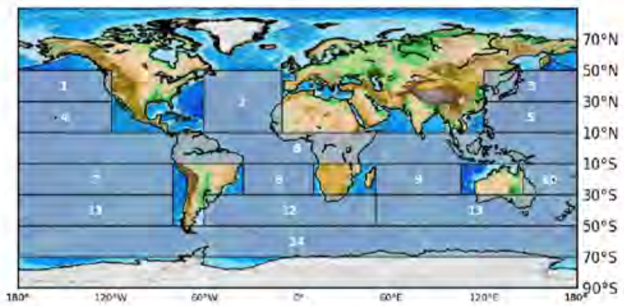
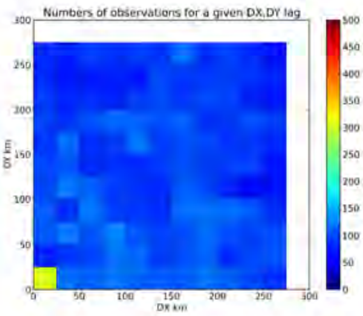
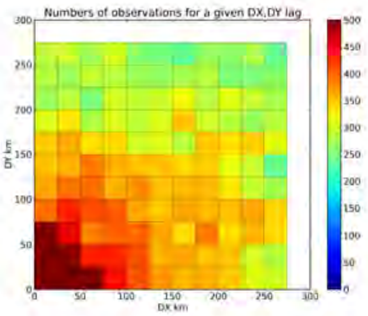


Figure 4.6 – *Ensemble des zones où les échelles spatiales de corrélation sont estimées.*

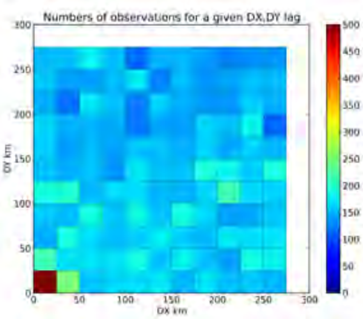
Les figures 4.7 et 4.8 représentent le nombre de paires de profils par classe de distance (N, défini en section 4.2) pour plusieurs zones situées dans l'océan Pacifique, Atlantique, au latitudes tropicales et aux hautes latitudes de l'hémisphère sud cumulés sur la période 2006 – 2009. On note une densité 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-



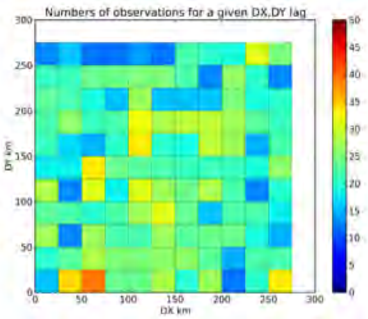
(a) zone 1, Pacifique Nord-Est



(b) zone 3, Pacifique Nord-Ouest



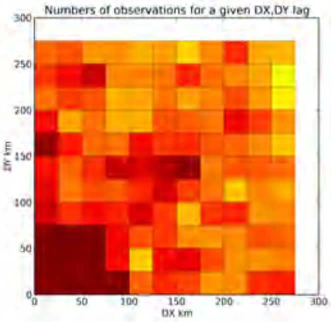
(c) zone 7, Pacifique Sud-Est



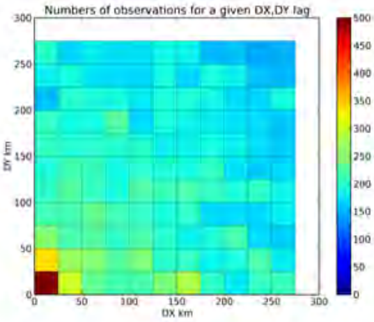
(d) zone 10, Pacifique Sud-Ouest

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

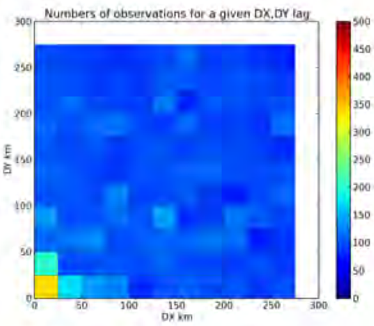
ception des zones 8 et 10 dont la densité d'observations est faible, la répartition des paires de profils des zones examinées est continue et décroissante lorsque la distance augmente. Les zones 8 et 10 ne seront pas prises en compte lors de l'étude, et confirment la nécessité de travailler avec des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



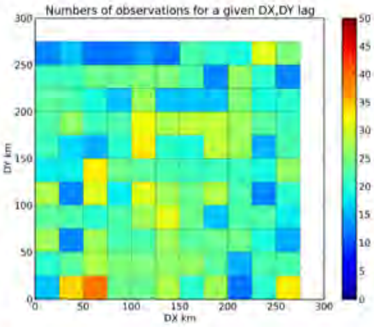
(a) zone 6, Equateur



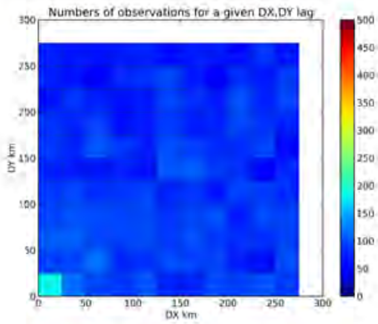
(b) zone 14, Antarctique



(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.

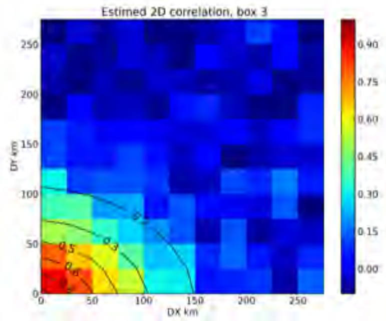
Dans les sections suivantes nous allons rendre compte des tendances globales des variations des échelles spatiales en température et en salinité issues des erreurs modèles (section 4.3.3). Ces estimations seront ensuite comparées à celles issues des anomalies par rapport à la climatologie mensuelle Levitus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

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

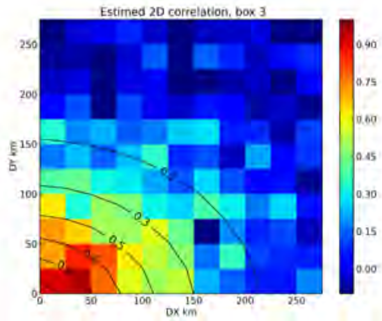
A titre d'exemple, les figures 4.9a et 4.9b présentent les champs de corrélation spatiales en température pour une zone située dans l'océan Nord Pacifique à 200 et 1000 m. Comme décrit en section 4.3.1 la zone 3 dispose d'un large nombre de paires de profils permettant d'obtenir des estimations statistiques 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.

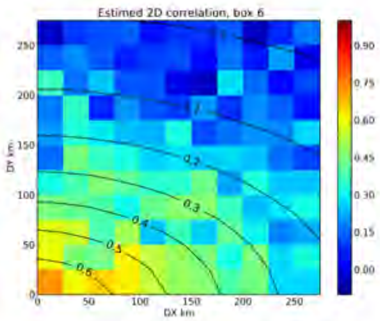
Les figures 4.12, 4.15 et 4.16 présentent les distributions verticales des échelles zonales et méridionales estimées à partir des erreurs modèles (courbe bleu) pour plusieurs zones. Pour des profondeurs supérieures à 800-1000 m, les échelles spatiales zonales et méridionales en température ont tendance 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.*

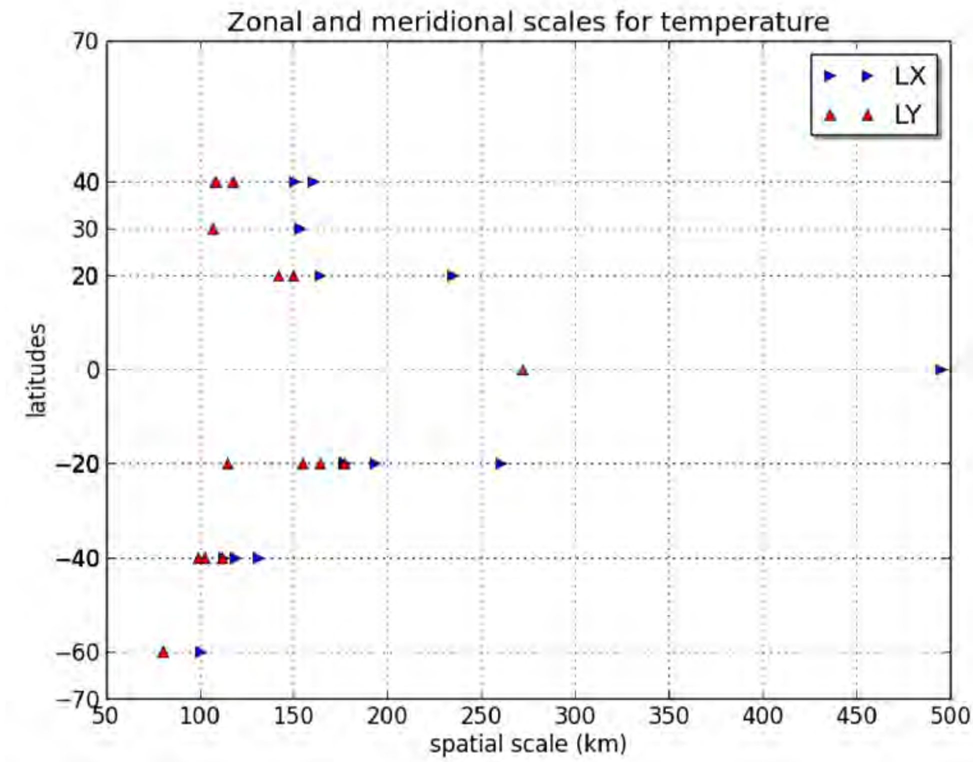


Figure 4.10 – *Echelles spatiales zonales (bleu) et méridionales (rouge) des erreurs modèles en température en fonction des latitudes à* 200 *m sur la période* 2006−2009.

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]).

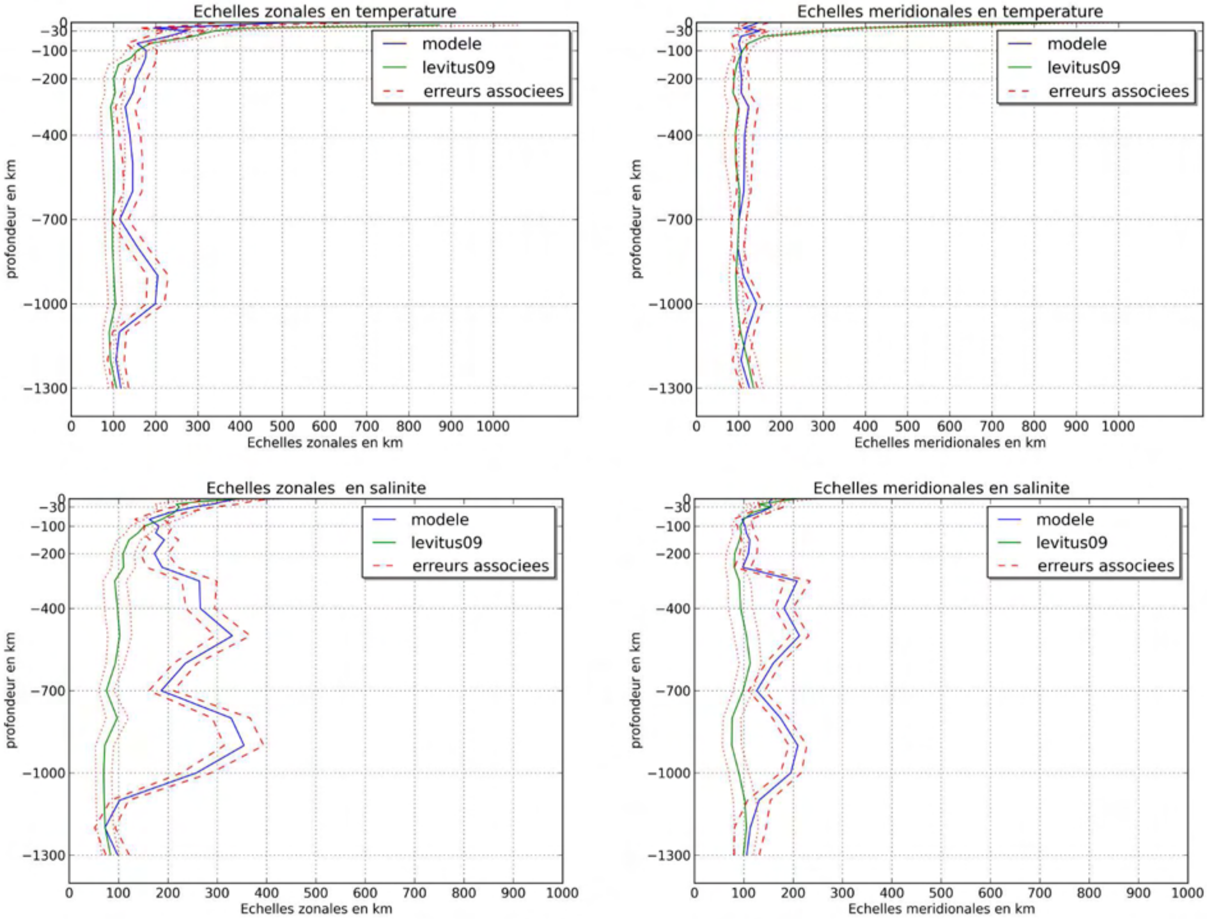


Figure 4.11 – *Corrélation estimée en température (haut) et en salinité (bas) à partir des erreurs modèles (bleu) et pour le signal ([climatologie – observations], en vert) dans l'océan Atlantique Nord (zone 2). Les incertitudes liées à l'ajustement des échelles sont aussi représentées (pointillés rouge)*

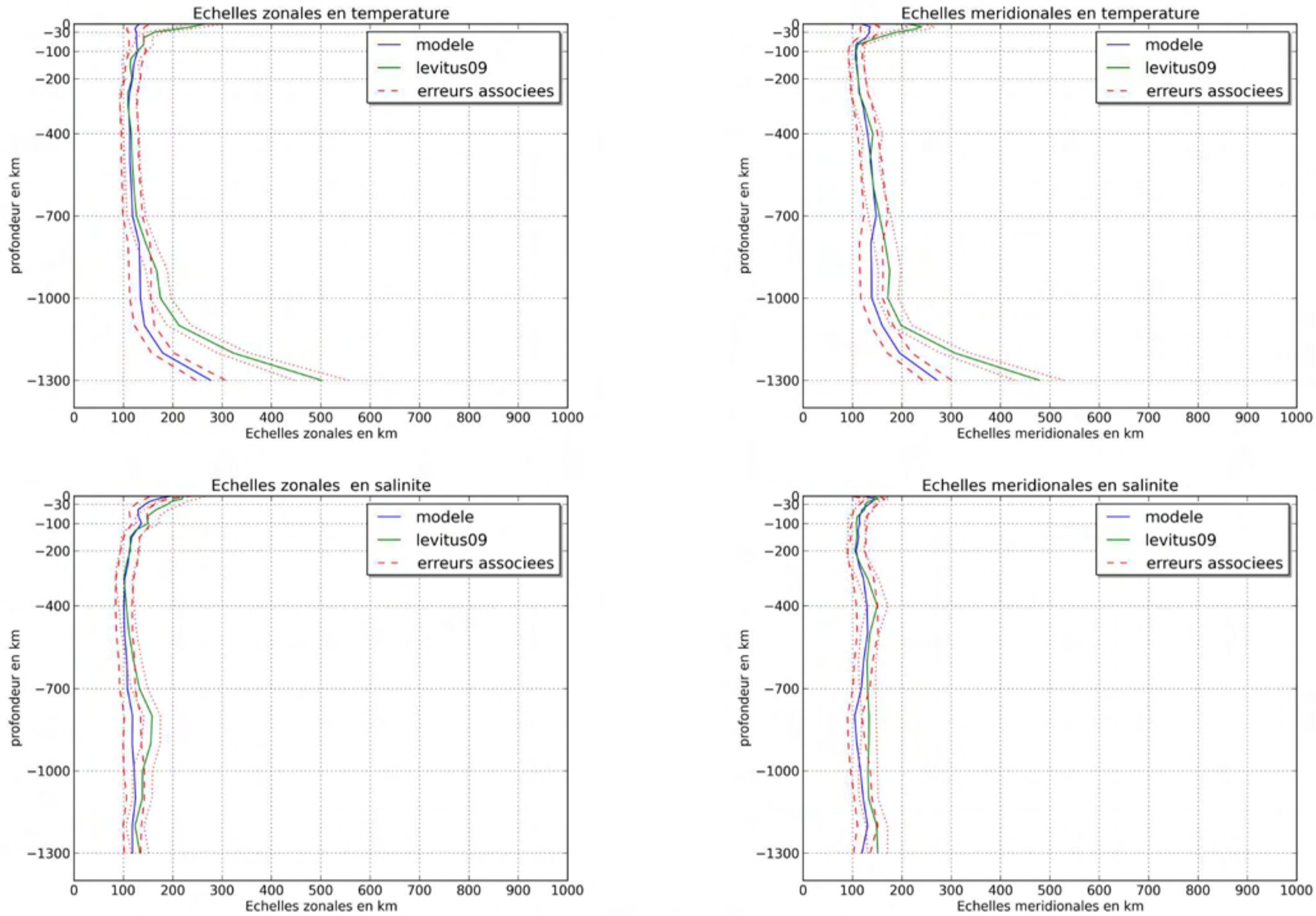


Figure 4.12 – *Corrélation estimée en température (haut) et en salinité (bas) à partir des erreurs modèles (bleu) et pour le signal ([climatologie − observations], en vert) dans l'océan Atlantique Sud (zone 12). Les incertitudes liées à l'ajustement des échelles sont aussi représentées (pointillés rouge)*

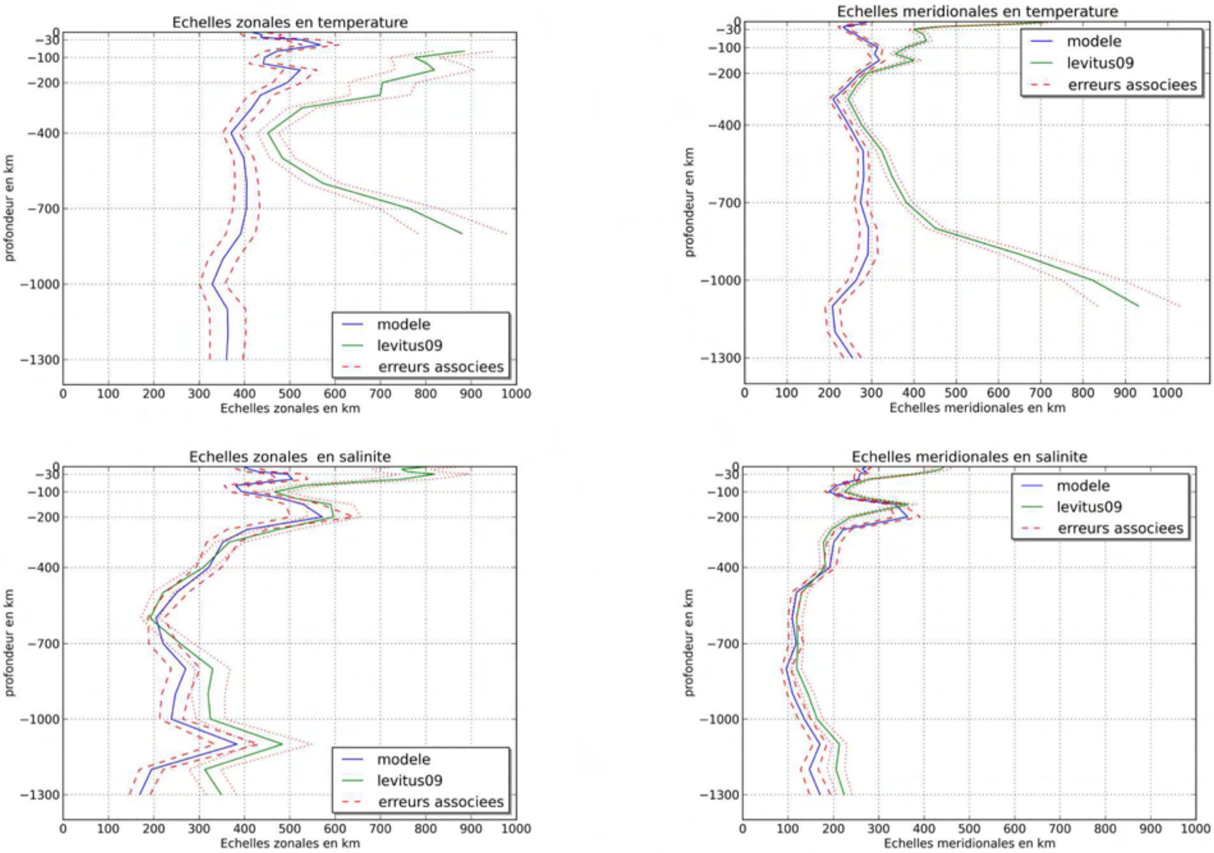


Figure 4.13 – *Corrélation estimée en température (haut) et en salinité (bas) à partir des erreurs modèles (bleu) et pour le signal ([climatologie − observations] en vert) dans la région tropicale (zone 6). Les incertitudes liées à l'ajustement des échelles sont aussi représentées (pointillés rouge)*

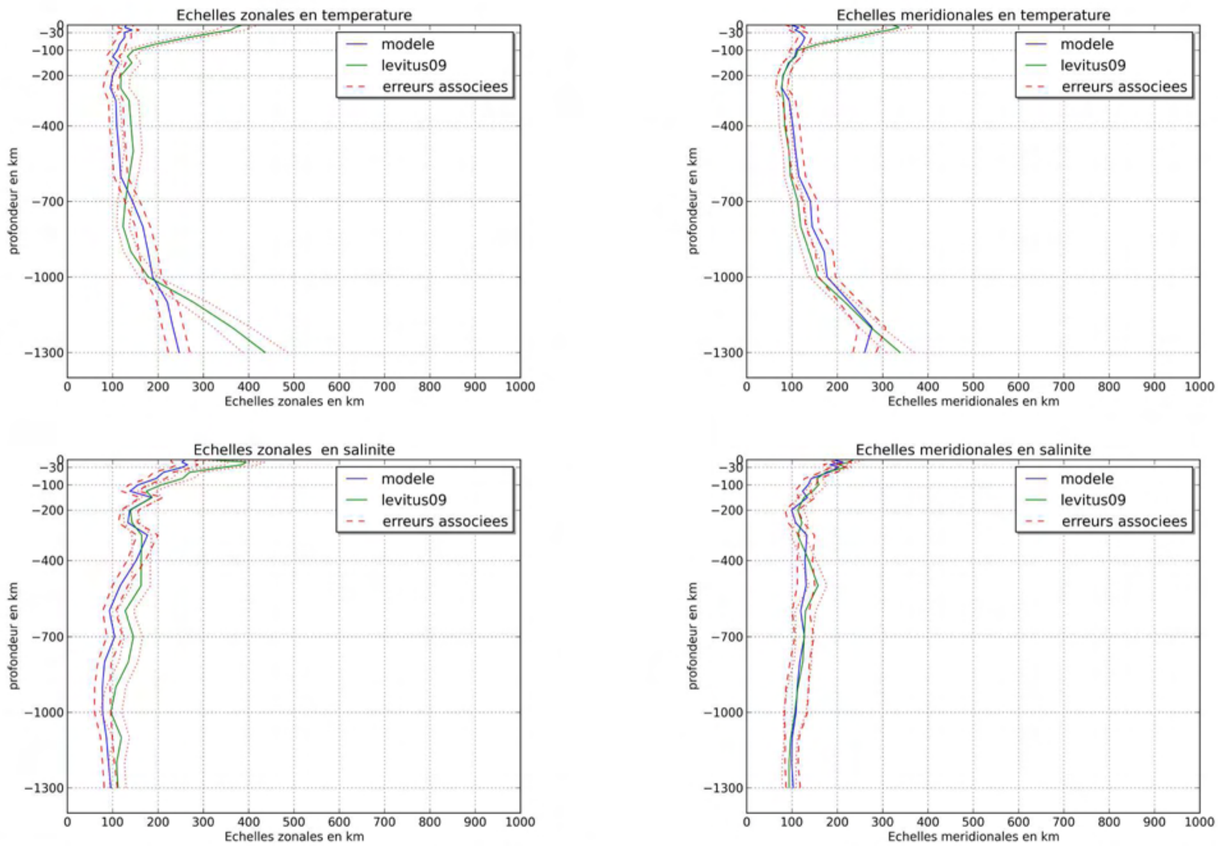


Figure 4.14 – *Corrélation estimée en température (haut) et en salinité (bas) à partir des erreurs modèles (bleu) et pour le signal ([climatologie – observations] en vert) aux hautes latitudes de l'hémisphère Sud (zone 14). Les incertitudes liées à l'ajustement des échelles sont aussi représentées (pointillés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

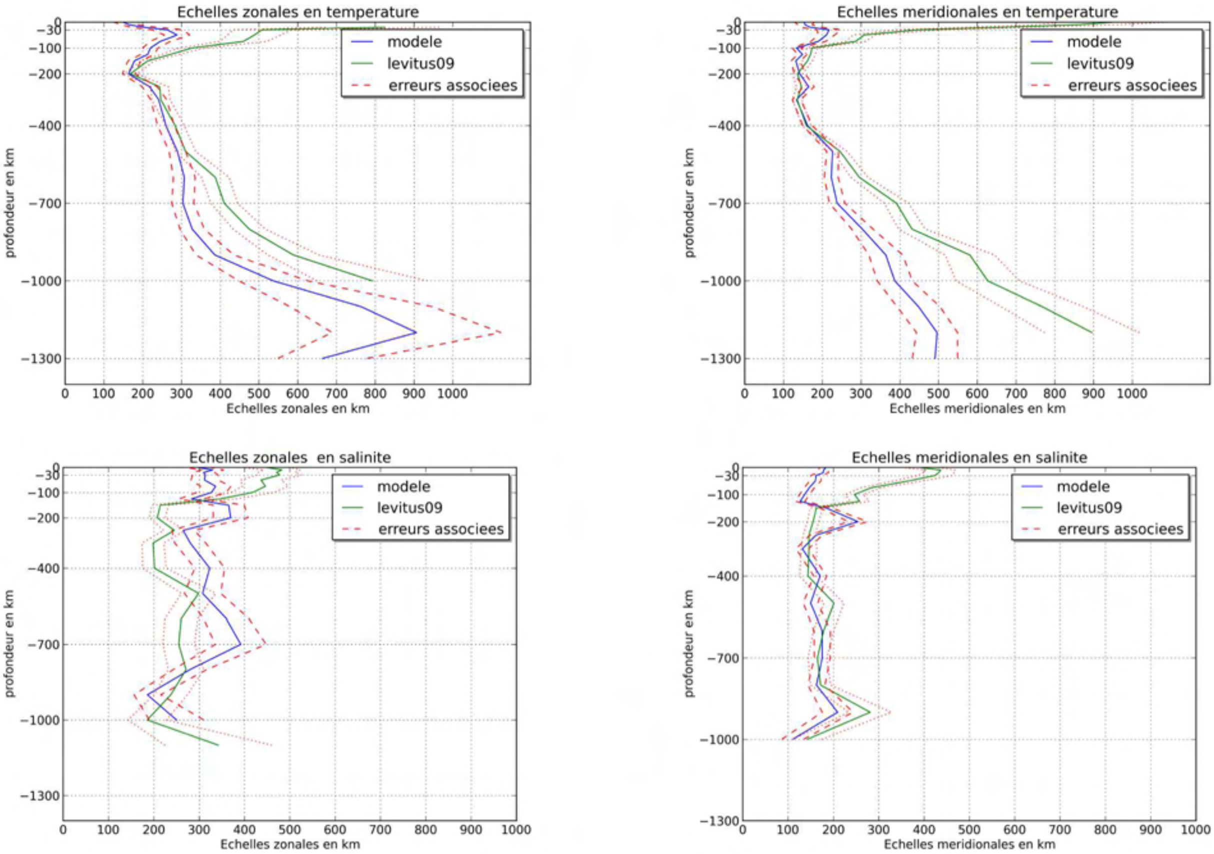


Figure 4.15 – *Corrélation estimée en température (haut) et en salinité (bas) à partir des erreurs modèles (bleu) et pour le signal [climatologie – observations] (vert) dans l'océan Nord-Est Pacifique (zone 4). Les incertitudes liées à l'ajustement des échelles sont aussi représentées (pointillés rouge)*

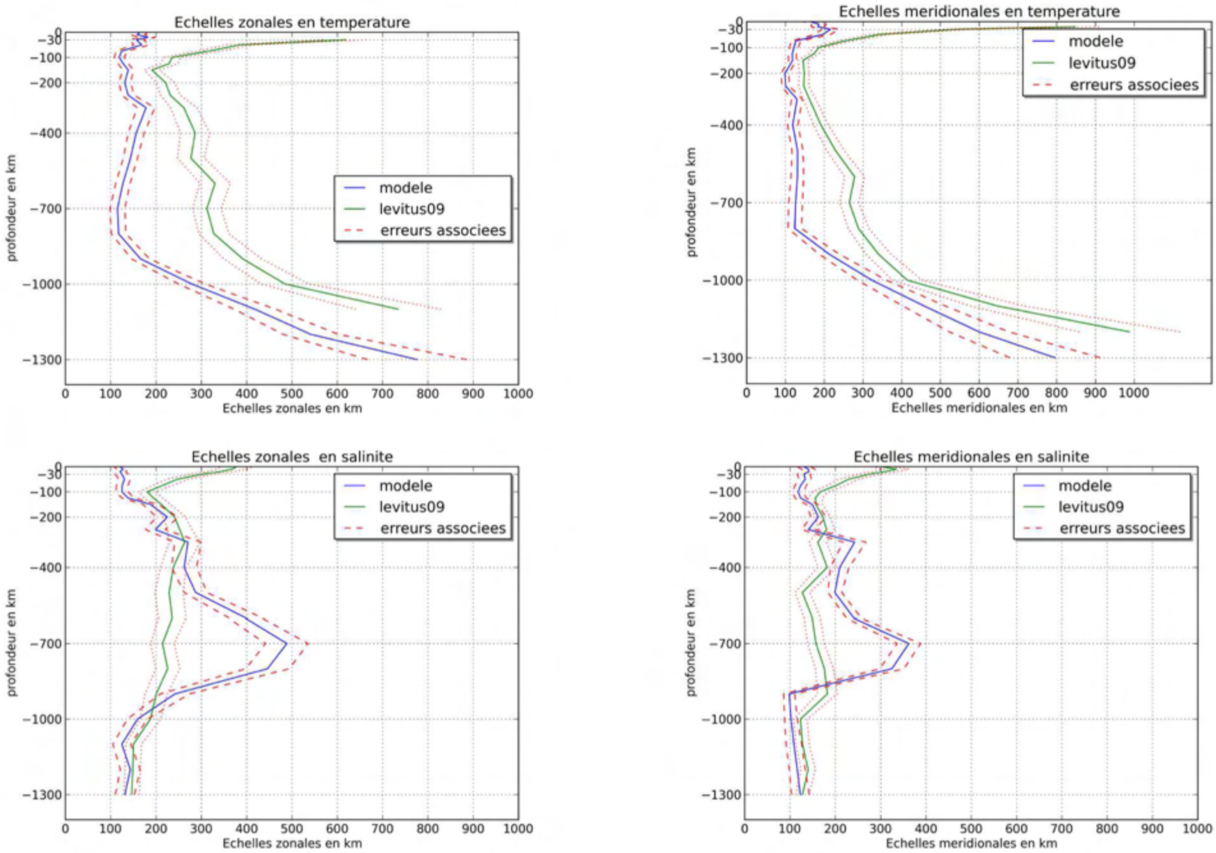


Figure 4.16 – *Corrélation estimée en température (haut) et en salinité (bas) à partir des erreurs modèles (bleu) et pour le signal [climatologie – observations] (vert) dans l'océan Sud-Est Pacifique (zone 11). Les incertitudes liées à l'ajustement des échelles sont aussi représentées (pointillés rouge)*

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:

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

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

   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 $\mathcal{E}$ here, is then derived following the Isserlis theorem (Bendat and Piersol, 1986) and expressed as:

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

   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 $dx_{ij} \gg dx$ or $dy_{ij} \gg dy$), the formal error variance on the correlation is simply equal to the following expression:

   $Var(\mathcal{E}) = 1/N\ [1+cor(dx,dy)]$.

   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).

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

   $\mathrm{Cor}(dx, dy) = 1$ 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

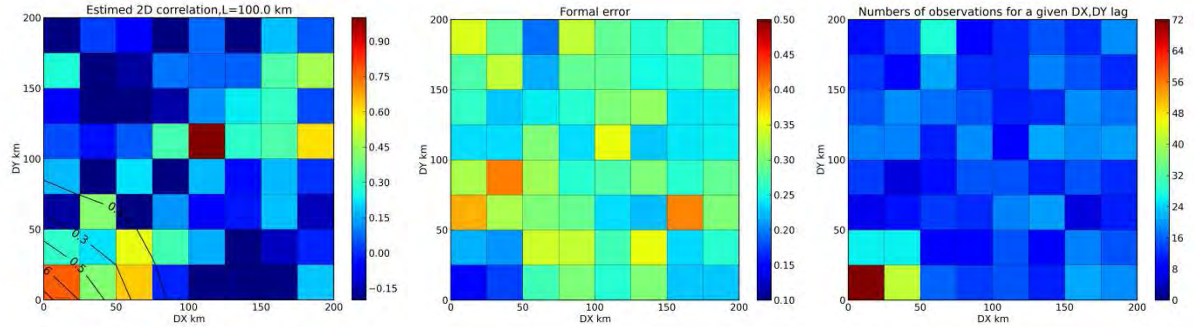


Figure 1a: Estimated 2D covariance fields for the L=100 km simulation for the 2005 Argo sampling (left), associated formal errors (middle) and number of observation pairs (right).

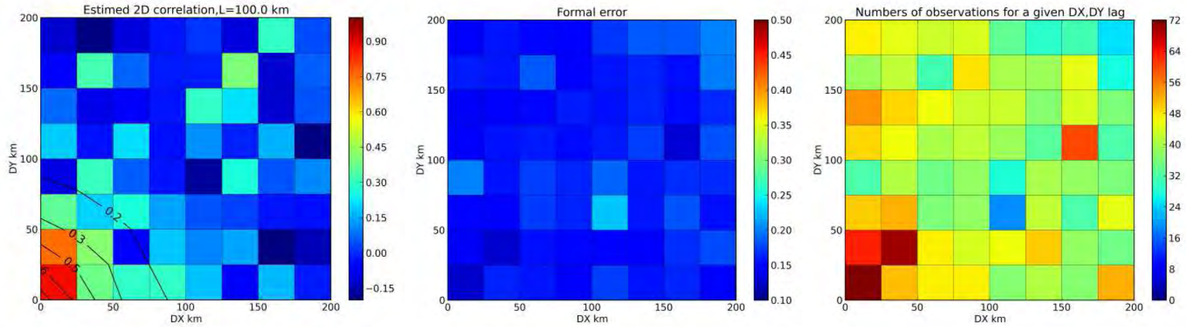


Figure 1b: Estimated 2D covariance fields for the L=100 km simulation for the 2013 Argo sampling (left), associated formal errors (middle) and number of observation pairs (right).

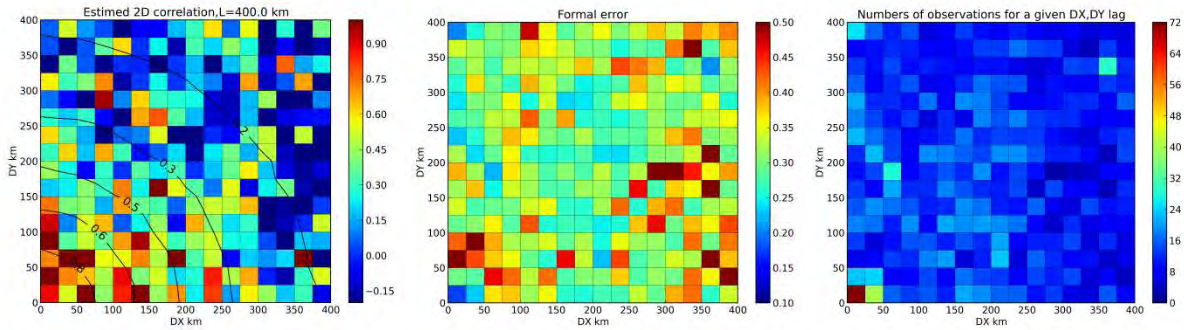


Figure 2a: Estimated 2D covariance fields for the L=400 km simulation for the 2005 Argo sampling (left), associated formal errors (middle) and number of observation pairs (right).

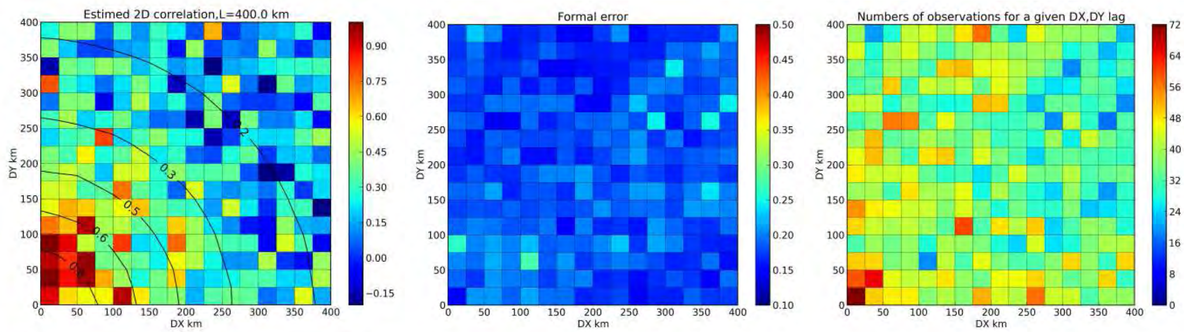


Figure 2b: Estimated 2D covariance fields for the L=400 km simulation for the 2013 Argo sampling (left), associated formal errors (middle) and number of observation pairs (right). Black isolines (left) correspond to the adjusted covariance model.

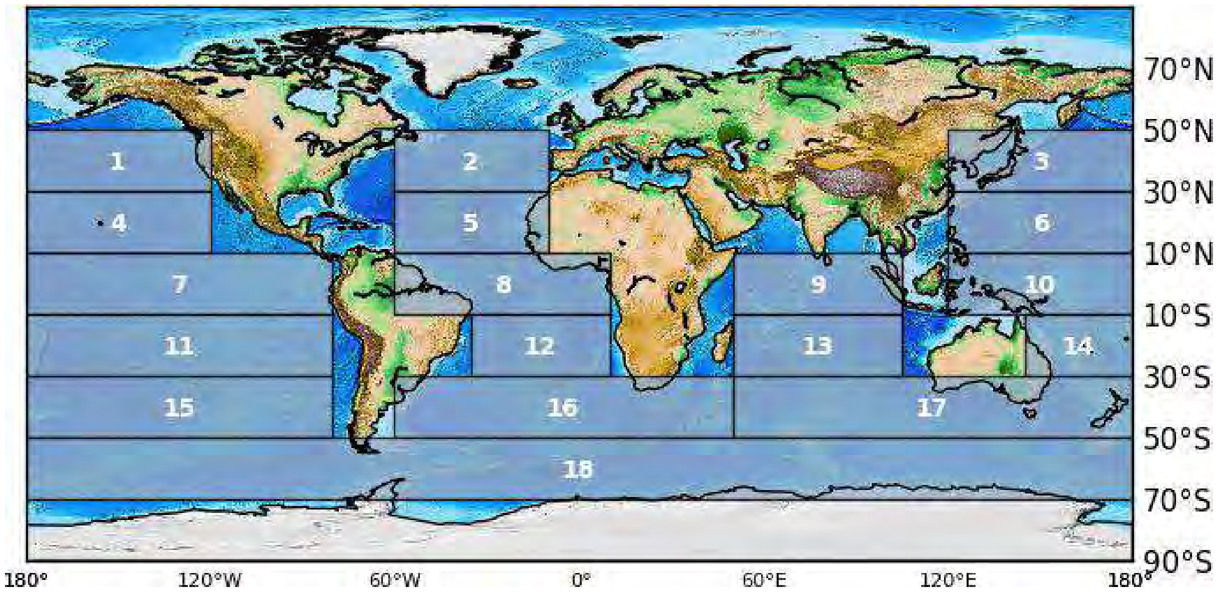


Figure 3: Large scale areas where temperature and salinity spatial correlations were calculated.

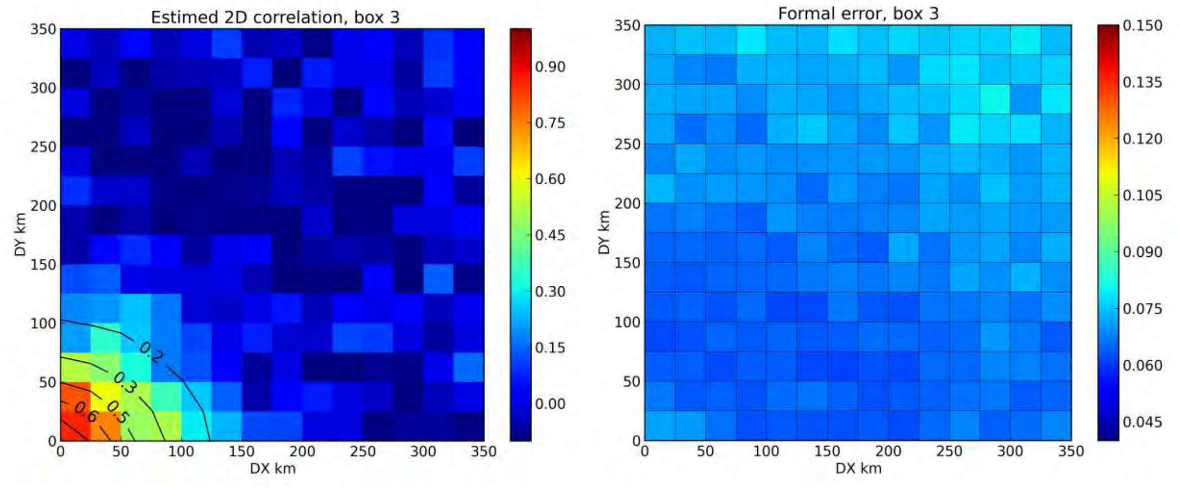


Figure 4: 2D covariance calculated in a 20°x60° area (box number 3) in the North Pacific Ocean (left) and associated formal error (right) for temperature at 200 m. Black isolines (left) correspond to the adjusted covariance model.

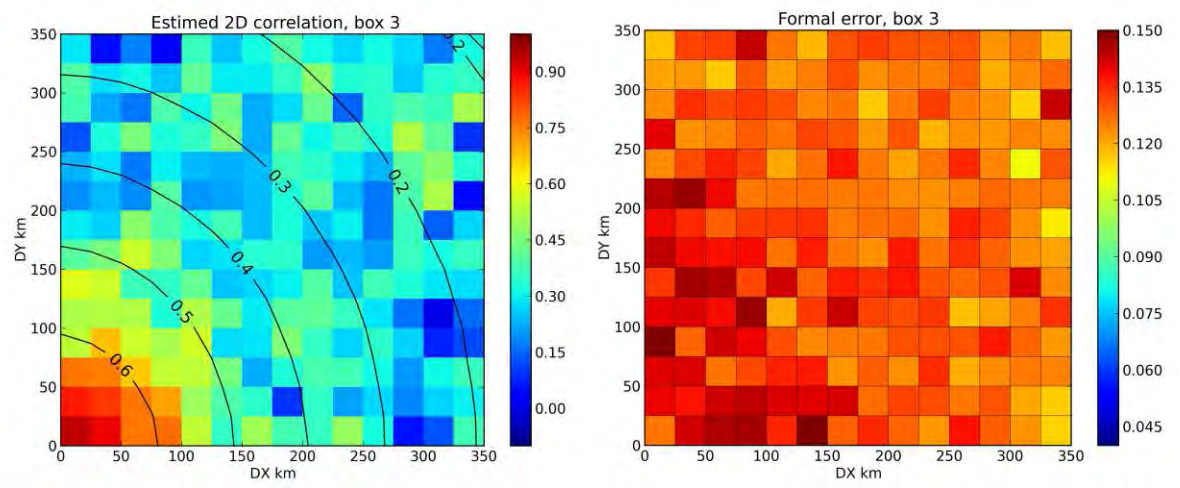


Figure 5: 2D covariance calculated in a 20°x60° area (box number 3) in the North Pacific Ocean (left) and associated formal error (right) for temperature at 1000 m. Black isolines (left) correspond to the adjusted covariance model.

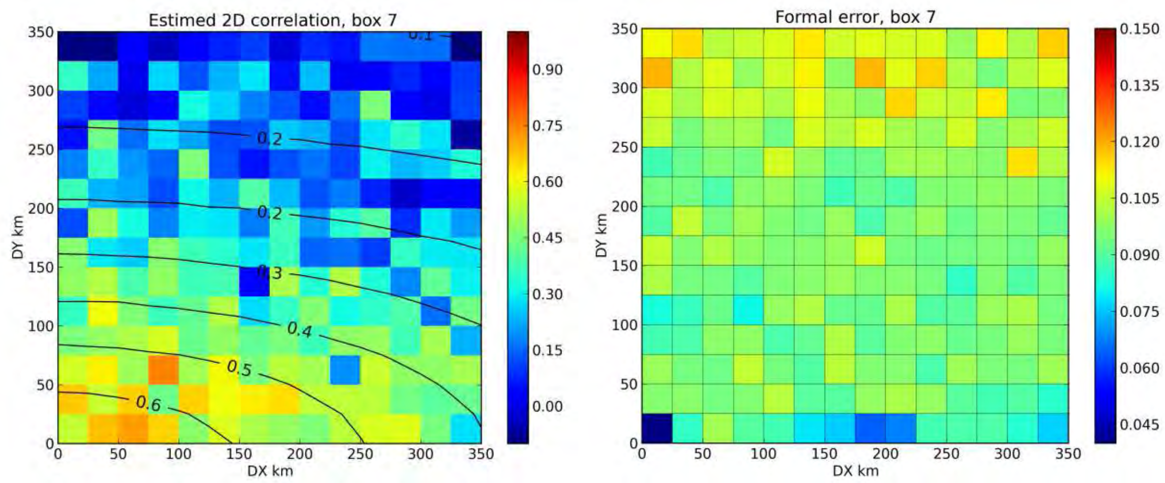


Figure 6: 2D covariance calculated in a 100°x20° area (box number 7) in the Equatorial Pacific Ocean (left) and associated formal error (right) for temperature at 200 m. Black isolines correspond to the adjusted covariance model.

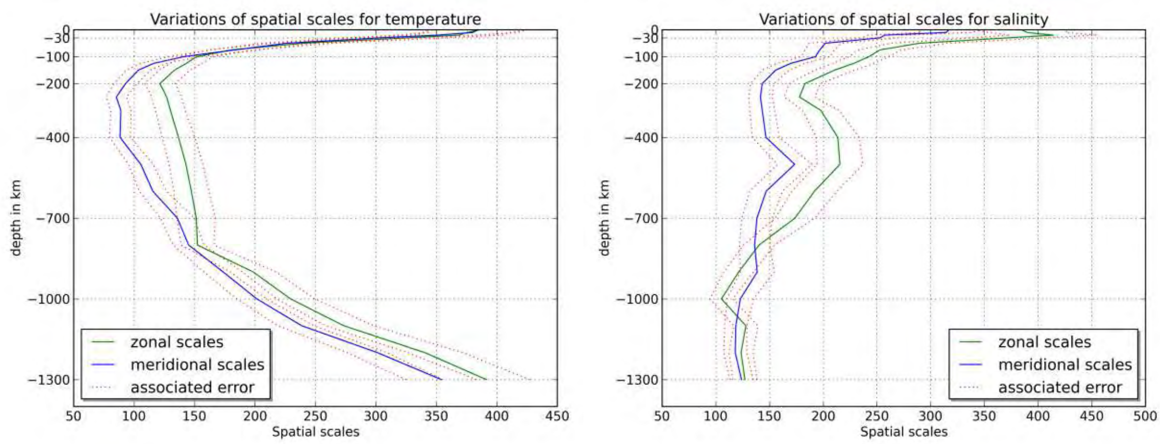


Figure 7: Variations of zonal and meridional spatial scales for temperature (left) and salinity (right) according to depth for box 18 (High latitude Southern Hemisphere). Dotted lines represent standard fitting errors.

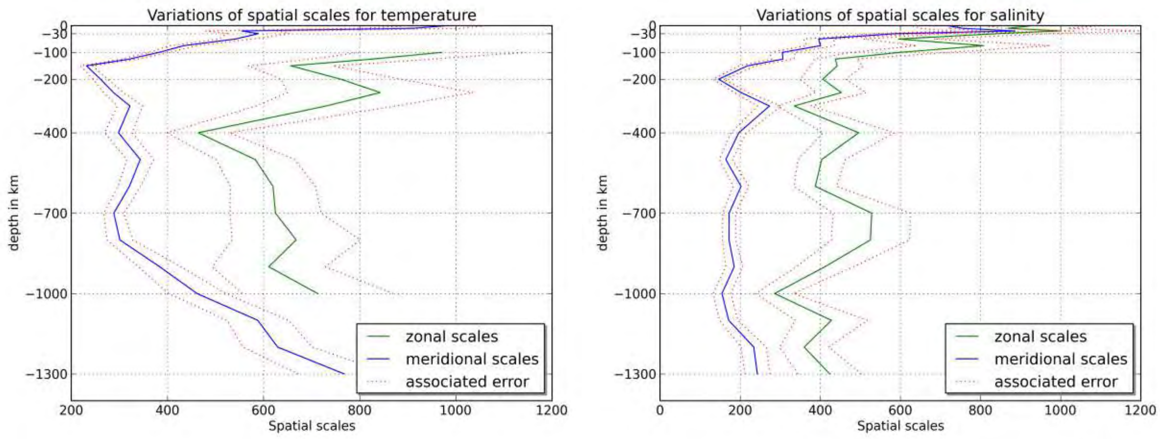


Figure 8: Variations of zonal and meridional spatial scales for temperature (left) and salinity (right) according to depth for box 9 (Equatorial Indian Pacific). Dotted lines represent standard fitting errors.

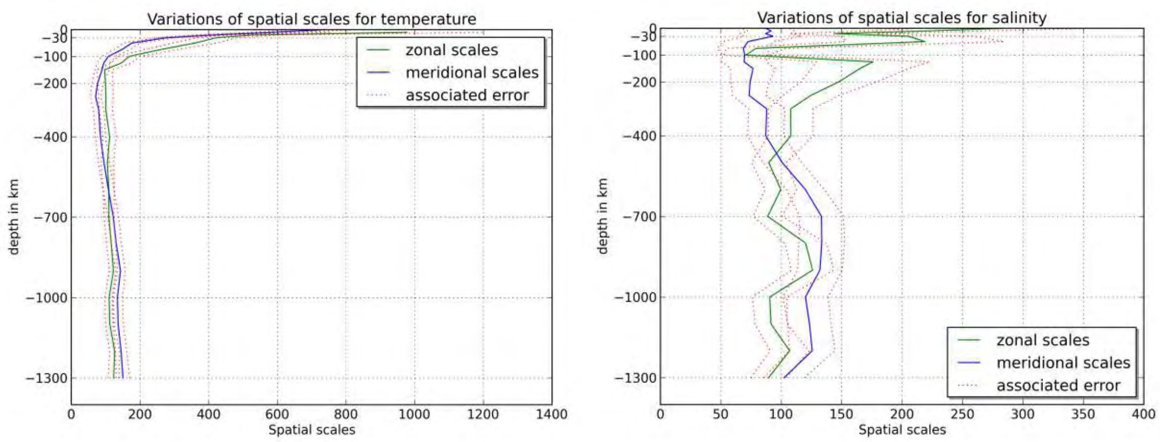


Figure 9: Variations of zonal and meridional spatial scales for temperature (left) and salinity (right) according to depth for box 2 (mid latitude North Atlantic). Dotted lines represent standard fitting errors.

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 $\mathbf{cov}$.
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.