

THÈSE

Pour obtenir le grade de
Docteur

Délivré par
UNIVERSITE DE PERPIGNAN VIA DOMITIA

Préparée au sein de l'école doctorale
Energie Environnement - ED305

Et de l'unité de recherche
IMAGES_Espace-DEV

Spécialité :
INFORMATIQUE

Présentée par
MOUSSA Hadjer

**Traitement automatique de données
océanographiques pour l'interpolation de la
CO₂ de surface dans l'océan atlantique
tropical, en utilisant les données satellitaires**

Soutenue le 19/09/2016

devant le jury composé de :

- | | |
|------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------|
| Mme. El Jai Marie-Claude
Maître de conférences HDR, Université de Perpignan Via Domitia,
France | Membre du
jury |
| Mme. GOYET Catherine
Professeur, Université de Perpignan Via Domitia, France | Directrice |
| Mme. LE BRIS Nadine
Professeur, Laboratoire d'Écogéochimie des Environnements
Benthiques, Banyuls, France | Présidente
du jury |
| Mme. LEFÈVRE Nathalie
Chargé de recherche, Université Pierre et Marie Curie, France | Rapporteur |
| Mme. SOUVERMEZOGLOU Ekaterini
Directeur de Recherche, Institute of Oceanography, Hellenic
Centre for Marine Research, Grèce | Rapporteur |
| M. TOURATIER Franck
Maître de conférences HDR, Université de Perpignan Via Domitia,
France | Membre du
jury |



À mes très chers parents, Kacem et Fatiha, vous qui avez sacrifié toute votre vie pour notre bonheur, vous m'avez toujours soutenu, encouragé, et épaulé pour que je puisse réussir et atteindre mes objectifs,

À ma très chère sœur Kaouther pour ses précieux conseils et son soutien moral. À mon très cher frère Omrane, qui sait toujours comment apporter de la joie au sein de la famille,

À ma chère belle-famille BENALLAL, Nasser, Hadja-Arbia, et Chahinez, vous qui m'avez toujours considéré comme votre propre fille,

Et enfin, à mon très précieux époux, Anis, toi qui a toujours su être un parfait mari, un frère, un ami ... jamais je n'oublierai le grand soutien que tu m'as apporté durant ces trois années de thèse.

À vous tous, je dédie ce travail ...

Remerciements

J'adresse mes premiers remerciements au Ministère Algérien de l'Enseignement Supérieur et de la Recherche Scientifique, pour le financement de ma thèse, ainsi qu'au Ministère des Affaires Étrangères et le Consulat Algérien à Montpellier pour leurs supports administratifs.

Merci aussi aux membres du SURI et spécialement Mme Gisèle CANAL pour tout l'effort qu'elle a fourni pour notre intégration au sein de l'UPVD.

Un grand merci à ma directrice de thèse Mme Catherine GOYET, qui m'a donné l'opportunité de réaliser cette thèse de doctorat, ce qui m'a permis de découvrir le monde de l'océanographie. Je la remercie aussi pour son aide et son orientation au cours de ces trois années de thèse.

Je voudrais aussi remercier Mme Nathalie LEFÈVRE et Mme Katerina SOUVERMEZOGLOU d'avoir accepté d'être les rapporteurs de cette thèse. Je tiens aussi à remercier Mme Marie-Claude EL JAI, Mme Nadine LE BRIS, et M Franck TOURATIER pour avoir accepté de faire partie de mon jury.

Merci à toute l'équipe IMAGES_Espace-DEV, et plus particulièrement à Virginie CASADEI et Cédric FALCO pour leur gentillesse et leur soutien moral, mais aussi à l'équipe du BAE, et plus particulièrement Florance VOUBE et Christophe CANAL.

Un très grand merci à tous mes amis : Rawa HASSAN EL FALLAH, Elissar GEMAYEL, Abed EL RAHMAN HASSOUN, Rafife NHEILI, Chemseddine CHOHRA, Hugues DE LASSUS SAINT-GENIES, Abdelkrim GACEM CHAOUCH, Jannine Marquez LENCINA AVILA. Et enfin un très très grand merci pour ma meilleure amie Nasmine DAMOUCHE, que je considère comme ma propre sœur, merci pour ta serviabilité, ta gentillesse et ton altruisme.

Résumé

Ce travail de thèse s'inscrit dans un cadre multidisciplinaire, ayant pour but d'exploiter les connaissances qu'on a en terme de traitement informatique de données, afin d'apporter une contribution à l'interpolation, avec une bonne résolution spatiale (4 km × 4 km) et sur une grande échelle temporelle, d'un paramètre océanographique nécessaire pour l'étude du cycle du carbone dans l'océan : la fugacité du CO₂ (fCO₂). Notre étude se concentre sur l'océan Atlantique tropical (de 23,5°S-70°O à 23,5°N-15°E) de 2002 à 2013, pour toutes les saisons de l'année.

Des données de 3 différentes provenances ont été utilisées dans cette étude : 1) *in situ* (SST ; *sea surface temperature*, SSS ; *sea surface salinity*, et fCO₂) de la base de données (BD) SOCAT V.3; 2) satellitaires (SST, Chl-*a* ; chlorophylle-*a*, et SSS) des capteurs MODIS-A, Sea-WIFS, et SMOS ; et 3) assimilées (SSS) de la BD SODA V.2.2.4.

Chaque saison a été traitée à part. On a commencé par une étape de classification des données basée sur la SST, en utilisant la méthode "*K-means*". Pour chaque classe, on a réalisé une interpolation de la fCO₂ basée sur les données *in situ*, en utilisant les "réseaux de neurones artificiels" avec une architecture de type *feedforward*, et un apprentissage avec la méthode *backpropagation*. L'algorithme proposé permet de créer un ensemble de 20 réseaux de neurones, avec un nombre de neurones cachés allant de 1 à 20, et de choisir celui garantissant au mieux l'étape de généralisation qui, est basée sur les données satellitaires.

Les résultats de tous les réseaux de neurones créés ont été analysés, ce qui a permis d'en tirer un certain nombre de conclusions. Celles-ci concernent principalement les causes de l'augmentation des résidus, essentiellement dû à la faible densité des faibles et des fortes valeurs de fCO₂, mais aussi aux changements de la variabilité de celle-ci (et/ou des autres paramètres) par rapport à la tendance générale de l'ensemble de toutes les données. Les résultats obtenus (des RMSEs (*root mean square error*) allant de 8,8 à 15,7 μatm suivant les classes) nous ont permis de confirmer l'importance de : traiter chaque saison à part, passer par une étape de classification des données, et enfin de baser le choix du meilleur réseau de neurone sur les résultats de la généralisation.

Le produit final de cette étude est constitué d'un ensemble de 138 fichiers CSV (*Comma-separated values*), chaque fichier correspond à un seul mois, et contient la compilation des données satellitaires (et assimilées) de : SST, Chl-*a*, et SSS, ainsi que la fCO₂ de surface, interpolées avec une résolution spatiale de 4 km × 4 km, et une précision ne dépassant pas 15,7 μatm, représentant la période qui s'étend de juillet 2002 à décembre 2013.

Abstract

This thesis work is part of a multi-disciplinary framework, aimed to exploit our knowledge in terms of data processing, in order to contribute to the interpolation, with a good spatial resolution ($4 \text{ km} \times 4 \text{ km}$) and a large temporal scale, of an ocean parameter required to the study of the ocean's carbon cycle : the CO_2 fugacity (fCO_2). Our study focuses on the tropical Atlantic ocean (from $23,5^\circ\text{S}$ - 70°O to $23,5^\circ\text{N}$ - 15°E) from 2002 to 2013, for all seasons of the year.

Data from 3 different sources were used in this study : 1) *in situ* (SST ; sea surface temperature, SSS ; sea surface salinity, and fCO_2) from the SOCAT V.3 database ; 2) satellite (SST, Chl-*a* ; chlorophyll-*a*, and SSS) from the MODIS-A, Sea-WIFS, and SMOS sensors ; and 3) assimilated (SSS) from the SODA V.2.2.4 database.

We processed every season separately. We began by a data classification step based on SST, using the “*K-means*” method. For every class, we achieved an fCO_2 interpolation based on *in situ* data using “artificial neural networks” with a *feedforward* architecture, and the *backpropagation* training method. The proposed algorithm allows to create a set of 20 neural networks, with a hidden neuron number ranging from 1 to 20, and to choose the one that best ensure the generalization step, which is based on satellite data.

Results of all created neural networks were analyzed, allowing to draw a number of conclusions. These mainly concern the causes of the residuals augmentation, primarily due to the low density of the low and the high fCO_2 values, but also to change in variability thereof (and/or the other parameters) compared to the general tendency of global dataset. The obtained results (RMSEs (*root mean square error*) ranging from 8,8 to 15,7 μatm according to classes) allowed us to confirm the importance of : process each season separately, pass through a data classification step, and finely choose the best neural network on the basis of the generalization results.

The final product of this study consists of a set of 138 CSV (*Comma-separated values*) file, each file represents one month, and contains the compilation of satellite (and assimilated) data of : SST, Chl-*a*, and SSS, as well as the sea surface fCO_2 , interpolated with a $4 \text{ km} \times 4 \text{ km}$ spatial resolution, and a uncertainty not more than 15,7 μatm , for the period from July 2002 to December 2013.

Table des matières

INTRODUCTION GÉNÉRALE	1
1 LE DIOXYDE DE CARBONE ET L'OCÉAN	3
1.1 Du CO ₂ gazeux au CO ₂ dissous	4
1.2 Le carbone total (C _T)	5
1.3 Le pH dans l'océan	5
1.4 L'alcalinité totale (A _T)	6
1.5 La fugacité du CO ₂ (fCO ₂)	6
1.5.1 Mécanisme thermodynamique : <i>pompe de solubilité</i>	7
1.5.2 Mécanismes biologiques : <i>pompe biologique/contre-pompe des carbonates</i>	7
1.5.3 Mécanisme physique : <i>circulation thermohaline</i>	8
1.6 La température des océans	9
1.7 La salinité des océans	10
1.8 Puits et sources de CO ₂	10
1.8.1 Impacts du CO ₂ sur l'écosystème marin	11
1.9 L'état actuel de l'océan quant au CO ₂ dissous	12
2 ZONE D'ÉTUDE : ATLANTIQUE TROPICAL	15
2.1 Zone de Convergence InterTropicale : ITCZ	17
2.2 Les courants de surface	18
2.2.1 Courant équatorial Sud (SEC)	19
2.2.2 Courant Nord Brésil (NBC)	19
2.2.3 Contre courant équatorial Nord (NECC)	20
2.2.4 Courant équatorial Nord (NEC)	20
2.3 Zones d'upwelling	20
2.4 La SST et les modes de variabilité climatique	21
2.4.1 Mode équatorial	21
2.4.2 Mode dipolaire (ou inter-hémisphérique)	22
2.5 La salinité de surface (SSS)	22
2.6 La Chlorophylle- <i>a</i>	23
2.7 La fugacité du CO ₂	24
2.8 La base de données SOCAT	25

3	L'OCÉAN VU DE L'ESPACE	27
3.1	Les débuts de la télédétection satellite	27
3.2	Principe de fonctionnement	28
3.2.1	Le rayonnement électromagnétique	30
3.2.2	Spectre électromagnétique	31
3.2.3	Signature spectrale	32
3.3	Classification des capteurs de télédétection	33
3.4	L'exploitation des données	34
3.5	La télédétection en océanographie	36
3.5.1	Couleur de l'océan	36
3.5.2	Température de surface (SST)	36
3.5.3	Salinité de surface (SSS)	37
4	MÉTHODES	39
4.1	Exemples d'études traitant l'interpolation de la $f\text{CO}_2$	39
4.2	Les réseaux de neurones artificiels	43
4.2.1	Principe de fonctionnement	45
4.2.2	Le perceptron multi-couches : MLP	47
4.3	Le partitionnement de données	49
4.3.1	K -means	49
4.4	langage de programmation : MATLAB [®]	51
5	PRÉSENTATION DES DONNÉES	53
5.1	Données <i>in situ</i>	55
5.1.1	Prétraitement des données <i>in situ</i>	55
5.2	Données satellitaires : Chl- <i>a</i> , SST, et SSS	74
5.2.1	Chl- <i>a</i>	75
5.2.2	SST	81
5.2.3	SSS	82
5.3	Données assimilées : SSS	83
5.4	Regroupement des données	84
5.5	Classification des données	86
5.5.1	Données satellitaire : SST	86
5.5.2	Données <i>in situ</i>	88
6	INTERPOLATION DE LA FUGACITÉ DU CO_2	91
6.1	L'algorithme suivi	91
6.2	Construction des réseaux de neurones	94
6.2.1	Printemps : Classe 1 (C1)	94
6.2.2	Printemps : Classe 2 (C2)	101
6.2.3	Printemps : Classe 3 (C3)	110
6.2.4	Hiver : Classe 1 (C1)	117
6.2.5	Hiver : Classe 2 (C2)	123
6.2.6	Hiver : Classe 3 (C3)	130
7	ÉLABORATION DES CARTES DE $f\text{CO}_2$	139

CONCLUSIONS ET PERSPECTIVES	151
BIBLIOGRAPHIE	157
ANNEXES	172
A CODES MATLAB	173
A.1 Spécifications techniques	173
A.2 Lecture de fichier de type NetCDF	173
A.3 Formatage des données <i>in situ</i>	173
A.3.1 Diverses structurations des données <i>in situ</i>	175
A.4 Prétraitement des données <i>in situ</i>	177
A.4.1 Bathymétrie et SSS	177
A.4.2 $f\text{CO}_2$ et distance à la terre	177
A.5 Formatage des données de SST et Chl- <i>a</i> de MODIS-A	178
A.5.1 Lecture et structuration de fichiers HDF	178
A.6 Regroupement de SST et de Chl- <i>a</i> satellitaires	180
A.7 Formatage des données de SSS de SMOS	181
A.8 Formatage des données de SSS SODA	182
A.9 Regroupement données <i>in situ</i> et SST/Chl- <i>a</i> de MODIS-A	184
A.9.1 Fonction de calcul de la distance euclidienne	185
A.10 Regroupement des données <i>in situ</i> et SSS de SODA	185
A.11 Regroupement des données <i>in situ</i> et SSS de SMOS	186
A.12 Classification des données satellitaires de SST	187
A.13 Classification des données <i>in situ</i>	189
A.13.1 Domaines de validation	190
A.14 Répartitions des BDs pour la construction des RNs	190
A.15 Construction des réseaux de neurones	191
A.15.1 Choix du RN optimal	193
A.16 Élaboration des fichiers CSV	193
A.16.1 Regroupement de SST-Chl- <i>a</i> avec SSS	193
A.16.2 Calcul de la $f\text{CO}_2$	196
A.16.3 Production des fichiers CSV	198
B VARIABILITÉS DES PARAMÈTRES : été, automne	201
B.1 Été	201
B.1.1 Tableaux	201
B.1.2 Figures	204
B.2 Automne	208
B.2.1 Tableaux	208
B.2.2 Figures	211
C INTERPOLATION DE LA $f\text{CO}_2$: été, automne	215
C.1 Été	215
C.1.1 Classe 1 et 2	215
C.1.2 Classe 3	223

C.2	Automne	230
C.2.1	Classe 1 et 2	230
C.2.2	Classe 3	237
D	TABLEAUX	245
D.1	Hiver	245
D.1.1	SST	245
D.1.2	SSS	246
D.1.3	Chl- <i>a</i>	246
D.1.4	fCO ₂	247
D.2	Printemps	247
D.2.1	SST	247
D.2.2	SSS	248
D.2.3	Chl- <i>a</i>	248
D.2.4	fCO ₂	249
D.3	Été	249
D.3.1	SST	249
D.3.2	SSS	250
D.3.3	Chl- <i>a</i>	250
D.3.4	fCO ₂	251
D.4	Automne	251
D.4.1	SST	251
D.4.2	SSS	252
D.4.3	Chl- <i>a</i>	252
D.4.4	fCO ₂	253
E	CONTRIBUTIONS SCIENTIFIQUES	255
E.1	Publication 1 : A comparison of multiple non-linear regression and neural network techniques for sea surface salinity estimation in the tropical Atlantic ocean based on satellite data	255
E.2	Publication 2 : Satellite-derived CO ₂ fugacity in surface seawater of the tropical Atlantic ocean using feedforward neural network	265
E.3	Publication 3 : Ocean salinity from satellite-derived temperature in the Antarctic Ocean	279
E.4	Publication 4 : Satellite-derived CO ₂ flux in the surface seawater of the austral ocean south of Australia	288
E.5	Publication 5 : Evolution of the human impact on oceans (en cours de préparation)	306
F	Copyright, Figure 2.6	309
	TABLE DES FIGURES	324
	LISTE DES TABLEAUX	331

INTRODUCTION GÉNÉRALE

Les progrès scientifiques et techniques qu'a connue l'humanité au cours des derniers siècles, n'ont pas été sans conséquences sur l'environnement. Au delà des avancées technologiques, cette évolution a, et continue à coûter cher à notre planète. Le phénomène qui préoccupe le plus, non seulement la communauté scientifique, mais aussi les politiques et même le grand public, est la forte augmentation des émissions anthropiques du dioxyde de carbone (CO_2), qui est l'une des causes principales du réchauffement climatique.

La redistribution de ces émissions de CO_2 se fait entre l'atmosphère, la biosphère, et l'océan. Néanmoins, bien qu'ils couvrent une proportion de l'ordre de 70% de la surface de la Terre, le rôle des océans dans la régulation du taux de CO_2 rejeté dans l'atmosphère, reste encore méconnu par le grand public. En effet, les océans absorbent l'équivalent de 25 à 30% par an du CO_2 anthropique, ce qui a beaucoup contribué à la diminution du niveau de CO_2 dans l'atmosphère. Ce processus d'absorption est surveillé de près par les scientifiques, car la conséquence directe de ce mécanisme est l'apparition du phénomène appelé "Acidification des Océans" (AO), qui est déclenché par la diminution du pH océanique, et qui peut avoir des conséquences dramatiques sur l'écosystème marin. Ainsi, mieux comprendre le cycle global du carbone dans l'océan, est, de nos jours, quelque chose d'une importance primordiale.

Bien que la quantité des données océanographiques a beaucoup augmentée durant les deux dernières décennies, tout les chercheurs s'accordent sur le fait qu'il est nécessaire d'avoir encore plus de mesures, dans le but d'obtenir des résultats plus précis. En effet, des données avec une fine résolution spatiale, réparties sur une large couverture temporelle, permettent d'avoir une vue globale et plus précise sur la dynamique des océans, et aussi de réduire les contradictions qu'on peut avoir dans les résultats trouvés par les chercheurs. Ce critère est aujourd'hui disponible grâce aux satellites, qui représentent un atout majeur et motive les études océanographiques.

Cette thèse a donc pour but de joindre les bénéfices de la précision des données *in situ*, et les avantages qu'offrent les mesures satellitaires, en terme de quantité d'observations de la surface océanique, afin de cartographier un des plus importants paramètres nécessaires pour l'étude du cycle du carbone dans l'océan, à savoir la "fugacité du CO_2 " ($f\text{CO}_2$). Cette étude doit donc être abordée sous une perspective multidisciplinaire.

Notre étude se focalise sur l'océan Atlantique tropical, qui est connu par ces processus qui sont fortement dynamiques. Bien qu'il est considéré comme une source de CO_2 pour

l'atmosphère, il est encore très difficile d'établir des conclusions précises sur le rôle de ce bassin océanique quant à la régulation du taux de CO₂ atmosphérique. Il serait donc très utile d'apporter une contribution à la création des cartes de fCO₂ d'une bonne résolution spatiale, et réparties sur plusieurs années. On a donc proposé d'exploiter les performances d'une méthode très puissante, en terme d'approximation de fonctions non linéaires, mais aussi en généralisation, dans le but de réaliser une interpolation de la fCO₂ de surface, et ainsi de la cartographier, en utilisant les données satellitaires. La méthode adoptée est issue du domaine de l'"intelligence artificielle" (IA), et inspirée du fonctionnement d'un système biologique très complexe, à savoir le cerveau humain. Cette méthode est dite "réseaux de neurones artificiels".

On présentera dans ce mémoire un ensemble de sept chapitres. Les quatre premiers constituent la partie théorique de l'étude, et les trois autres la partie pratique. Dans le **Chapitre 1**, on parlera de la relation qui existe entre le dioxyde de carbone et l'océan, en expliquant les principaux mécanismes qui concernent l'absorption de ce gaz par l'océan. Le **Chapitre 2** est consacré à la zone d'étude; l'océan Atlantique tropical, on abordera les principales caractéristiques de ce bassin, notamment les modes de variabilité des principaux paramètres océanologiques. Le **Chapitre 3** est dédié au domaine de la télédétection par satellite, on évoquera le principe général du fonctionnement des satellites, et comment les avantages de celle-ci sont explorés pour l'océanographie. Le **Chapitre 4** s'articule en deux parties, dans la première on abordera quelques unes des plus importantes études traitant l'interpolation de la fCO₂ de surface, la deuxième partie est consacrée à la méthode adoptée, à savoir les réseaux de neurones artificiels.

La partie pratique de ce manuscrit est organisée sous forme d'étapes, représentant toute la démarche suivie dans cette étude. On présentera donc un ensemble d'outils et de techniques informatiques qui permettront de gérer et de traiter, de façon automatique, des données océanographiques. On présentera dans le **Chapitre 5** les différentes données utilisées au cours de cette étude, en montrant, entre autre, toutes les techniques de structuration et de préparation des données pour l'étape d'interpolation de la fCO₂. Cette dernière est présentée dans le **Chapitre 6**, où on parlera de l'algorithme adopté pour cette interpolation, des résultats obtenus, ainsi que leur interprétations. Enfin, le **Chapitre 7** contiendra les étapes finales suivies pour l'élaboration des cartes de fCO₂, à partir des données satellitaires.

On terminera par un chapitre (**CONCLUSIONS ET PERSPECTIVES**) dédié à toutes les conclusions qu'on a pu tirées de cette étude, ainsi qu'aux perspectives que ce travail à permit d'engendrer.

Tous les programmes MATLAB[®] utilisés, de la première jusqu'à la dernière étape de la démarche proposée, sont présentés dans l'**Annexe A**.

Enfin, Toute la contribution scientifique qu'on a apportée sous forme d'articles, est présentée dans l'**Annexe E**.

Chapitre 1

LE DIOXYDE DE CARBONE ET L'OCÉAN

Le carbone est un élément chimique qui est naturellement présent dans notre planète. Il suit un certain itinéraire appelé "*cycle du carbone*", ce cycle est illustré dans la **Figure 1.1**. Dans son état naturel, ce cycle présente un certain équilibre qui permet de stabiliser le taux de carbone présent dans les quatre réservoirs naturels : lithosphère, biosphère, hydrosphère, et atmosphère. Cependant, les activités humaines excessives (déforestation, combustion des énergies fossiles, industrialisation ...) perturbent cet équilibre, et modifie la concentration du CO₂ rejeté dans l'atmosphère.

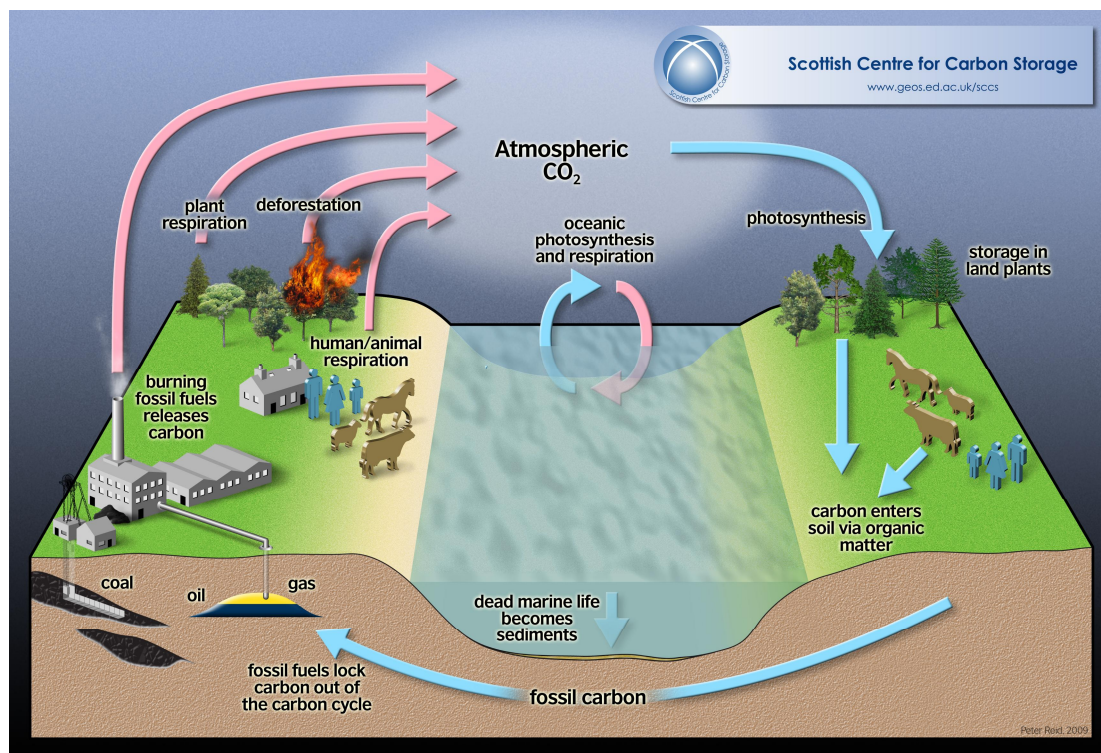


FIGURE 1.1 – Illustration du cycle du carbone en présentant les principaux processus affectant ce cycle. [Harry, 2009]

Le total des émissions cumulées entre 1750 et 2013 a été estimé à 580 ± 70 PgC (Péta gramme de carbone = 10^{15} gramme), partagé entre : l'atmosphère avec un taux de 250 ± 5 PgC, le lithosphère avec 160 ± 70 PgC, et l'océan avec 170 ± 20 PgC [Le Quéré *et al.*, 2015].

D'après les observations atmosphériques et océaniques, ainsi que les résultats des modèles numériques, l'océan absorbe environ 2 PgC.an^{-1} [Battle *et al.*, 2000; Orr *et al.*, 2001], soit 25 à 30% du CO_2 anthropique [Sarmiento *et al.*, 2010; IPCC, 2014]. Ce processus naturel d'absorption a considérablement diminué le niveau de gaz à effet de serre dans l'atmosphère et a minimisé les impacts du réchauffement climatique, mais en augmentant, d'un taux qui de plus croît d'une année à une autre, la capacité des océans à absorber ce CO_2 anthropique est en train de changer. Plusieurs modèles de prédictions ont montré que l'efficacité de ces puits de carbones peut diminuer dans les années à venir [Sarmiento et Le Quéré, 1996; Matear et Hirst, 1999; Joos *et al.*, 1999; Le Quéré *et al.*, 2015]. Ceci aura des implications sur l'équilibre générale du cycle de carbone, et donc sur les changements climatiques.

Le processus d'absorption de CO_2 , les mécanismes jouant un rôle dans la modification du taux de carbone dans l'océan, ainsi que les impacts qui en résultent, vont être expliqués dans ce chapitre.

1.1 Du CO_2 gazeux au CO_2 dissous

On reconnaît le CO_2 comme étant un gaz qui est stable, mais une fois dans l'eau, et après dissolution, plusieurs processus chimiques, biologiques et physiques, le rendent instable. Il peut se trouver sous quatre formes :

$$\left\{ \begin{array}{l} \text{CO}_{2(\text{aq})} \\ \text{HCO}_3^- \\ \text{CO}_3^{2-} \\ \text{H}_2\text{CO}_3 \end{array} \right.$$

Les réactions entre ces espèces chimiques sont régies par des équilibres chimiques représentés sous forme d'un système d'équations dit *système des carbonates* :



Du fait que H_2CO_3 est très instable et n'est donc pas mesurable, on se réfère à la somme :

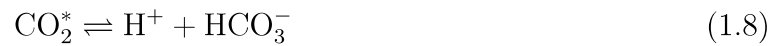
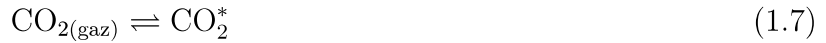
$$[\text{CO}_2^*] = [\text{H}_2\text{CO}_3] + [\text{CO}_{2(\text{aq})}] \quad (1.5)$$

1.2 Le carbone total (C_T)

Le carbone inorganique total dissous (C_T) représente 95% du carbone se trouvant dans l'océan. Le $CO_{2(aq)}$ représente 1% de la somme totale, les bicarbonates (HCO_3^-) constituent environ 90% de C_T , les carbonates (CO_3^{2-}) 9%, et l'acide carbonique (H_2CO_3) représente environ 0,001% de la concentration du CO_2 dissous. En prenant en considération l'équation 1.5, le C_T se définit par la somme :

$$C_T = [CO_2^*] + [HCO_3^-] + [CO_3^{2-}] \quad (1.6)$$

L'unité utilisée pour cette grandeur est : $\mu\text{mol.kg}^{-1}$. Les réactions (1.1), (1.2), et (1.3) sont redéfinies en fonction de CO_2^* :



Ces équilibres obéissent à la loi de la thermodynamique, et sont définis par les constantes de dissociation K_1 et K_2 qui dépendent de la température, la salinité, ainsi que la pression :

$$K_1 = [H^+][HCO_3^-]/[CO_2] \quad (1.9)$$

$$K_2 = [H^+][CO_3^{2-}]/[HCO_3^-] \quad (1.10)$$

Il existent plusieurs expressions empiriques de ces constantes apparentes en fonction de la température et de la salinité [Hansson, 1973a; Mehrbach *et al.*, 1973; Dickson et Millero, 1987; Goyet et Poisson, 1989; Roy *et al.*, 1993].

1.3 Le pH dans l'océan

Le pH représente la concentration en ions hydrogène ($[H^+]$) dans l'eau de mer, il existe trois échelles pour représenté le pH :

1. Échelle des protons libres [The Free scale; Khoo *et al.*, 1977; Ramette *et al.*, 1977] :

$$pH_F = -\log[H^+]_F \quad (1.11)$$

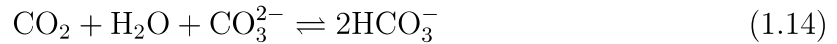
2. Échelle de l'eau de mer [The Seawater scale; Dickson, 1984; Dickson et Millero, 1987] :

$$pH_{sws} = -\log[H^+]_{sws} = -\log\{[H^+]_F + [HSO_4^-] + [HF]\} \quad (1.12)$$

3. Échelle totale [The Total scale; Hansson, 1973b; Dickson, 1993] :

$$pH_T = -\log[H^+]_T = -\log\{[H^+]_F + [HSO_4^-]\} \quad (1.13)$$

L'apport de CO_2 par transfert de CO_2 atmosphérique, par respiration, ou par oxydation de la matière organique, va entraîner une transformation des carbonates en bicarbonates, ce qui provoque une diminution du pH. Inversement, la diminution de CO_2 provoque une augmentation de pH. L'eau océanique est de nature basique. Dans l'océan ouvert, le pH des eaux de surface varie entre 7,9 et 8,3, l'eau devient moins basique quand la valeur de son pH baisse. Toutefois, l'océan est connu par son *effet tampon*, qui permet, dans le cas d'ajout de CO_2 atmosphérique, d'associer les ions CO_3^{2-} au CO_2 dissous et neutralise son effet sur le pH. Ceci s'effectue suivant l'équation :



1.4 L'alcalinité totale (A_T)

L'alcalinité totale d'une eau, peut être définie comme étant sa capacité de résister aux attaques acides (une baisse du pH), elle est donc liée à son contenu en substances à caractère basique (« alcalin ») qui sont capables de se lier à un acide pour le neutraliser. Dans l'océan ouvert, l' A_T s'avère très corrélée avec la salinité et avec les facteurs physiques qui influent sur la variation de la salinité, principalement : l'évaporation et les précipitations, les mouvements des masses d'eau, et les apports fluviaux. Elle est définie ainsi :

$$\begin{aligned} A_T = & \left[\text{HCO}_3^- \right] + 2 \left[\text{CO}_3^{2-} \right] + \left[\text{B}(\text{OH})_4^- \right] + \left[\text{OH}^- \right] + \left[\text{HPO}_4^{2-} \right] \\ & + 2 \left[\text{PO}_4^{3-} \right] + \left[\text{SiO}(\text{OH})_3^- \right] + \left[\text{HS}^- \right] + \left[\text{NH}_3 \right] + \dots \\ & - \left[\text{H}^+ \right]_F - \left[\text{HSO}_4^- \right] - \left[\text{HF} \right] - \left[\text{H}_3\text{PO}_4 \right] - \dots \end{aligned} \quad (1.15)$$

1.5 La fugacité du CO_2 ($f\text{CO}_2$)

La $f\text{CO}_2$ est un paramètre très important pour le système océanique du carbone inorganique : 1) parce qu'elle permet de déterminer l'ampleur ainsi que la direction des échanges de CO_2 entre l'atmosphère et l'océan, et 2) parce qu'elle est considérée comme étant un bon indicateur de changements du cycle de carbone dans la couche de surface des océans, et ceci grâce à une loi dite *Loi de Henry*, qui relie la quantité du CO_2 dissous à la fugacité de CO_2 :

$$[\text{CO}_2^*] = \alpha \dots f\text{CO}_2 \quad (1.16)$$

α représente le *coefficient de solubilité* de CO_2 , défini en $\mu\text{mol.kg}^{-1}.\mu\text{atm}^{-1}$. Ce coefficient est calculé en fonction de la température et de la salinité, suivant l'algorithme proposée par Weiss [1974]. À partir de cette équation, la concentration du CO_2 dissous peut être calculée.

Les mécanismes naturels affectant la distribution de $f\text{CO}_2$ dans l'océan superficiel se résument en trois grands points : les changements de température, l'activité biologique, et la circulation océanique.

1.5.1 Mécanisme thermodynamique : *pompe de solubilité*

La température des eaux de surface est le facteur thermodynamique qui affecte le plus l'équilibre des carbonates, et donc la fCO₂. Un refroidissement de la zone superficielle entraîne l'augmentation de la solubilité du CO₂, et donc une grande absorption du CO₂ atmosphérique, d'où l'appellation "*pompe de solubilité*". La fCO₂ varie d'environ 4% par °C [Gordon et Jones, 1973; Copin-Montegut, 1988; Copin-montegut, 1989]. Différentes formules reliant les variations de fCO₂ à celles de la température ont été proposées [Gordon et Jones, 1973; Copin-Montegut, 1988; Copin-montegut, 1989; Goyet et Poisson, 1989; Goyet *et al.*, 1993]. Par la suite plusieurs méthodes ont été adoptées afin de cartographier la fCO₂ à partir des mesures de la température [Stephens *et al.*, 1995; Lefèvre et Taylor, 2002; Inoue *et al.*, 2003; Olsen *et al.*, 2003; Lefèvre *et al.*, 2005]

1.5.2 Mécanismes biologiques : *pompe biologique/contre-pompe des carbonates*

L'activité biologique a aussi son rôle dans la modification de la concentration des différents composants du système des carbonates. Ces mécanismes sont schématisés dans la **Figure 1.2**.

1.5.2.1 La photosynthèse

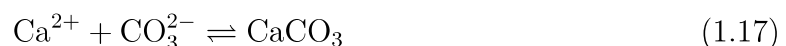
Ce processus s'effectue par le phytoplancton dans la couche appelée "*euphotique*", où la lumière est disponible et suffisante pour cette opération. La croissance de ces organismes marins nécessite la synthèse du carbone organique, afin de fabriquer leurs matières organiques à l'aide du CO₂, de lumière et de sels nutritifs (comme les nitrates dissous). Cette action se traduit donc par une décroissance de l'abondance de C_T. Ceci va provoquer une diminution de la concentration des ions HCO₃⁻, accompagnée d'une augmentation des ions CO₃²⁻. Ceci est appelé "*pompe biologique*". La photosynthèse a donc pour résultat, la diminution de la fCO₂ à la surface de l'océan.

1.5.2.2 La respiration

L'oxydation et la biodégradation de la matière organique formée en surface par la photosynthèse, est assurée par la respiration (en surface, ou au cours de la sédimentation dans les eaux profondes). Ce mécanisme compense la photosynthèse en permettant l'augmentation de la quantité du CO₂.

1.5.2.3 Formation et dissolution des carbonates

Plusieurs organismes marins (foraminifère, coccolithophoridés, coraux ...) construisent leurs coquilles ou socles calcaires en utilisant le calcium des eaux de surface ainsi que les ions CO₃²⁻ pour former le carbonate de calcium CaCO₃ selon l'équation :



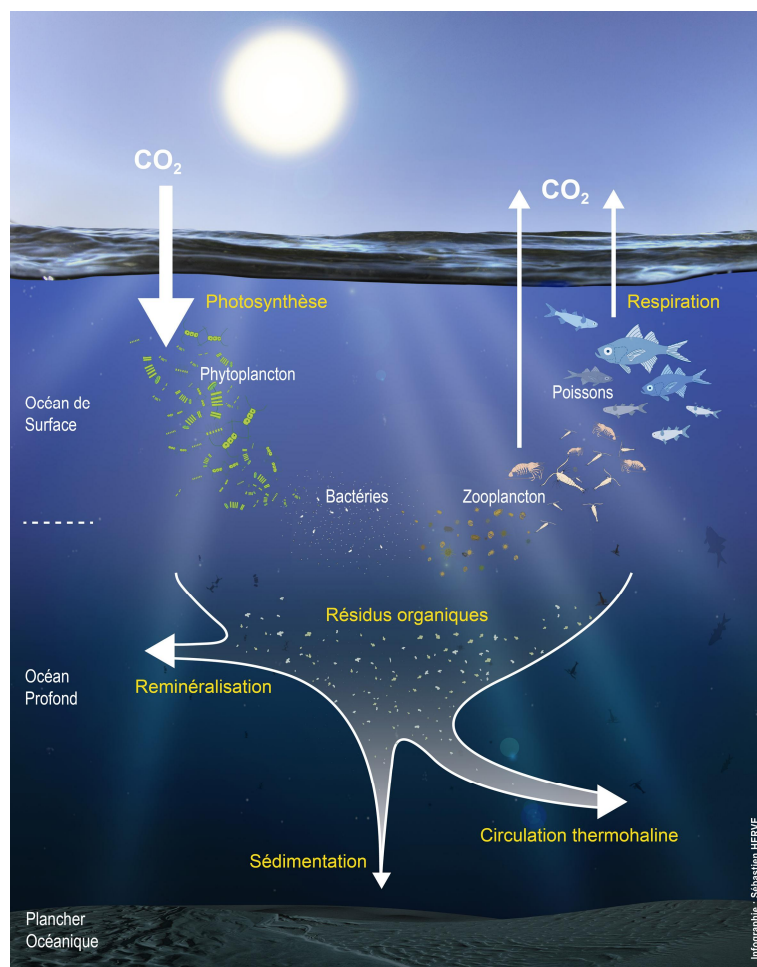


FIGURE 1.2 – Représentation des trois processus biologiques affectant la $f\text{CO}_2$: la photosynthèse, la respiration, et la formation et la dissolution des carbonates. [Gorgues et Planquette, 2016]

Ceci va diminuer les carbonates dans la couche de surface, ce qui va donc entraîner une diminution du C_T et de l' A_T . On a ainsi une augmentation de la $f\text{CO}_2$ en surface. Ce mécanisme est appelé "*contre-pompe des carbonates*". Cependant, la sédimentation de la matière organique et la dissolution de CaCO_3 dans les eaux profondes conduit à une augmentation de l'alcalinité.

À l'échelle globale de l'océan, ces organismes marins utilisent quatre fois plus de carbone pour produire de la matière organique que pour construire leurs squelettes, ainsi, on peut dire que les mécanismes biologiques cités ci-dessus, ont plutôt tendance à diminuer la $f\text{CO}_2$ qu'à l'augmenter.

1.5.3 Mécanisme physique : *circulation thermohaline*

Ce processus est assuré par la circulation à grande échelle dans l'océan mondial dite : "*circulation thermohaline*" (**Figure 1.3**) qui est liée à la température et à la salinité des masses d'eau. La température de l'eau diminue aux pôles et sa salinité augmente, ce qui va augmenter sa densité, cette masse d'eau va donc plonger aux profondeurs en

provoquant la séquestration du CO_2 . Elle va ensuite continuer son chemin vers les zones tropicales, où il y aura une augmentation de sa température et donc une diminution de sa densité, d'où, la remontée de cette eau vers la surface avant de repartir vers les pôles. Ce cycle d'eau permet d'aspirer une partie du CO_2 atmosphérique.

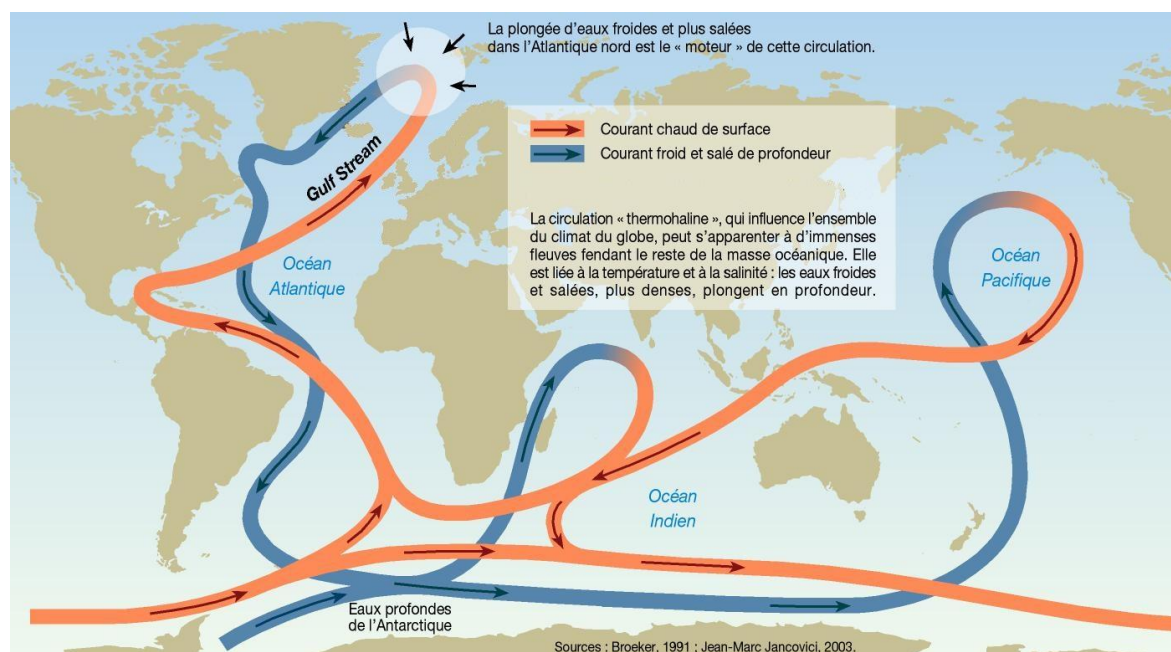


FIGURE 1.3 – Circulation thermohaline présentée par les courants marins chauds (en surface) en orange, et en bleu les courants froids (en profondeur) [Source : goudouly. over-blog.com]

1.6 La température des océans

La température est une grandeur physique familière, elle est considérée comme l'un des paramètres les plus importants de l'océan. Les premiers mètres de la surface de la mer se réchauffent principalement sous l'effet des rayons du soleil. D'autres phénomènes contrôlent aussi les variations de la température, notamment : les échanges de chaleur avec l'atmosphère sous forme de rayonnement infrarouge, les échanges d'eau avec le reste du système océanique, les effets du vent en surface...

Les températures de surface se répartissent suivant les latitudes en : eaux chaudes (jusqu'à 30°C) en zones tropicales, et eaux froides (jusqu'à -2°C) dans les hautes latitudes. On enregistre de faibles variations annuelles, de l'ordre de 2°C à l'équateur et de 8°C à 40° de latitude.

L'effet des variations de températures dans l'océan, se traduit par des modifications des constantes apparentes K_1 , K_2 , et donc des déplacements des équilibres chimiques. Son effet apparaît également (comme on l'a mentionné dans la **section 1.5**) dans la variation de la solubilité du CO_2 , α , qui augmente avec la diminution de la température.

1.7 La salinité des océans

La quantité moyenne de sel présente dans l'océan global, est de l'ordre de 35 g dans chaque 1000 g d'eau de mer. Cette quantité varie de 33 à 37 dans l'océan ouvert. Dans les mers semi fermées, on retrouve des salinités beaucoup plus faibles, comme par exemple en Mer Baltique, où la salinité est de moins de 10, ou en Mer Noire où elle varie de 18 à 22, et des salinités plus fortes, comme en Mer Méditerranée, où elle va de 37 à 39, et en Mer Rouge, où elle dépasse 40. Dans la couche de surface, les eaux sont plus salées dans les zones tropicales, et moins salées à l'équateur et aux hautes latitudes.

Plusieurs processus naturels sont considérés comme régulateurs de sel, en gardant l'équilibre de ce qui rentre et ce qui sort de l'océan. L'océan se débarrasse donc annuellement d'une quantité de sel égale à celle que lui apportent les cours d'eau. L'altération superficielle des roches produit des minéraux, dont le sel, qui est ensuite transporté vers l'océan à travers les rivières. L'évaporation de l'eau à la surface de l'océan ainsi que la formation des glaciers, contribuent à l'augmentation de cette quantité. Ces facteurs sont contrebalancés par des phénomènes qui réduisent la salinité, à savoir, les précipitations de pluie et de neige, la fonte des glaces, ainsi que l'apport continu d'eau douce venant des rivières.

De petites variations dans la concentration de ces sels dissous, a de grands effets sur le cycle de l'eau et la circulation dans l'océan. La présence de sel dans l'océan, engendre des modifications dans certaines propriétés, telles que la densité, le point de congélation, la température du maximum de densité, la conductivité ... Les variations de température et de salinité sont fortement liées entre elles.

1.8 Puits et sources de CO₂

À l'interface air-mer, l'échange de CO₂ entre l'océan et l'atmosphère s'effectue d'une façon continue. Ce flux d'échange de gaz dépend de la différence des fugacités de CO₂ entre l'océan et l'atmosphère ($\Delta f\text{CO}_2$), il se calcul ainsi :

$$F = k \cdot \alpha \cdot (f\text{CO}_2^{\text{mer}} - f\text{CO}_2^{\text{air}}) \quad (1.18)$$

où k est la vitesse de transfert du CO₂, plusieurs formules existent pour calculer ce paramètre [Liss et Merlivat, 1986; Wanninkhof, 1992; Wanninkhof et McGillis, 1999; Nightingale *et al.*, 2000; Jean-Baptiste *et al.*, 2002; Ho *et al.*, 2006; Sweeney *et al.*, 2007; Wanninkhof *et al.*, 2009; Ho *et al.*, 2011; Wanninkhof, 2014] qui est lié à des facteurs océaniques comme la présence de glaces qui freinent les échanges air-mer, et atmosphériques comme les vents qui les intensifient. α est le coefficient de solubilité (**Section 1.5 § 1**). À l'équilibre et en l'absence du facteur anthropique, le flux résultant est nul. Le signe de $\Delta f\text{CO}_2$ détermine le sens du flux, celui-ci dépend de l'état de saturation des eaux superficielles par rapport à l'atmosphère, le flux suit donc le sens qui a tendance à réduire l'écart. Quand le $\Delta f\text{CO}_2$ est positif, le flux est donc vers l'atmosphère, et l'océan est considéré comme une "source" de carbone, dans le cas contraire, le flux est vers l'océan, on parle donc d'un "puits" de carbone.

Les études ont montré que la concentration du CO₂ dissous à la surface des océans présente une importante variabilité spatiotemporelle [Komhyr *et al.*, 1985; Lee *et al.*, 1998; Takahashi *et al.*, 1997, 2009]. En général, dans les deux hémisphères, les zones entre 14° et 50° sont considérées comme les principaux puits de CO₂ [Takahashi *et al.*, 1997, 1999, 2009, 2014; Rödenbeck *et al.*, 2013; Zeng *et al.*, 2014], avec un flux de -0,70 PgC.an⁻¹ au nord et -1,05 PgC.an⁻¹ au sud [Takahashi *et al.*, 1999]. Comme expliqué dans la **Section 1.5.3**, la forte densité des eaux de surface dans les hautes latitudes fait que les masses d'eau chargées de CO₂ plongent en profondeurs, ce qui va entraîner une séquestration du CO₂. Contrairement à ces zones, les régions équatoriales se comportent comme des sources de CO₂ pour l'atmosphère [Takahashi *et al.*, 1997, 1999, 2009, 2014; Rödenbeck *et al.*, 2013; Zeng *et al.*, 2014]. Ceci s'explique par le fait que les eaux qui ont plongé au fond, vont remonter à la surface des océans, dans les régions dites "zones d'*upwelling*" (voir **Figure 1.4**). Ces eaux riches en CO₂ font que ces zones là deviennent des sources de CO₂ pour l'atmosphère.

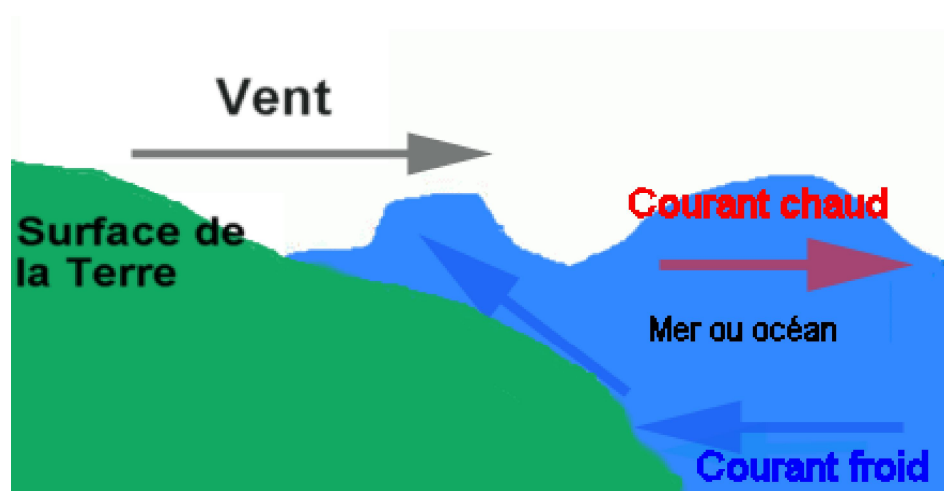


FIGURE 1.4 – Illustration du processus de l'Upwelling, ou remontée d'eau froide. [Source : <http://www.meteolafleche.com/>]

1.8.1 Impacts du CO₂ sur l'écosystème marin

La forte présence de CO₂ dans l'océan résulte en "*l'acidification des océans*". Comme présenté dans la **Figure 1.5**, ce phénomène se traduit par la diminution progressive du pH de l'océan au fil des décennies (**Section 1.3**). En 250 ans, l'acidification des océans a augmenté de 30% (baisse du pH de 8,2 à 8,1 depuis 1751 ; Gattuso et Lavigne, 2009 ; Dupont et Pörtner, 2013 ; IPCC, 2014, et une diminution de plus de 0,4 unité est prévue pour la fin de ce siècle ; Jacobson, 2005). Cette acidification n'est pas sans conséquence pour l'écosystème marin, quand le CO₂ est absorbé par l'océan, il y a non seulement une diminution du pH, mais aussi une diminution de la concentration des ions CO₃²⁻ et du CaCO₃. Ces composants sont nécessaires à plusieurs organismes marins pour fabriquer leur squelette ou coquilles calcaires, tels que les coraux, les moules, les huîtres ... Ce phénomène a des effets dramatiques sur toute la chaîne alimentaire dans l'océan, et donc sur tout l'écosystème marin. Ainsi, la compréhension du flux de CO₂ échangé entre

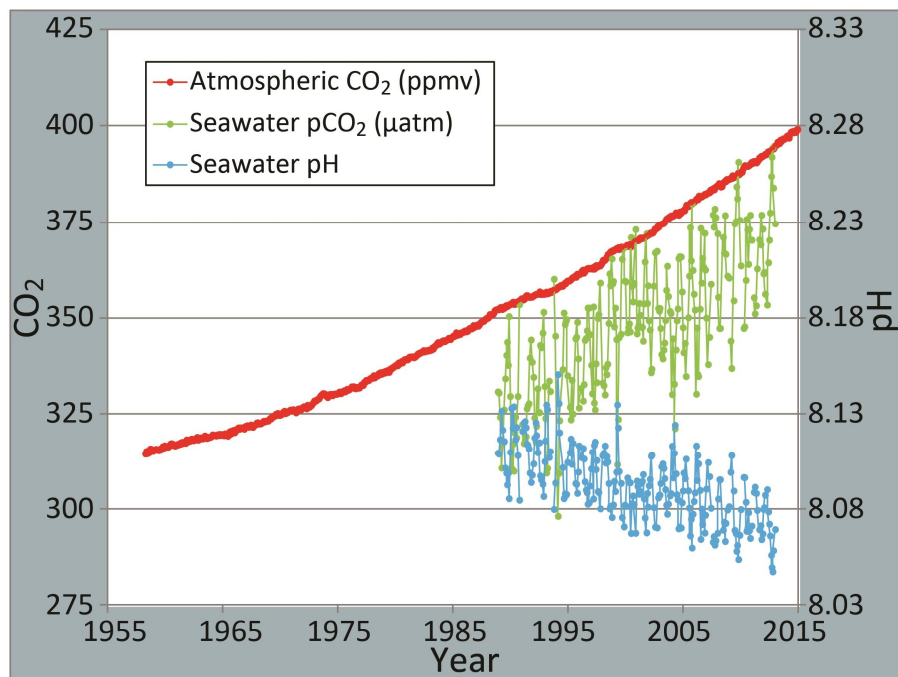


FIGURE 1.5 – Graphe représentant l'augmentation du CO₂ dans l'atmosphère associée à une augmentation de fCO₂ de l'océan, et donc une diminution du pH (Station Mauna Loa). [PMEL Carbon Group, 2016]

l'océan et l'atmosphère est essentiel pour évaluer le rôle de l'océan dans la variabilité et les dynamiques du climat terrestre et marin.

1.9 L'état actuel de l'océan quant au CO₂ dissous

Les taux variables de CO₂ échangé entre l'océan et l'atmosphère, représente l'un des facteurs clés quant à la régulation de l'augmentation de la concentration du CO₂ dans l'atmosphère. Les estimations faites par Le Quéré *et al.* [2015], ont montré que les émissions de CO₂ issues de la combustion des énergies fossiles ainsi que la production du ciment (les deux principales sources de CO₂ anthropique) ont augmenté d'environ 65% par rapport aux années 1990, ils ont atteint $9,9 \pm 0,5$ PgC en 2013, 2,3% de plus que l'année 2012 (d'après Friedlingstein *et al.* [2014], ce taux a augmenté d'une moyenne de 2,5% par année durant la dernière décennie). Un taux de $10,1 \pm 0,6$ PgC a été prédit pour l'année 2014. Les émissions issues du changement d'utilisation des sols, et principalement la déforestation, ont été estimées à $0,9 \pm 0,5$ PgC, un taux qui reste plus ou moins stable d'une année à une autre ($0,9 \pm 0,5$ PgC en 2011 [Le Quéré *et al.*, 2012]). La contribution de cette source à l'augmentation du CO₂ anthropique, a diminué de 20% (1990 - 2000) à 12% en 2008 [Le Quéré *et al.*, 2009].

Le taux d'augmentation de la concentration du CO₂ atmosphérique a atteint $5,4 \pm 0,2$ PgC en 2013, une valeur plus élevée que la moyenne calculée pour la décennie 2004 - 2013, qui est de $4,3 \pm 0,1$ PgC. Les observations ainsi que les modèles océaniques ont montré que les puits océaniques de CO₂ ont atteint $2,9 \pm 0,5$ PgC en 2013, un taux plus élevé que la

moyenne estimée pour 2004 - 2013 ($2,6 \pm 0,5$ PgC). En ce qui concerne les puits terrestres, le taux était de $2,5 \pm 0,9$ PgC en 2013, ce qui est inférieur à la moyenne de 2004 - 2013 qui était de $2,9 \pm 0,8$ PgC.

Toutes ces estimations du taux des émissions de CO₂ ainsi que la capacité des puits à ralentir la vitesse du changement climatique, représentent le fruit des efforts de la communauté scientifique, qui tâchent par tous les moyens de faire des calculs en combinant les mesures, les estimations statistiques, ainsi que les résultats des modèles océanographiques. Néanmoins, ces estimations sont sujettes à de grandes incertitudes, liées à la grande variabilité que présentent les processus physiques et biologiques, et aussi au manque de données en terme de temps et d'espace. Dans cette thèse nous allons nous intéresser à la partie océan. Afin d'estimer la capacité des océans à agir comme un puits pour l'augmentation du CO₂ atmosphérique, il faut surveiller la variation de la fCO₂ de surface des océans, sur des échelles de temps mensuelles et saisonnières. L'étude que je présente dans cette thèse se focalise sur la partie tropicale d'un océan bien particulier, considéré comme le plus important puits de CO₂, assurant environ 60% de l'absorption de l'océan global [Takahashi *et al.*, 1997]. Cet océan est "l'*Atlantique*".

Chapitre 2

ZONE D'ÉTUDE : ATLANTIQUE TROPICAL

L'océan Atlantique est le deuxième plus grand océan de la planète après l'océan Pacifique, il s'étend sur une superficie d'environ 106.400.000 km², avec une profondeur moyenne avoisinant les 4000 m, et un volume d'environ 323.600.000 km³. D'après Takahashi *et al.* [2002] au Nord de 50°S, cet océan représente approximativement 24 % de la superficie de l'océan global, et absorbe environ 60% de la quantité global du CO₂ atmosphérique, ce qui le rend un des plus importants puits de CO₂ pour la planète [Sabine *et al.*, 2004; Mikaloff Fletcher *et al.*, 2006, 2007; Gruber *et al.*, 2009; Wanninkhof *et al.*, 2010]. L'étude faite par Landschützer *et al.* [2013] sur la zone qui s'étend de 44°S à 79°N, entre 1998 et 2007, a montré que la moyenne décennale du flux de CO₂ absorbé par cette région océanique était de $0,45 \pm 0,15$ PgC.an⁻¹, qui se distribue sur trois zone : la région se trouvant au Nord de 18°N, présente un puits de $0,39 \pm 0,10$ PgC.an⁻¹, la zone entre 18°N et 18°S est une source de $0,11 \pm 0,07$ PgC.an⁻¹, et enfin, au Sud de 18°S, un puits de $0,16 \pm 0,06$ PgC.an⁻¹ a été trouvé.

Dans cette étude je vais me concentrer sur la partie tropicale de cet océan, qui est bornée par les deux tropiques : Cancer et Capricorne, de 23,5°S à 23,5°N (voir **Figure 2.1**). Cette zone est caractérisée par des processus fortement dynamiques, causés par les variations saisonnières des courants de surface. Comme on le voit sur la **Figure 2.2**, plusieurs études ont montré que la partie équatoriale de l'océan Atlantique présente tout au long de l'année, une source de CO₂ pour l'atmosphère [Takahashi, 1961; Takahashi *et al.*, 1999, 2009, 2014; Goyet *et al.*, 1998; Lefèvre *et al.*, 1998; Landschützer *et al.*, 2013; Zeng *et al.*, 2014], causée par l'upwelling équatorial qui ramène à la surface les eaux riches en CO₂ (**Section 1.8**). Cependant, plusieurs études ont montré que la région se trouvant à l'embouchure du fleuve de l'Amazone (rectangle noir dans la **Figure 2.2**), se comporte en été comme un puits, en éliminant une partie du CO₂ atmosphérique [Nittrouer et DeMaster, 1996; Ternon *et al.*, 2000; Fratantoni et Glickson, 2002; Körtzinger, 2003; Lefèvre *et al.*, 2010; Park et Wanninkhof, 2012; Lefèvre *et al.*, 2014]. Ce phénomène s'explique par la forte activité biologique qui se produit quand l'eau du fleuve Amazone se mélange avec celle de l'océan. Ce processus induit une diminution du carbone inorganique total et donc de la fCO₂ dans l'eau de surface [DeMaster *et al.*, 1996; Ternon *et al.*, 2000].

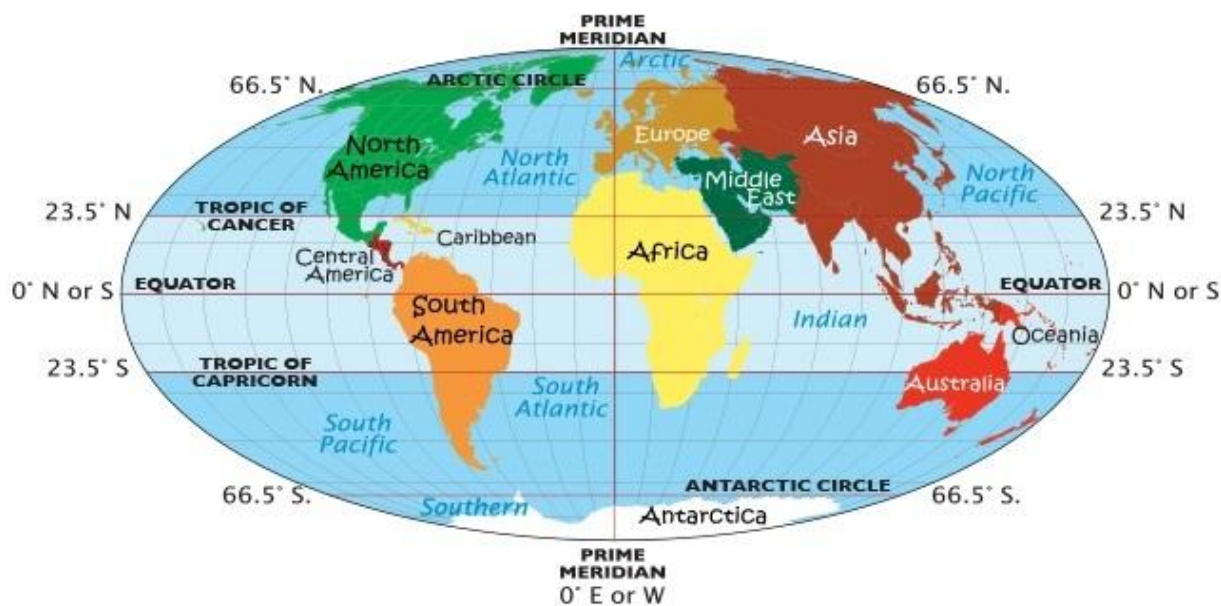


FIGURE 2.1 – Carte du monde représentant les lignes imaginaires de la terre. [Worldatlas, 2016]

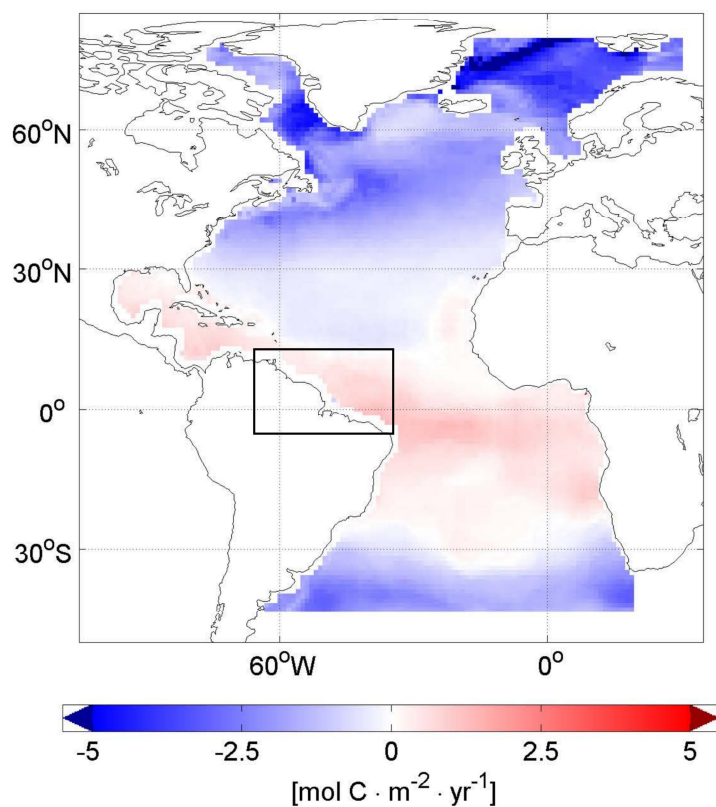


FIGURE 2.2 – Flux de CO_2 dans l'océan Atlantique. Les valeurs négatives représentent les puits de CO_2 . Le carré noir représente l'embouchure du fleuve de l'Amazonie. [Figure 6.(b) dans Landschützer *et al.*, 2013]

La **Figure 2.3** montre bien qu'aux mois de mai et juin, le flux négatif du CO₂ est associé avec la diminution de la fCO₂ de surface.

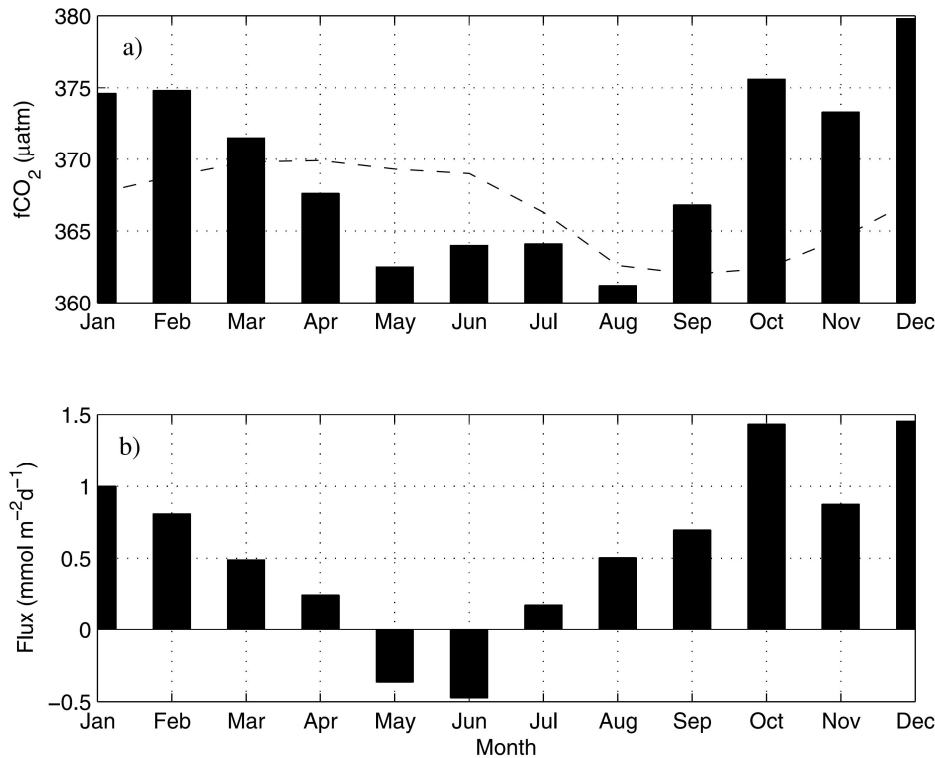


FIGURE 2.3 – Relation entre fCO₂ et le flux air-mer de CO₂ dans l'ouest de l'océan Atlantique tropical : (a) Les moyennes mensuelles de fCO₂, le trait en pointillé représente la fCO₂ atmosphérique. (b) Flux mensuel de CO₂, les valeurs négatives représentent des zones puits de CO₂ vis à vis de l'océan. [Figure 7. dans Lefèvre *et al.*, 2010]

Dans ce chapitre, on va aborder les caractéristiques principales de cette zone, qui vont nous permettre par la suite, de déterminer les processus à prendre en compte lors de l'interpolation de la fCO₂.

Avant d'expliquer la distribution des principaux paramètres (SST, SSS, fCO₂, ...), on doit parler des courants de surface qui subissent la forte influence des courants atmosphériques et en particulier de l'"*InterTropical Convergence Zone*" (ITCZ, Zone de Convergence InterTropicale).

2.1 Zone de Convergence InterTropicale : ITCZ

Comme illustré dans la **Figure 2.4**, les rayons du soleil réchauffent les eaux superficielles, ce qui forme une zone humide de basse pression. Quand les alizés du Nord et du Sud se rencontrent près de l'équateur, ils convergent en poussant cet air humide vers le haut. La vapeur d'eau va ainsi être condensée et forme une bande de fortes précipitations autour du globe (bande rouge dans la **Figure 2.4**). La position de l'ITCZ connaît un cycle saisonnier qui suit la position de la terre par rapport au soleil (l'inclinaison de l'axe de la terre). Elle atteint sa position la plus au Nord entre juillet et septembre ($\sim 5^\circ\text{N}$ à

l'Ouest du bassin, et $\sim 15^\circ\text{N}$ à l'Est, **Figure 2.5** en rouge), et sa position la plus au Sud entre janvier et mars [$\sim 0^\circ\text{N}$ à l'Ouest du bassin, et $\sim 5^\circ\text{N}$ à l'Est, **Figure 2.5** en bleu ; Merle *et al.*, 1979; Mitchell et Wallace, 1992; Biasutti *et al.*, 2003]

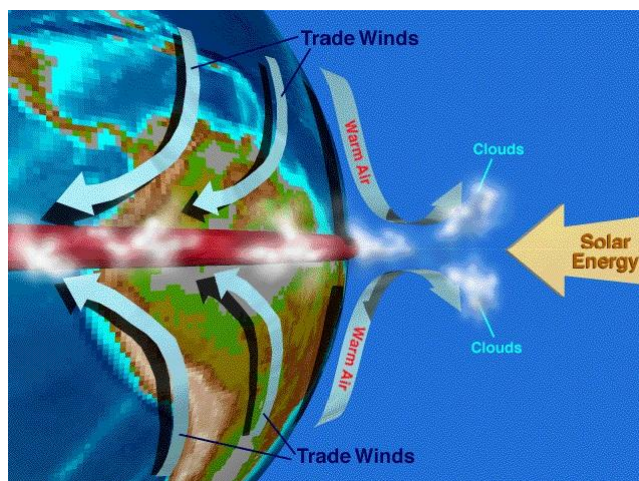


FIGURE 2.4 – La zone de convergence intertropicale. [Source : www.alevelgeography.com]

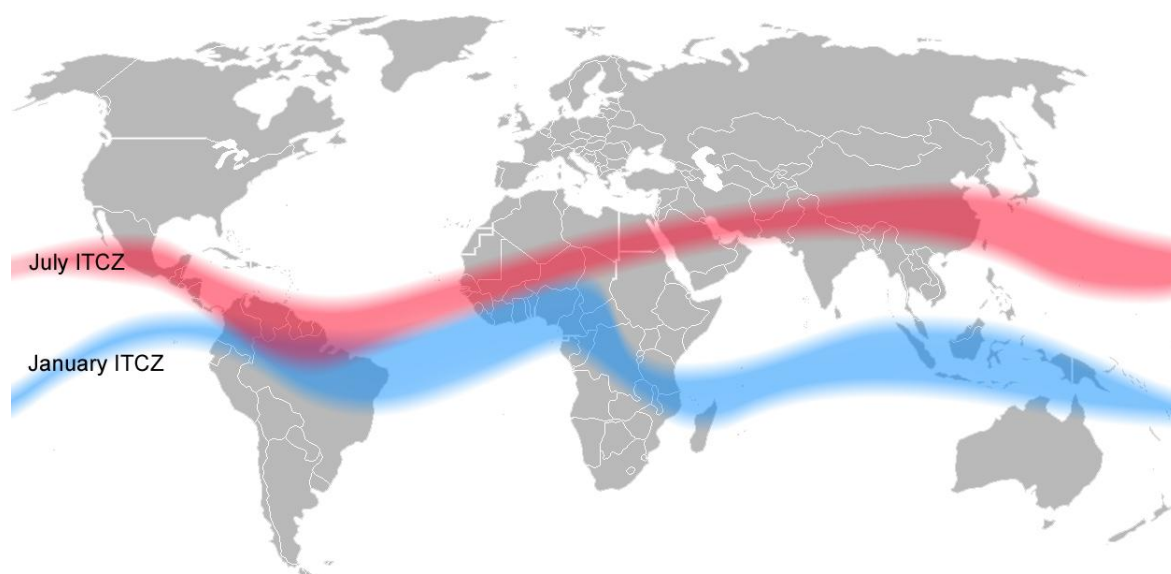


FIGURE 2.5 – La position de l'ITCZ pendant le mois de juillet et le mois de janvier. [Halldin, 2006]

2.2 Les courants de surface

Dans cette partie de l'océan Atlantique, les courants marins présentent de fortes variabilités saisonnières, dominées par les mouvements des alizés et par la position de l'ITCZ

[Philander et Pacanowski, 1986; Stramma et Schott, 1999]. La **Figure 2.6** présente les principaux courants de ce bassin :

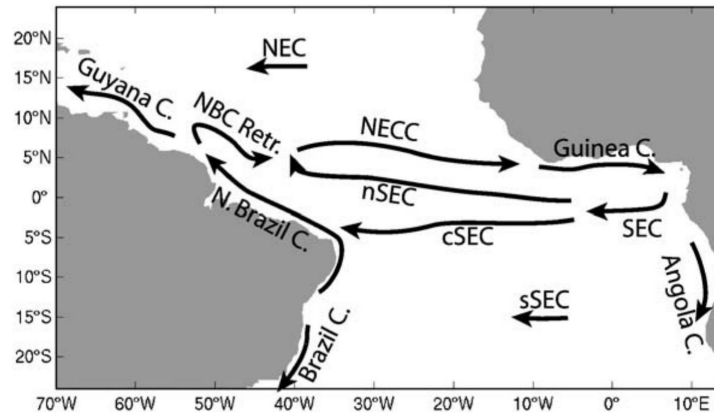


FIGURE 2.6 – Les principaux courants marins de la zone Atlantique tropical. [Figure 1. dans Lumpkin et Garzoli, 2005]

2.2.1 Courant équatorial Sud (SEC)

Le SEC (*South Equatorial Current*) s'écoule vers l'Ouest dans la bande 4°N et 15-25°S, et se sépare en 3 branches :

1. sSEC (*southern SEC*) : s'écoule au Sud de 10°S à une vitesse d'environ 10 à 15 cm.s^{-1} [Philander et Pacanowski, 1986], et forme le BC (*Brazil Current*) le long de la côte Sud Américaine.
2. cSEC (*central SEC*) : ce courant se trouve entre 3-5°S et 6-9°S [Peterson et Stramma, 1991]. Il se divise en deux parties à $\sim 5-6^\circ\text{S}$, une partie Nord-Ouest qui rejoint le NBC (*North Brazil Current*), et une partie Sud-Ouest qui se fusionne avec le flux Nord de la branche la plus au Sud du sSEC [da Silveira *et al.*, 1994].
3. nSEC (*northern SEC*) : s'écoule vers le Nord, et alimente la partie supérieure du NBC à 2-3°N, 45-46°O. Ce courant est très rapide au niveau de l'équateur [$37 \pm 10 \text{ cm.s}^{-1}$; Richardson et Walsh, 1986; Lumpkin et Garzoli, 2005].

2.2.2 Courant Nord Brésil (NBC)

Le NBC est un courant alimenté par le SEC, et plus précisément par le cSEC. Il commence généralement vers 10°S et transporte les eaux chaudes originaires du Sud de l'Atlantique, vers le Nord-Ouest le long de la côte Brésilienne. Après avoir traversé l'équateur, une partie de ce courant est transportée vers l'Est et forme le NECC (*North Equatorial CounterCurrent*), et une deuxième partie continue vers le Nord-Ouest et forme le GC (*Guyana Current*) et le CC [*Caribbean Current*; Richardson et Walsh, 1986]. Les vitesses de ce courant sont de l'ordre de 60 à 100 cm.s^{-1} [Bourles *et al.*, 1999].

2.2.3 Contre courant équatorial Nord (NECC)

Le NECC est formé principalement par le NBC, et aussi par une partie du nSEC [Richardson et Walsh, 1986]. Il est fortement lié à la variation saisonnière de l'ITCZ. Pendant l'été-automne, quand l'ITCZ est à sa position la plus au Nord, le NECC s'écoule vers l'Est entre ~ 3 et 10°N et alimente une partie du GC (*Guinea Current*), durant cette période, son intensité atteint son maximum [$>50 \text{ cm.s}^{-1}$; Arnault, 1987]. Durant l'hiver-printemps, l'ITCZ est à son extrême Sud, et l'intensité du NECC atteint son minimum [5 à 10 cm.s^{-1} ; Lumpkin et Garzoli, 2005].

Dans l'océan Atlantique tropical, les plus fortes variabilités saisonnières, ont été enregistrées dans le NECC et dans la partie équatoriale du SEC [Peterson et Stramma, 1991].

2.2.4 Courant équatorial Nord (NEC)

Le NEC (*North Equatorial Current*) est un courant large et alimenté principalement par les eaux froides venant du Nord-Est de l'océan Atlantique. Il s'écoule vers l'Ouest au Nord de 10°N , avec une vitesse variant de 10 à 12 cm.s^{-1} en printemps-automne, et qui augmente jusqu'à 15 cm.s^{-1} en été boréal [Richardson et Walsh, 1986; Arnault, 1987]. La globalité de ce flux est dirigée vers le Nord-Ouest où elle alimente le GC et le CC. Cependant, une partie de l'eau subie une réflexion cyclonique vers l'Est et rejoint le NECC [Bourles *et al.*, 1999].

2.3 Zones d'upwelling

La partie Est de l'océan Atlantique tropical est caractérisée par deux types d'upwellings [Voituriez et Herbland, 1982] :

1. des upwellings côtiers permanents (au nord de 20°N et au sud de 15°S), et des upwellings côtiers saisonniers (entre 12 et 20°N), principalement au sud du Cap-Vert (côte Sénégalaise) et au niveau du Golfe de Guinée; la côte Nord du Golfe de Guinée (Côte d'Ivoire, Bénin, ...), et la côte Sud du Golfe de Guinée (Gabon, Angola, ...; voir **Figure 2.7**).
2. un upwelling équatorial (au niveau de la bande équatoriale).

Ces upwellings induisent une diminution de la SST de l'ordre de 5 à 7°C [Hastenrath et Lamb, 1978], et une augmentation de la Chl-*a* qui peut dépasser $0,8 \text{ mg.m}^{-3}$ [Parard, 2011]. On distingue :

1. Upwelling du Sénégal : prend place entre la Mauritanie et la Guinée (**Figure 2.7**) entre octobre et juin.
2. Upwelling du Golfe de Guinée :
 - a) Côtier : deux upwellings se produisent le long de la côte; 1) au Nord de l'équateur entre 8°O et 2°E sur une échelle saisonnière. Entre juin et octobre l'upwelling est fort, et il est plus faible entre décembre et mars. 2) au Sud de l'équateur, il se produit entre juillet et septembre.
 - b) Équatorial : il se produit entre juillet et septembre dans la bande équatoriale 2°N - 4°S , et il s'étend jusqu'à 20°O .



FIGURE 2.7 – Le Golfe de Guinée qui se situe dans la côte Ouest du continent Africain. [Source : www.thisissierreleone.com]

2.4 La SST et les modes de variabilité climatique

Dans l'océan Atlantique tropical, il y a deux modes qui dominent la variabilité de la SST : un mode équatorial, et un mode dipolaire.

2.4.1 Mode équatorial

Appelé aussi "Atlantic Niño" en raison de sa similarité avec le phénomène ENSO (El Niño-Southern Oscillation) de l'océan Pacifique tropical (pour plus de détail sur ENSO voir Wang *et al.* [2012]). Ce mode se manifeste suivant une échelle de temps inter-annuelle (allant de 2 à 5 ans [Carton *et al.*, 1996]), et résulte en une alternance de deux épisodes de SST, chaud et froid. À l'équateur, les plus fortes variabilités de SST apparaissent à l'Est du bassin tropical. L'épisode chaud (en printemps boréal) est associé à une diminution de l'intensité des vents équatoriaux, ce qui provoque un approfondissement de la thermocline¹, et une diminution de l'intensité des upwellings dans le Golfe de Guinée, et donc une augmentation de la SST (elle atteint son maximum [$>27^{\circ}\text{C}$]) dans cette zone.

À l'inverse, l'épisode froid (en été boréal) est lié à une intensification des alizés le long de l'équateur, ceci est associé à une remontée de la thermocline et une intensification des upwellings côtiers, la SST va donc atteindre son minimum le long de la côte Ouest de l'Afrique [Carton *et al.*, 1996; Carton et Zhou, 1997; Xie et Carton, 2004]. Dans cette même période, une langue d'eau froide se forme tout au long du bassin près de l'équateur (**Figure 1.b** dans Xie et Carton [2004]).

1. Une très fine couche qui marque la séparation entre les eaux de surfaces (chaudes) et les eaux profondes (froides).

2.4.2 Mode dipolaire (ou inter-hémisphérique)

Ce mode exprime la relation qui existe entre la variabilité de la SST et la position de l'ITCZ. Il se produit sur une échelle de temps qui va de trois à dix ans, et se caractérise par une migration anormale de l'ITCZ qui se déplace au-delà des positions saisonnières habituelles. Un déplacement inhabituel vers le Nord, provoque l'augmentation de la SST au Nord et sa diminution au Sud, ce qui entraîne une sécheresse au Nord-Est du Brésil, et de forte précipitations au Sahel [Kushnir *et al.*, 2006]. Le mécanisme inverse apparaît quand l'ITCZ se déplace vers le Sud.

La SST varie d'un courant marin à un autre, on peut citer :

- SEC : La température de surface moyenne enregistrée dans le SEC, varie entre 26 et 28°C.
- NBC : La moyenne de SST varie de 22 à 29°C.
- NECC : En printemps-été la SST dépasse 24°C. À la fin de l'été, l'eau devient plus froide, et la SST est inférieur à 15°C.
- NEC : Durant l'été-automne, la SST connaît des valeurs entre 30 et 32°C. Pendant l'hiver, la SST varie entre 24 et 28°C.

2.5 La salinité de surface (SSS)

Dans l'océan Atlantique tropical, on constate une forte SSS (>37) au niveau des eaux subtropicales des deux hémisphères (Nord et Sud), ces dernières sont séparées par la partie équatoriale qui connaît des salinités entre 34,5 et 35 [Tzortzi *et al.*, 2013]. Il existe une variabilité saisonnière dans certains endroits du bassin [Reverdin *et al.*, 2007]. Les plus fortes variabilités ont été observées dans la partie Ouest du Nord de l'océan Atlantique tropical, principalement dues à l'advection horizontale² et aux apports fluviaux de l'Amazone et de l'Orénoque [Foltz et McPhaden, 2008]. Dans ce bassin, l'eau est plus salée en janvier (elle atteint un minimum de 35), contrairement au mois de juillet (elle descend jusqu'à 29,7). La variabilité de SSS est aussi forte au niveau de l'Est du bassin, le cycle saisonnier est dû aux apports fluviaux du fleuve du Niger et du Congo, à l'advection horizontale et verticale³, mais aussi aux précipitations causées par l'ITCZ. La SSS descend jusqu'à 28,8 en janvier, et elle est plus élevée en mois de juillet (elle atteint un minimum de 33,5). Au niveau des deux bassin (Est et Ouest) la différence de SSS entre le mois de janvier et le mois de juillet dépasse 6,5 [Figure 2.8 ; Tzortzi *et al.*, 2013]. Au Nord et au Sud les variations sont moins importantes, et sont principalement dues à l'évaporation et aux précipitations [Tzortzi *et al.*, 2013; Da-Allada *et al.*, 2013].

Les valeurs de SSS propres à chaque courant sont les suivantes :

- SEC : La SSS enregistré dans le sSEC est aux alentours de 37, et celle du cSEC et du nSEC est de l'ordre de 36.
- NBC : La moyenne de SSS est comprise entre 35 et 36,7.

2. Transport horizontal de l'eau (à travers les courants marins) qui diminue la salinité.

3. Transport vertical de l'eau (à travers les courants marins) qui augmente la salinité.

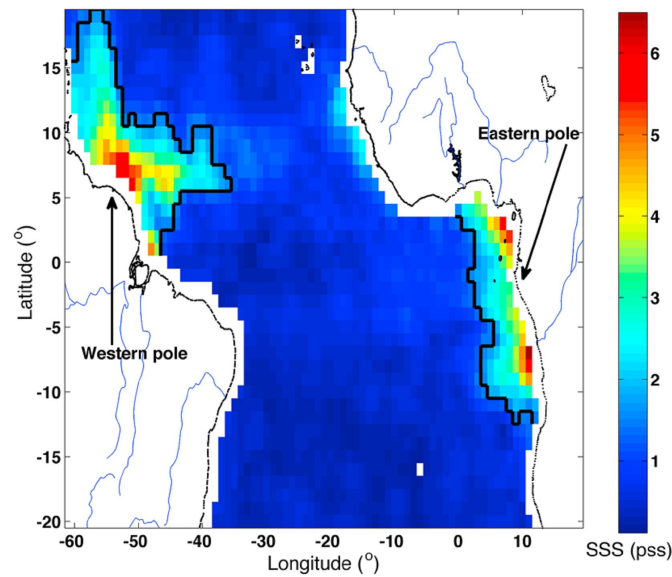


FIGURE 2.8 – La différence de SSS entre janvier et juillet (maximum moins minimum) dans l’océan Atlantique tropical, les deux lignes noires en gras représentent les deux régions où la différence de SSS dépasse 1,5 unité. [Figure 3. dans Tzortzi *et al.*, 2013]

- NECC : La SSS est généralement faible en printemps-été ($\sim 34,8$), et plus élevée en automne-hiver ($\sim 36,2$).
- NEC : Durant l’été-automne, la SSS est autour de 36,4. Pendant l’hiver, la SSS est d’environ 35,2.

2.6 La Chlorophylle-*a*

La Chlorophylle-*a* (Chl-*a*) représente le pigment le plus important présent lors de la photosynthèse du phytoplancton. Sa concentration est utilisée comme un indicateur de biomasse et de production primaire⁴ du phytoplancton [Ryther et Yentsch, 1957]. La Chl-*a* connaît de fortes variabilités spatiotemporelles. Les zones subtropicales sont connues pour leur faibles concentrations en Chl-*a*. Les fortes concentrations en Chl-*a* se trouvent au niveau des régions côtières (présence d’un taux élevé de sels nutritifs), ainsi que dans les zones d’upwellings.

Dans l’océan Atlantique tropical, la distribution de Chl-*a* est reliée aux eaux déversées par les deux fleuves Congo et Amazone, qui, tout au long de l’année contiennent de très fortes valeurs de Chl-*a* (entre 5 et 10 $\text{mg}\cdot\text{m}^{-3}$), ainsi qu’à l’upwelling équatorial et les upwellings côtiers. Dans la partie équatoriale de l’océan Atlantique, les variabilités saisonnières de la production primaire sont considérables. Pendant la saison froide (de juin à octobre), l’upwelling équatorial est présent, et la Chl-*a* à la surface de l’océan est maximale. Durant la saison chaude (entre octobre et mai), la Chl-*a* à la surface de l’eau est faible. À l’Est du bassin, dans la zone 0-10°S, 0-20°O, les concentrations de Chl-*a* sont comprises entre 0,16 $\text{mg}\cdot\text{m}^{-3}$ (entre février et avril) et 0,52 $\text{mg}\cdot\text{m}^{-3}$ (en août). Dans la partie Ouest du bassin, à 10°N-10°S, 20°O-30°O, les variabilités sont moins importantes,

4. La formation de la matière organique lors de la photosynthèse.

la concentration minimale ($0,15 \text{ mg.m}^{-3}$) survient en mois d'avril, et la concentration maximale ($0,24 \text{ mg.m}^{-3}$) apparaît en août [Perez *et al.*, 2005].

2.7 La fugacité du CO_2

Depuis les années 1980, les chercheurs scientifiques ont commencé à s'intéresser à la variation de la $f\text{CO}_2$ dans l'océan Atlantique tropical, ce qui a permis de classer cette zone parmi les zones caractérisées par une $f\text{CO}_2$ élevée. Andrié *et al.* [1986] ont analysé les données des campagnes en mer de juillet 1982 à août 1984, le long de 4°O , 22°O et 35°O entre 5°N et 5°S . Ils ont observé que la $f\text{CO}_2$ au Sud est généralement supérieure à la $f\text{CO}_2$ au Nord. Cela est dû, d'une part à l'upwelling équatorial, et d'une autre part à la SST (responsable de la diminution ou de l'augmentation de la solubilité du CO_2). En 1986, Oudot et Andrié ont analysé les variabilités spatiotemporelles de $f\text{CO}_2$ dans la même zone, en janvier et juillet 1983 [Oudot et Andrié, 1986]. Selon leurs résultats, pendant les deux saisons, chaude et froide (janvier et juillet), les eaux de surface de l'Est du bassin, se réchauffent quand elles sont entraînées vers l'Ouest par le SEC, ce qui résulte en une $f\text{CO}_2$ plus élevée à l'Ouest qu'à l'Est. Ces résultats ont été confirmés par la suite par Goyet *et al.* [1998], qui ont montré que cette même zone s'est comportée comme une source de CO_2 pour l'atmosphère pendant le printemps de 1994. Néanmoins, ils ont trouvé que la $f\text{CO}_2$ dans le NECC était faible et inférieure à la $f\text{CO}_2$ atmosphérique, ce qui rend ce courant un puits de CO_2 pour l'atmosphère. Les travaux de Lefèvre *et al.* [1998] ont abouti à la même conclusion quant à la faible présence de CO_2 dans le NECC (durant plusieurs saisons de l'année 1995), où un $\Delta f\text{CO}_2$ de $-70 \text{ } \mu\text{atm}$ a été mesuré près de 8°N . Cette faible $f\text{CO}_2$ est due d'une part à la faible SSS causée par les fortes précipitations associées à l'ITCZ, et d'autre part à la forte activité biologique.

En 1995, la partie Nord-Est de l'océan Atlantique tropical a fait l'objet d'une étude de Copin-Montégut et Avril 1995. Les analyses de ces données ont montré que la $f\text{CO}_2$ dans la zone côtière, représente de très fortes variabilités spatiotemporelles, influencées par l'upwelling de Mauritanie. En septembre-octobre 1991, la $f\text{CO}_2$ a varié entre 360 et $410 \text{ } \mu\text{atm}$, et en juin 1992, on a enregistré des $f\text{CO}_2$ comprises entre 300 et $450 \text{ } \mu\text{atm}$, représentant une très grande variabilité ($\sim 80 \text{ } \mu\text{atm}$ sur 44 heures d'observation). Dans l'océan ouvert, les valeurs de $f\text{CO}_2$ ainsi que sa variabilité sont généralement moins importantes. Elles sont dues à la variabilité de la SST. En septembre-octobre 1991 la $f\text{CO}_2$ était comprise entre 365 et $390 \text{ } \mu\text{atm}$, alors qu'en mai-juin, la $f\text{CO}_2$ variait entre 330 et $360 \text{ } \mu\text{atm}$.

Ternon *et al.* [2000] ont étudié la zone Ouest du bassin (la côte Sud-Américaine) pendant l'été-automne 1995. Ils ont enregistré de très faibles valeurs de $f\text{CO}_2$ ($<150 \text{ } \mu\text{atm}$) associées à de faibles SSS (<20), causées par les eaux déversées par le fleuve de l'Amazone. Au Nord de 5°N , au niveau de la réflexion du NBC, ces eaux se propagent jusqu'à 2000 km vers l'Est. Ces résultats ont été confirmés par la suite [Körtzinger, 2003; Lefèvre *et al.*, 2010].

La zone entre 19°N - 20°N , 65°O - 68°O entre 2000-2009, a été étudiée par Park et Wankhof en 2012, la $f\text{CO}_2$ de surface semblait avoir un cycle saisonnier régulier, contrôlé par les variations de SST ainsi que par les processus biologiques. Les valeurs maximales étaient constatées en été, où $\Delta f\text{CO}_2$ était positif, donc la zone s'est comportée comme

source. Pendant l'hiver, la $f\text{CO}_2$ était faible et le $\Delta f\text{CO}_2$ était négatif, ce qui signifie que la zone était un puits de CO_2 .

Lefèvre *et al.* [2014] ont analysé la distribution de $f\text{CO}_2$ au niveau des principaux courants, à l'Ouest et au centre du bassin (6°S - 15°N , 52°O - 24°O), en mars-avril 2009 et juillet-août 2010. Ils ont constaté qu'en mars-avril la $f\text{CO}_2$ est dominée par la SST, alors que la SSS a une grande influence en juillet-août. La $f\text{CO}_2$ varie entre des valeurs inférieures à $280 \mu\text{atm}$ (près de la Guyane Française), et des valeurs qui dépassent $420 \mu\text{atm}$ (dans le NEC en mois d'août). En raison des variations saisonnières, au Nord de 10°N la $f\text{CO}_2$ est élevée en août, et est faible en mars, quant à la partie qui est au Sud de $\sim 8^\circ\text{S}$, la $f\text{CO}_2$ est élevée en mars, et est faible en août. À l'échelle régionale, les variabilités de $f\text{CO}_2$ sont plus importantes. Le NBC, le SEC, et la partie Sud du NECC, sont caractérisés par la forte présence de CO_2 ($f\text{CO}_2^{\text{eau}} > f\text{CO}_2^{\text{air}}$). Cependant, la partie Nord du NECC a une $f\text{CO}_2^{\text{eau}}$ inférieure à celle de l'air. Ceci est dû aux eaux venant de l'Amazone, à la présence de l'ITCZ en juillet-août, ainsi qu'au refroidissement de l'eau au mois de mars. Affectée par la SST, au mois de mars, la $f\text{CO}_2^{\text{eau}}$ est généralement inférieur à la $f\text{CO}_2^{\text{air}}$ au niveau du NEC, et y est légèrement supérieur en août.

Afin d'avoir une quantification globale ou régionale du flux de CO_2 dans l'océan avec de bonnes précisions, il faut disposer d'un ensemble d'observations de $f\text{CO}_2$ ayant une résolution spatiotemporelle assez importante. La collection qui regroupe le plus de données de $f\text{CO}_2$ de surface est SOCAT [Bakker *et al.*, 2014], pour *Surface Ocean CO₂ Atlas*. Les données sont disponibles sur le site web : <http://www.socat.info/>.

2.8 La base de données SOCAT

L'idée de la création de SOCAT est née lors du workshop "*Surface Ocean CO₂ Variability and Vulnerability*" (SOCOVV) à UNESCO Paris, en avril 2007 [IOCCP, 2007], dans le but de mettre en place un ensemble global de données standardisées de $f\text{CO}_2$ de surface, accessible au public.

Comme on le voit dans la **Figure 2.9**, le nombre de mesures de CO_2 a largement augmenter durant les deux dernières décennies [Sabine *et al.*, 2010]. Dans cette étude j'ai utilisé deux versions : 1) la SOCAT V.1.4 publiée en septembre 2011, et contient 6.3 millions de données de $f\text{CO}_2$ de l'océan global, entre 1968 et 2007 [Pfeil *et al.*, 2013; Sabine *et al.*, 2013]. 2) la SOCAT V.3, publiée en septembre 2015 lors de la "*SOLAS Open Science Conference*" à l'Université de Kiel en Allemagne [SOCAT et SOCOM, 2015; Bakker *et al.*, 2016]. Cette base de données (BD) contient un nombre total de 14,5 millions de données de $f\text{CO}_2$, réparties sur la période allant de 1957 à 2014. La partie qu'on a extraite de la BD globale s'étend de $23,5^\circ\text{S}$ - 70°O à $23,5^\circ\text{N}$ - 15°E .

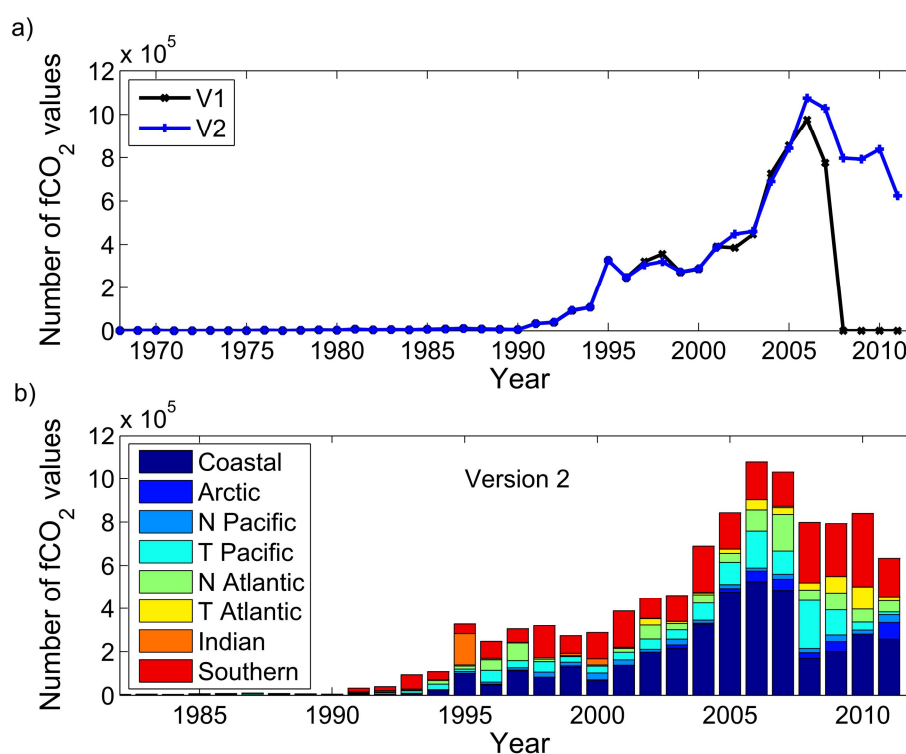


FIGURE 2.9 – Nombre de données de fCO₂ de surface. (a) Nombre par année dans les deux versions 1 et 2 de SOCAT, (b) Nombre par région par année dans la version 2. [Figure 1. dans Bakker *et al.*, 2014]

Bien que les données *in situ* représentent une mesure directe du paramètre qu'on souhaite étudier, leur seul inconvénient est leur couverture spatiotemporelle qui est très réduite. Il nous est donc difficile d'avoir une information complète sur la distribution spatiale et temporelle du paramètre en question. Cependant, l'autre source de données qui permet de pallier à cet inconvénient est "le satellite".

Ayant la capacité d'échantillonner tout le globe en seulement quelques jours, jusqu'à plusieurs fois par jour, les satellites nous offrent la possibilité d'avoir des données de plusieurs paramètres utilisés en océanographie, avec des résolutions spatiales de quelques kilomètres. Ainsi, pour bien profiter des avantages des données satellitaires, les données *in situ* sont utilisées comme une référence qui sert non seulement à les valider, mais aussi à modéliser des quantités non mesurables par satellites, comme la fCO₂.

Chapitre 3

L'OCÉAN VU DE L'ESPACE

L'expression anglaise "*remote sensing*", est apparue dans les années soixante aux États-Unis. En 1973 en France, le Centre National d'Études Spatiales (CNES, [cnes.fr/fr](http://cnes.fr)) reconnaît officiellement le terme *télé-détection*. D'après le Centre canadien de cartographie et d'observation de la Terre [2015], la définition du terme *télé-détection* est :

« La télé-détection est la technique qui, par l'acquisition d'images, permet d'obtenir de l'information sur la surface de la Terre sans contact direct avec celle-ci. La télé-détection englobe tout le processus qui consiste à capter et à enregistrer l'énergie d'un rayonnement électromagnétique émis ou réfléchi, à traiter et à analyser l'information, pour ensuite mettre en application cette information. »

3.1 Les débuts de la télé-détection satellite

La télé-détection satellite a commencé avec l'ère de l'exploration de l'espace (les années cinquante), durant laquelle les premiers satellites ont été lancés. Le monde de la télé-détection est marqué par de grandes dates, citons-en quelques unes :

- 1957 : Lancement du premier satellite appelé **Sputnik 1** (terme russe signifiant 'satellite') par l'Union Soviétique (URSS).
- 1960 : Mise en orbite du premier satellite météorologique **Tiros** (*Television Infrared Observation Satellite*) par la NASA.
- 1972 : Lancement du premier satellite dédié à l'observation des ressources terrestres, **ERTS-1** (*Earth Resources Technology Satellite*) par la NASA, rebaptisé **Landsat** en 1975.
- 1978 : Le premier satellite spécialisé dans l'observation de l'océan, **Seasat**, fut lancé par la NASA.
- 1986 : Début de la commercialisation des images satellite (par la société Spot Image), après le lancement de **SPOT-1** (*Système Probatoire d'Observation de la Terre*) par le CNES.

- 1991 : Lancement du satellite européen **ERS-1** (*European Remote-sensing Satellite*) par l'ESA (*European Space Agency*) dans le but de l'étude de l'environnement global de la planète.

Le développement des technologies de la télédétection continue toujours, et est marqué par l'amélioration continue de la résolution spatiale des capteurs, ainsi que les techniques d'acquisition des images, et ceci grâce au développement des technologies de l'information, et à l'augmentation de la puissance de calcul des ordinateurs.

L'atmosphère a été le premier domaine ayant bénéficié des avantages de la télédétection, avec des applications directes sur la météorologie et la climatologie, notamment pour des prévisions de l'évolution du climat. Les études cartographiques, géographiques et géologiques, elles aussi, reposent beaucoup sur les images satellitaires, qui sont par la suite numérisées et sont intégrées dans ce qu'on appelle les SIG¹ (*Systèmes d'Information Géographiques*).

Les satellites offrent une vision synoptique de l'océan, donc on parle de résolution spatiale et temporelle assez importante. Ces deux critères jugés très importants pour les études océanographiques (qui se limitent à la couche de surface), sont impossibles d'obtenir à travers les moyens traditionnels, à savoir, les bateaux. En océanographie, l'utilisation de la télédétection a commencé avec les lancements faits par la NASA en 1978, des deux premiers satellites dédiés à l'observation de l'océan : **Seasat** et **Nimbus-7**. Ces satellites ont permis d'avoir pour la première fois, des mesures de quelques propriétés de la couche superficielle de l'océan, principalement la SST, la Chl-*a*, les courants marins, l'altimétrie, et les vents de surface [Barale *et al.*, 2010]. Depuis ce temps-là, le nombre de satellites lancés pour contribuer à l'observation de l'océan, a largement augmenté.

Dans ce chapitre, nous allons voir le fonctionnement principal des capteurs satellites et les principaux types de capteurs qui existent, on parlera aussi, brièvement, des principaux paramètres océanographiques, pouvant être observés via les satellites.

3.2 Principe de fonctionnement

Le principe général du fonctionnement de la télédétection, peut être résumé en 7 grandes étapes (**Figure 3.1**) :

1. Illumination : une source d'énergie est nécessaire pour illuminer la cible qu'on veut détecter. Comme illustré dans la **Figure 3.2**, le soleil représente la source principale de lumière en télédétection de type **passive**, pour laquelle les observations ne peuvent pas se faire pendant la nuit, ou quand il y a des couvertures nuageuses. En télédétection **active**, le capteur produit sa propre énergie (via des radars ou des lasers) qui est dirigée vers la cible. La nuit ainsi que les nuages ne présentent pas un obstacle pour ce type de télédétection.

1. « Un système informatique qui permet à partir de diverses sources, de rassembler, d'organiser, de gérer, d'analyser, de combiner, d'élaborer et de présenter des informations localisées géographiquement, contribuant notamment à la gestion de l'espace. » (Société Française de Photogrammétrie et de Télédétection, 1989)

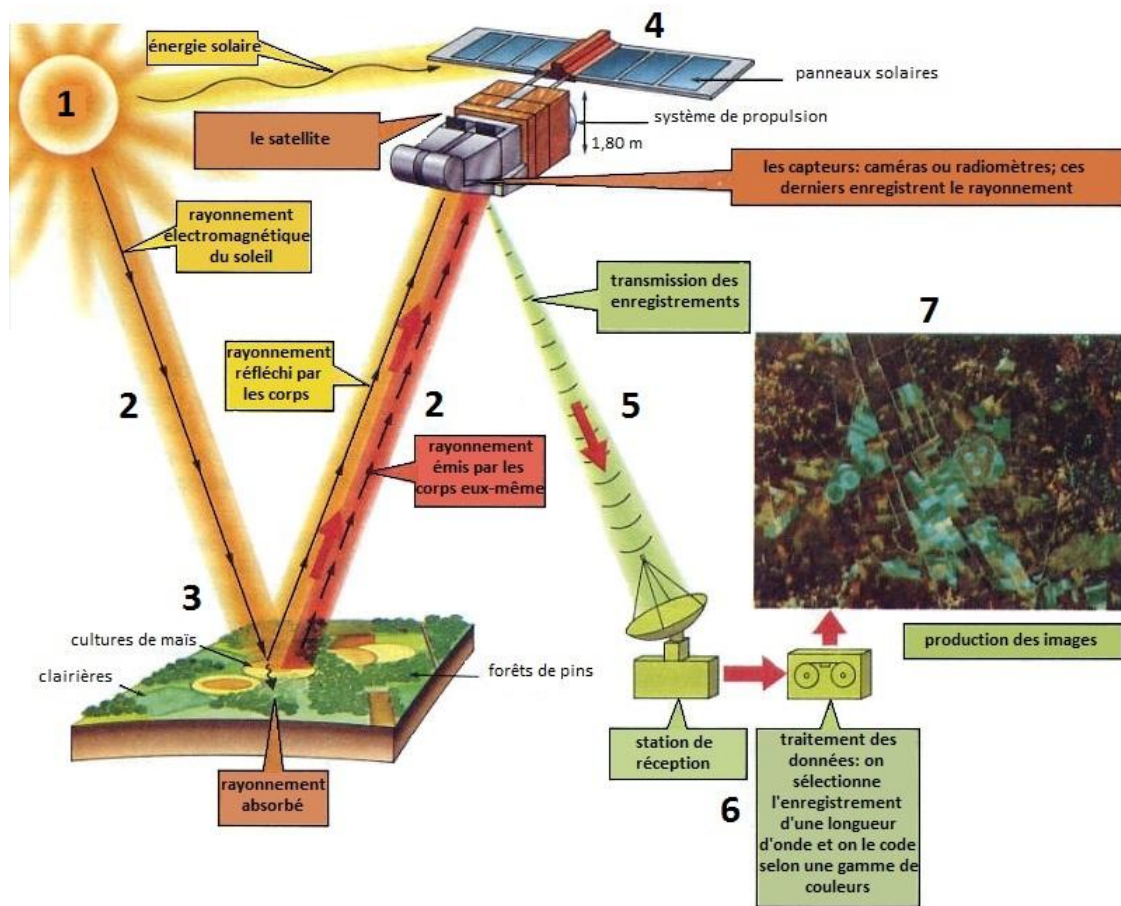


FIGURE 3.1 – Les principales étapes de la télédétection par satellite. 1 : Illumination, 2 : Interaction 1, 3 : Interaction 2, 4 : Enregistrement, 5 : Transmission, 6 : Analyse, 7 : Application. [Source : www.alertes-meteo.com]

- Interaction 1 : tout au long du parcours source-cible et cible-capteur, il y a une interaction entre le rayonnement (émis par la source) et l'atmosphère.
- Interaction 2 : l'énergie du rayonnement émis par la source, est *absorbée*, *transmise* ou *réfléchi*. Ceci dépend de la nature du rayonnement, et des caractéristiques de la surface de la cible.
- Enregistrement : le signal (énergie) émis (ou réfléchi) par la cible va être capté par le capteur (satellite).
- Transmission, réception et traitement : le signal ainsi reçu par le capteur, va être transmis à des stations au sol, décodé, et enfin transformé en images numériques ou photographiques.
- Analyse : afin d'obtenir des informations sur les caractéristiques de la cible captée, les images enregistrées sont analysées et interprétées (visuellement et/ou numériquement).
- Application : cette étape consiste en l'utilisation des informations extraites des images, pour des applications qui diffèrent d'un domaine à un autre.

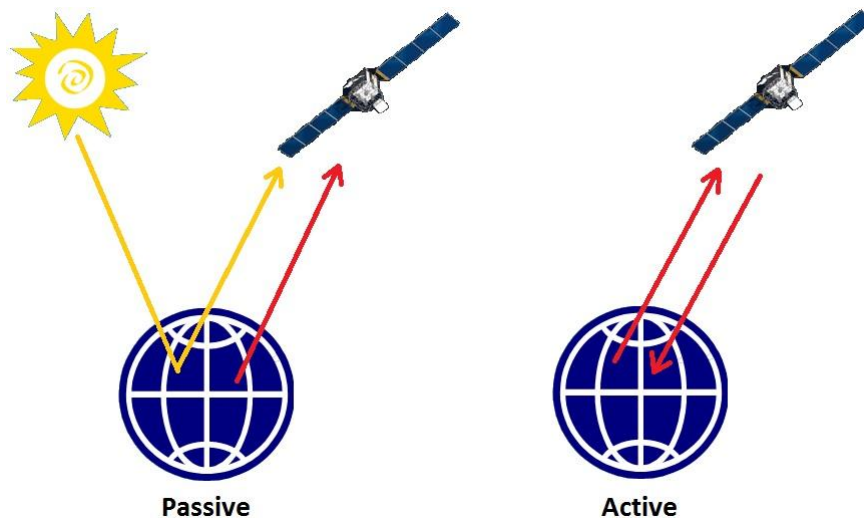


FIGURE 3.2 – Les deux modes de télédétection : passive et active.

Pour mieux comprendre ce mécanisme, on va détailler un peu plus les notions de bases :

3.2.1 Le rayonnement électromagnétique

Le rayonnement électromagnétique émis ou réfléchi par l'objet qu'on souhaite capter, contient des informations sur cet objet. Une fois ce rayonnement mesuré par le capteur, les informations vont être exprimées en images représentant cet objet. D'où l'appellation *radiomètres imageurs* pour désigner les capteurs de télédétection.

En physique, le rayonnement électromagnétique représente un phénomène ondulatoire symbolisant la propagation d'une quantité d'énergie dans la nature. Comme on le voit dans la **Figure 3.3**, une onde électromagnétique peut être définie comme étant une succession de vibrations, correspondant à deux champs, électrique et magnétique, perpendiculaires à la direction de propagation. Chaque onde électromagnétique est caractérisée par :

- La *longueur d'onde* λ : appelée aussi *amplitude*, représente la distance qui sépare deux crêtes successives. Elle est exprimée par le **mètre** ou ses multiples (principalement le **micromètre** [μm] et le **nanomètre** [nm]).
- La *fréquence* ν : exprime le nombre de vibrations effectuées par seconde. Elle est mesurée en **Hertz** (Hz) ou ses multiples (généralement en **kilohertz** [kHz], **mégahertz** [MHz] ou **gigahertz** [GHz]). Il existe une relation entre λ et ν , illustrée par l'équation :

$$\lambda \cdot \nu = c \quad (3.1)$$

c (pour célérité) représente la vitesse de propagation de l'onde électromagnétique, qui n'est d'autre que la vitesse de la lumière ($c = 299.792.458 \text{ m}\cdot\text{s}^{-1}$).

- L'*énergie* : le rayonnement électromagnétique transporte de l'énergie de façon discrète, sous forme de paquets d'énergie véhiculés par des entités élémentaires dites

photons. La quantité d'énergie (E) transportée est relative à la fréquence de l'onde, et est exprimée par l'équation suivante :

$$E = h \cdot \nu \quad (3.2)$$

h est la constante de Planck qui équivaut à $6,63 \cdot 10^{-34}$ J.s ($m^2.kg.s^{-1}$). Ainsi, les rayonnements de haute fréquence transportent plus d'énergie que les rayonnements de faible fréquence.

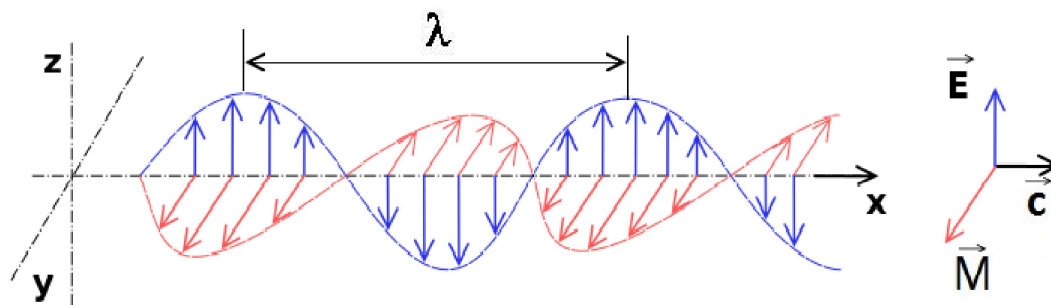


FIGURE 3.3 – Les composantes d'une onde électromagnétique. [Boutet, 2007]

La classification des ondes électromagnétiques en fonction de leurs longueurs d'onde, de leurs fréquences ou bien de leurs énergies, constitue ce qu'on appelle "*spectre électromagnétique*" (**Figure 3.4**).

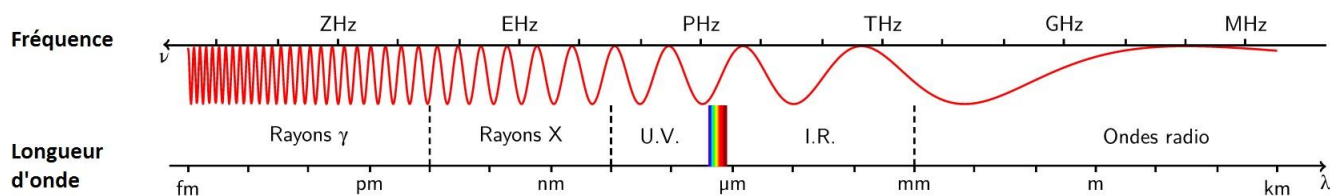


FIGURE 3.4 – Spectre électromagnétique : longueur d'onde et fréquence de chaque bande. [ABEL, 2012]

3.2.2 Spectre électromagnétique

Il existe plusieurs catégories d'ondes électromagnétiques utilisées en télédétection, dont une visible par l'être humain :

- **Les ondes hyperfréquences (ondes radio, micro-ondes)** : concernent les ondes ayant les plus basses fréquences, et transportant peu d'énergie. Elles sont utilisées pour la diffusion d'information (télévision, radio et téléphone). Les longueurs d'ondes de cette bande varient de 1 mm à 100.000 km.
- **L'infrarouge (IR)** : concerne le rayonnement qui est émis par tout corps ayant une température supérieur au zéro absolu (~ -273 °C). Quelques-unes des bandes spectrales de l'infrarouge sont utilisées en télédétection afin de mesurer la température des surfaces océaniques et terrestres. On distingue 4 types d'infrarouge :

- IR Proche (*near IR*) : 0,7 - 1,6 μm
- IR moyen (*middle IR*) : 1,6 - 4 μm
- IR thermique (*thermal IR*) : 4 - 15 μm
- IR lointain (*far IR*) : 15 - 1000 μm

- **La lumière visible** : représente une partie très limitée du spectre, elle est perceptible par l'œil humain. L'ensemble de couleurs qui constituent la lumière blanche (violet, bleu, vert, jaune, orange et rouge) est visible dans cette partie du spectre. Le rayonnement solaire est maximal dans cette bande.

Longueur d'onde pour chaque couleur :

- Violet : 0,400 - 0,446 μm
- Bleu : 0,446 - 0,500 μm
- Vert : 0,500 - 0,578 μm
- Jaune : 0,578 - 0,592 μm
- Orange : 0,592 - 0,620 μm
- Rouge : 0,620 - 0,700 μm

- **Les ultraviolets (UV)** : cette bande représente les plus petites longueurs d'ondes utilisées par les capteurs de télédétection. En raison de la forte absorption atmosphérique, cette zone est beaucoup plus utilisée pour les études atmosphériques. Elle se compose de 4 parties :

- UV extrême (*extreme UV*) : 0,01 - 0,1 μm
- UV lointain (*far UV*) : 0,1 - 0,2 μm
- UV moyen (*middle UV*) : 0,2 - 0,3 μm
- UV proche (*near UV*) : 0,3 - 0,4 μm

3.2.3 Signature spectrale

Une fois que le rayonnement traverse l'atmosphère et atteint la cible, en fonction de la longueur de l'onde et de la nature de la surface de la cible (matériau, humidité, ...), un de ces trois modes d'interactions (ou leur combinaison) peut survenir : réflexion, diffusion et absorption. En télédétection, on s'intéresse à l'énergie réfléchie par la cible. La proportion d'énergie réfléchie dans une direction donnée est appelée *réflectance* (voir **Figure 3.5**). Celle-ci varie en fonction de la longueur d'onde. Ainsi, en mesurant cette variation selon plusieurs longueurs d'ondes, la *signature spectrale* de l'objet peut être construite. La comparaison des signatures spectrales de plusieurs objets captés permet de les distinguer les uns des autres.

Citons comme exemples :

La végétation : la chlorophylle contenue dans les feuilles est une molécule qui absorbe fortement les rayonnements aux longueurs d'onde du rouge et du bleu, et elle réfléchit le vert. Dans l'eau, la chlorophylle des algues absorbe plus de bleu et réfléchit

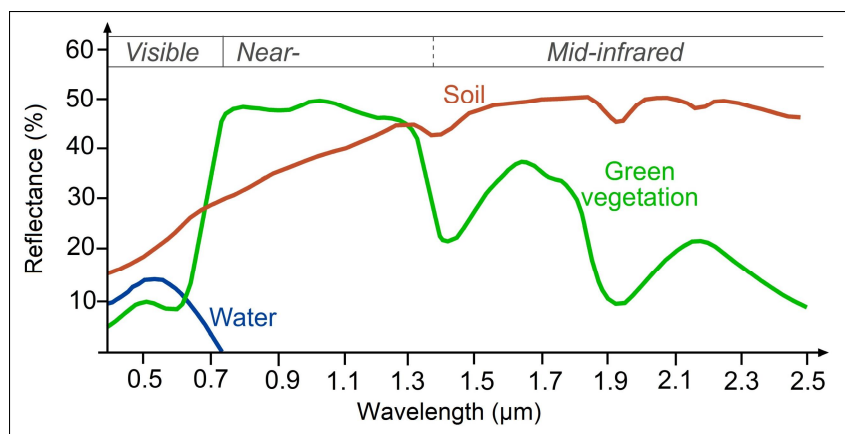


FIGURE 3.5 – Signature spectrale de l'eau, du sol, et de la végétation. [Source : www.seos-project.eu]

plus de vert. C'est pour cette raison que la végétation nous apparaît verte. Il est à noter aussi que les feuilles réfléchissent également dans le proche infrarouge (**Figure 3.6a**).

L'eau : absorbe les grandes longueurs d'onde du proche infrarouge et du visible, et réfléchit les petites longueurs d'ondes (bleu). C'est pourquoi la couleur de l'eau est bleu (dans le cas où elle contient des algues, elle apparaît bleu-verte ; **Figure 3.6b**).

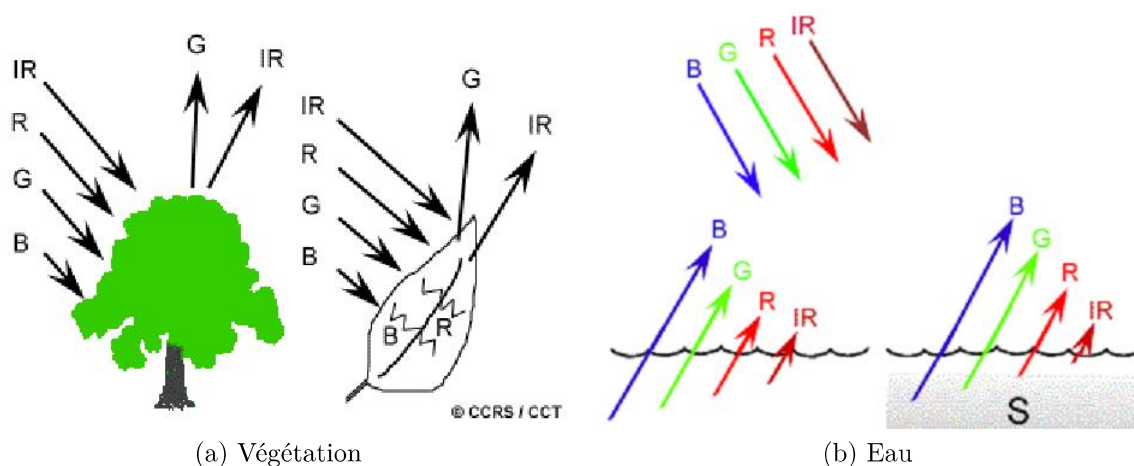


FIGURE 3.6 – Types de rayonnements absorbés et réfléchis par la végétation et par l'eau. S : surface de la mer [Centre canadien de cartographie et d'observation de la Terre, 2015]

3.3 Classification des capteurs de télédétection

Les capteurs des satellites peuvent être classés suivant deux critères, le premier est celui mentionné dans la **section 3.2** (passif et actif), et qui concerne le type de source de rayonnement utilisé. Le deuxième critère concerne la région du spectre utilisé pour capter la cible. Suivant ce critère, trois types de capteurs existent :

1. **Capteurs optiques** : appelés aussi *radiomètres imageurs*, ils fonctionnent dans les longueurs d'ondes se situant entre le visible et l'IR thermique. Ces capteurs produisent des images de la surface de la Terre en détectant les radiations solaires réfléchies par la cible. En fonction du nombre de bandes spectrales utilisées lors du processus de production d'images, ils peuvent être divisés en quatre types :

Panchromatique : le capteur utilise un seul canal correspondant à une seule bande spectrale. L'image résultante contient des informations sur la luminosité de l'objet détecté. Dans le cas où la bande spectrale correspond au visible, l'image résultante est une image en noir et blanc (Ex. : IKONOS PAN, SPOT HRV-PAN).

Multispectrale : le capteur utilise entre 3 et 10 canaux, chacun correspondant à une seule bande spectrale (étroite). L'image résultante contient des informations sur la luminosité et le spectre de l'objet observé (Ex. : LANDSAT MSS, SPOT HRV-XS).

Superspectrale : le capteur utilise entre 10 et 50 canaux pour des bandes spectrales très étroites, permettant d'avoir plus de précision sur les caractéristiques de l'objet observé (Ex. : MODIS, MERIS).

Hyperspectrale : le capteur utilise un nombre de capteurs qui va de 50 jusqu'à 300, pour des bandes spectrales plus étroites que celle du "superspectrale". Ainsi, les images produites contiennent encore plus d'informations sur les caractéristiques de l'objet détecté (Ex. : Hyperion).

2. **Capteurs infrarouges** : ces capteurs détectent l'énergie IR émise par la surface de la terre. Ils mesurent essentiellement la température de la surface (notamment la SST) et les propriétés thermiques de l'objet détecté (Ex. : AVHRR, MODIS).
3. **Capteurs à hyperfréquences (ou micro-ondes)** : ces capteurs ont la particularité de détecter les objets même dans de mauvaises conditions météorologiques (caractéristique des grandes longueurs d'onde). Néanmoins, comme les ondes de cette bande sont très longues, elles transportent moins d'énergie que les autres (**Section 3.2.1**). Ainsi, la résolution spatiale de ces capteurs est faible. Il en existe deux types :

Capteurs passifs à hyperfréquences : captent toute l'énergie émise par un objet dans les hyperfréquences.

Capteurs actifs à hyperfréquences : transmettent un signal micro-onde, et enregistrent le signal rétrodiffusé par la cible.

3.4 L'exploitation des données

Le signal électromagnétique reçu par le capteur est ensuite transmis sous forme numérique à une station de réception terrestre. Les données reçues sont alors traitées afin de corriger les distorsions atmosphériques et géométriques. Enfin, elles sont converties (suivant certains algorithmes) dans un format standard contenant des informations sur une ou plusieurs propriétés physiques de la surface observée.

L'exemple du *Ocean Biology Processing Group* (OBPG)

Ce groupe fait partie du *NASA's Goddard Space Flight Center* (<https://www.nasa.gov/centers/goddard/home/index.html>). Il a pour but de recueillir et de traiter les données provenant de plusieurs satellites d'observation de la Terre, afin de fournir aux utilisateurs des séries temporelles de paramètres (tels que la couleur de l'océan, SST, SSS, etc) de l'océan global, de façon continue, dans le but de faciliter la recherche et la mise en œuvre des applications qui ont trait aux changements climatiques, cycle du carbone, dynamiques des écosystèmes marins, etc. Les données fournies par ce groupe sont catégorisées suivant différents niveaux ; du niveau L0 (*Level 0*) qui représente les données sous format brut (et donc pas disponible pour tous les utilisateurs) jusqu'au niveau L4 (*Level 4*). Le niveau L3 représente les valeurs géophysiques des paramètres observés, organisées suivant une échelle spatiotemporelle uniforme. Le niveau L4 regroupe des données issues de modèles et d'analyses numériques. Les données utilisées dans notre étude sont de type L3 *Standard Mapped Image* (SMI).

Format L3 SMI

Ce format est une représentation sous forme d'un tableau de deux dimensions qui permet la visualisation du paramètre géophysique sous forme d'image. Les données sont disponibles sur plusieurs échelles de temps : un jour, huit jours, un mois, une saison, et une année. Chaque fichier SMI contient une seule image d'un seul paramètre, stockée dans un fichier HDF (*Hierarchical Data Format*). Ce dernier représente un format de fichier, développé par le *National Center for Supercomputing Applications* (NCSA, <http://www.ncsa.illinois.edu/>). Comme son nom l'indique, il est destiné à sauvegarder et à structurer de façon hiérarchique de très grandes quantités de données scientifiques. Le NCSA a développé une bibliothèque de routines et un ensemble de programmes et d'outils permettant la création et l'utilisation des fichiers HDF. Ce format est supporté par plusieurs langages informatiques : MATLAB, Java, Python, Le développement de cet outil est maintenant assuré par *The HDF Group* (THG, <http://www.hdfgroup.org/>).

Chaque fichier L3 SMI contient des informations qui peuvent être structurées en 5 sous-ensembles :

- **Mission et capteur** : on retrouve dans cette partie des informations sur la mission spatiale, le capteur qui fournit les données, la station terrestre qui les reçoit, et également sur le logiciel utilisé pour la création du produit ainsi que sur le processus de sa création.
- **Intervalle de temps** : les principales informations qu'on retrouve dans cette partie concerne l'intervalle de temps d'observation. Elle contient les dates (jour, mois, et année) et heures (heure, minute, seconde, et fraction de seconde) du début et de fin d'observation.
- **Coordonnées géographiques** : les informations géographiques (en latitude et longitude) sur toute la zone observée sont enregistrées ici, notamment la résolution spatiale du produit.
- **Description des données** : contient les informations sur le paramètre en question, ainsi que sur la taille du tableau.

- **Informations sur le "Scaling" :** cette partie concerne les données qui sont stockées sous un format réduit ('l3m_data' dans l'équation 3.3), et afin d'avoir la valeur géophysique du paramètre, il faut passer par une équation dite "Scaling equation" qui est sous la forme :

$$\text{Valeur} = (\text{Slope} \times \text{l3m_data}) + \text{Intercept} \quad (3.3)$$

On retrouve dans cette partie les valeurs du *Slope* et du *Intercept*, qui vont nous permettre de calculer les valeurs géophysiques du paramètre en question.

3.5 La télédétection en océanographie

Comme on l'a mentionné au début de ce chapitre, l'observation de l'océan par la voix satellitaire a commencé il y a déjà une quarantaine d'année. Les propriétés océanographiques pouvant être observées par satellite sont très nombreuses, notons comme exemples : Chl-*a*, SST, SSS, niveau de la mer, les courants marins, les vents de surface, Citons quelques uns de ces paramètres :

3.5.1 Couleur de l'océan

Cette propriété est obtenue en mesurant la réflectance de l'eau dans la partie visible du spectre, afin d'obtenir les concentrations du phytoplancton marin. En pratique c'est principalement celles de la Chlorophylle-*a* (Chl-*a*). Le premier capteur ayant fourni des données sur la couleur de l'océan fut le *Coastal Zone Color Scanner* (**CZCS** : 1978-1986). Le capteur qui a succédé au CZCS est le *Sea-viewing Wide Field-of-view Sensor* [**SeaWiFS** : 1997-2010; NASA Ocean Biology (OB), 2014c]. L'autre capteur qui est très utilisé par les chercheurs, est le *Moderate Resolution Imaging Spectroradiometer-Aqua* [**MODIS-A** : lancé en 2002; NASA Ocean Biology (OB), 2014b], qui offre des images avec deux résolutions spatiales : 4 km × 4 km, et 9 km × 9 km. Les données disponibles sont de juillet 2002 jusqu'à nos jours, avec une résolution temporelle qui va de 1 jour jusqu'à une année. **MODIS-A** comme **SeaWiFS**, offrent une précision de ±35% [Moore *et al.*, 2009]. La **Figure 3.7** présente un exemple d'une carte globale de Chl-*a*.

3.5.2 Température de surface (SST)

Les rayonnements électromagnétiques émis par l'océan se situent dans l'IR thermique et dans la bande des micro-ondes. La quantité de l'énergie reçue par le capteur, varie selon la température de la mer. La profondeur à laquelle la SST est détectée dépend du type du rayonnement émis; environ 20 μm pour l'IR thermique, et quelques millimètres pour les micro-ondes. Les premières observations de la SST par satellite furent en 1981, grâce au lancement du *Advanced Very High Resolution Radiometer* (**AVHRR**). Depuis, d'autres satellites en orbite permettent la détection de ce paramètre, comme le **MODIS-A** [NASA Ocean Biology (OB), 2014b] qui offre une précision d'environ ±0,5 °C, avec les mêmes résolutions spatio-temporelles que celles de la Chl-*a* (**Section 3.5.1**). La **Figure 3.8** montre une carte mensuelle de SST produite par MODIS-A.

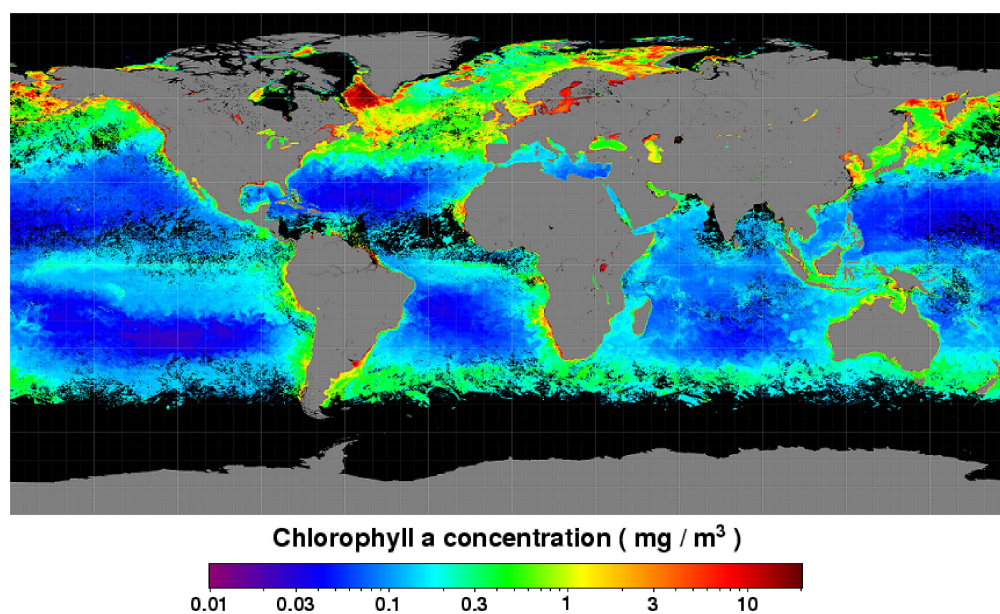


FIGURE 3.7 – Moyenne mensuelle de Chl-*a* de Mai 2015 fournie par MODIS Aqua avec une résolution de 4 km × 4 km. [Source : [NASA Ocean Biology (OB), 2014c]]

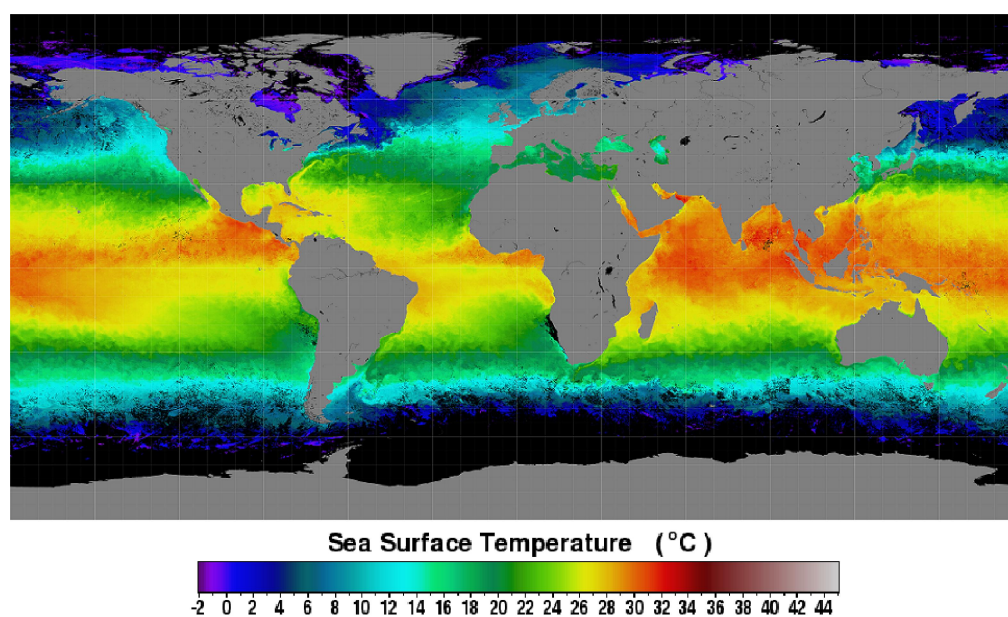


FIGURE 3.8 – Moyenne mensuelle de SST de Mai 2015 fournie par MODIS Aqua avec une résolution de 4 km × 4 km. [Source : [NASA Ocean Biology (OB), 2014b]]

3.5.3 Salinité de surface (SSS)

Contrairement à la SST qui est mesurée par satellite depuis plus de trois décennies, les technologies permettant l'observation de la SSS depuis l'espace sont encore dans leurs débuts. La salinité de la surface des océans est mesurée grâce à des capteurs micro-ondes opérant autour de 1,4 GHz. L'observation est faite jusqu'à une profondeur d'environ 1 à 2 cm [Barale *et al.*, 2010]. Le premier satellite d'observation de SSS est le *Soil Mois-*

ture and Ocean Salinity (SMOS), qui a été lancé en 2009 par l'European Space Agency (ESA, <http://www.esa.int/>). Les données (mensuelles) provenant de ce satellite ont une précision d'environ 0,3-0,4 au niveau des zones tropicales et subtropicales, et de 0,5 dans les régions froides [Boutin *et al.*, 2013; Kolodziejczyk *et al.*, 2016], avec une résolution spatiale qui varie de 25 à 50 km, et une résolution temporelle d'un minimum de 3 jours à l'équateur [ESA, 2013]. En 2011, un deuxième satellite a été lancé par la NASA ; **Aquarius** [NASA Ocean Biology (OB), 2014a], ce dernier offre une résolution spatiale d'environ 150 km × 150 km, sur une échelle temporelle d'un minimum de 1 jour. La précision de SSS sur tout le globe, et moyennée sur un mois, est de 0,2 [NASA Ocean Biology (OB), 2014a]. La **Figure 3.9** présente une carte mensuelle de SSS produite par Aquarius.

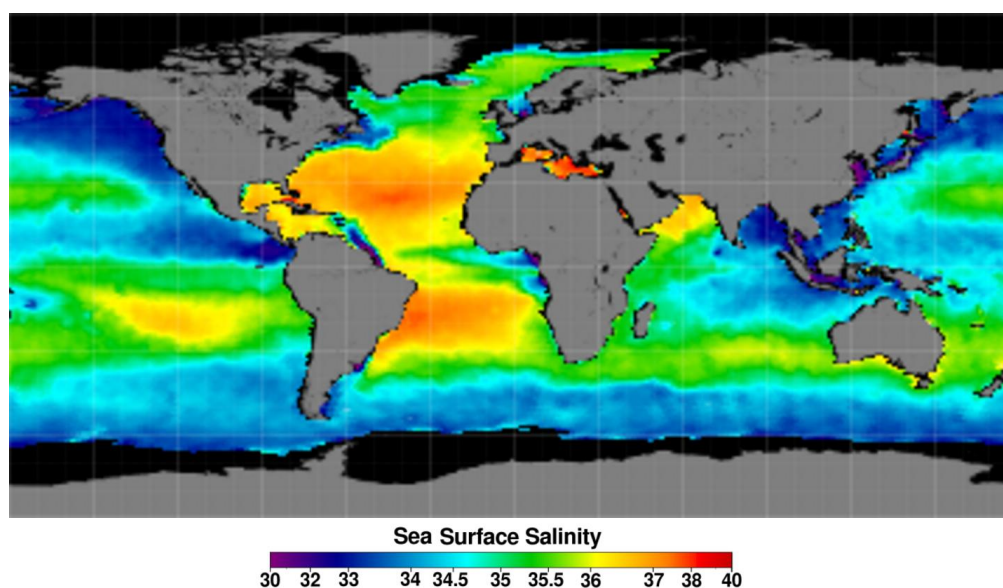


FIGURE 3.9 – Moyenne mensuelle de SSS de Mai 2015 fournie par Aquarius avec une résolution spatiale de 1 degré ($\sim 150 \text{ km} \times 150 \text{ km}$). [Source : [NASA Ocean Biology (OB), 2014a]]

Les données fournies pour les paramètres qui, de nos jours, peuvent être observés par satellite, sont basées sur la compréhension qu'on a sur le fonctionnement de tel ou tel processus naturel, et aussi sur les techniques mises en place pour permettre une acquisition d'informations avec de bonnes précisions, et de bonnes résolutions spatio-temporelles. Néanmoins, il existe des processus naturels qui ne transmettent pas de signal électromagnétique, et donc qui ne peuvent pas être observés par satellite. C'est le cas de la fugacité du CO_2 dans l'eau de surface des océans.

Plusieurs méthodes et modèles ont été développés afin d'interpoler la $f\text{CO}_2$ de surface, pour mieux comprendre ses variabilités, et ainsi, localiser les puits et sources de CO_2 dans l'océan, mais aussi pour avoir plus d'informations sur le flux air-mer de CO_2 .

Chapitre 4

MÉTHODES

La zone de l'océan Atlantique tropical a été largement étudiée par une grande communauté scientifique. En effet, l'interpolation de la fugacité du CO_2 dans l'eau de surface de cet océan, a fait l'objet de diverses études, élaborées par de nombreux chercheurs. Le choix de la méthode à adopter, et des paramètres d'estimation de $f\text{CO}_2$ à utiliser, est une question qui doit être posée avant de commencer cette étude. Ceci dans le but d'élaborer un modèle très efficace, c.à.d., ayant une grande précision dans l'estimation de la $f\text{CO}_2$.

Dans la première partie de ce chapitre, j'aborderai les études les plus pertinentes qui ont traité l'interpolation de $f\text{CO}_2$ dans plusieurs zones de l'océan global, en s'intéressant d'avantage à l'océan Atlantique. Cela va permettre d'élucider la question du choix de la méthode, et des paramètres d'estimation.

Dans la seconde partie du chapitre, j'expliquerai en détail la méthode choisie pour l'interpolation de la $f\text{CO}_2$.

4.1 Exemples d'études traitant l'interpolation de la $f\text{CO}_2$

Les études présentées dans cette section, n'ont pas forcément les mêmes objectifs. Ainsi, pour chaque étude, l'objectif principal ainsi que le(s) résultat(s) obtenu(s) sont brièvement expliqués.

1. Lefèvre *et al.* [2005] : "*A comparison of multiple regression and neural network techniques for mapping in situ $p\text{CO}_2$ data*"

- **Objectif(s) de l'étude** : comparaison entre la régression linéaire multiple (MLR ; Multiple Linear Regression) et les cartes auto-organisatrices (SOM ; Self Organizing Map, Kohonen [1982]), pour cartographier la $p\text{CO}_2$ de surface mensuelle, sur une échelle spatiale de $1^\circ \times 1^\circ$. La BD *in situ* utilisée est issue du projet CAVASSOO (Carbon Variability Studies by Ships of Opportunity ; tracer.env.uea.ac.uk/e072), et contenant un nombre de points de 104 000. La zone d'étude a été divisée en 3 régions biogéochimiques. Pour chaque région, un modèle a été développé.

- **Année(s)** : de 1995 à 1997.

- **Zone géographique** : gyre subpolaire de l'Atlantique Nord (50-70°N, 60-10°O).

- **Paramètre(s) utilisé(s)** : latitude, longitude, SST, et l'année.

- **Résultat(s) obtenu(s)** :

- * Les cartes produites par l'approche SOM, offrent plus de détail à petite échelle que celles produites par MLR.

- * La performance a été estimée en calculant l'écart-type résiduel (RSD)¹ entre le résultat souhaité (X_i) et le résultat obtenu Y_i . Lors de l'apprentissage du SOM, le RSD variait de 4 à 16 μatm (en fonction des mois).

- * Lors de la comparaison des deux méthodes (sur un plus petit set de données), SOM a donné de meilleurs précisions que MLR (RSD entre 3 et 11 μatm pour SOM vs. RSD entre 4 et 88 μatm pour MLR).

- **Problème(s) soulevé(s) et/ou suggestion(s)** : comme la zone a été divisée en plusieurs régions, contrairement à SOM, pour les cartes produites par MLR, la transition de pCO_2 entre ces régions là n'est pas graduelle ; le changement de pCO_2 est brusque d'une région à une autre (**Fig. 6** dans Lefèvre *et al.* [2005]).

2. Friedrich et Oschlies [2009] : "*Neural network-based estimates of North Atlantic surface pCO_2 from satellite data : A methodological study*"

- **Objectif(s) de l'étude** : estimation de pCO_2 mensuelle à partir de données satellitaires, en utilisant l'approche SOM. Les 740 000 données *in situ* utilisées ont une profondeur qui va de 10 à 250 m, et sont issues du projet Européen CARBOOCEAN (www.carboocean.org/). La résolution spatiale est de $0,5^\circ \times 0,5^\circ$.

- **Zone géographique** : océan Atlantique, entre 18°S et 70°N.

- **Année(s)** : 2005.

- **Paramètre(s) utilisé(s)** : SST et Chl-*a*.

- **Résultat(s) obtenu(s)** :

- * Un RMSE global de 19,0 μatm a été obtenu, lors de la construction du modèle.

- * Un RMSE de 21,1 μatm a été obtenu lors de l'interpolation de pCO_2 en utilisant des valeurs de climatologie pour la SST et la Chl-*a*, afin de remplacer les points satellitaires qui manquent (à cause des conditions météorologiques).

- **Problème(s) soulevé(s) et/ou suggestion(s)** : la couverture réduite des données *in situ*, le manque fréquent de points de données satellitaires, ainsi que la forte variabilité spatiale de pCO_2 dans cette région, sont les plus importants facteurs limitant les performances du modèle.

3. Telszewski *et al.* [2009] : "*Estimating the monthly pCO_2 distribution in the North Atlantic using a self-organizing neural network*"

- **Objectif(s) de l'étude** : cartographier la pCO_2 sur une échelle mensuelle, en utilisant la méthode SOM, avec une résolution spatiale de $1^\circ \times 1^\circ$. Les 137 000 données utilisées sont issues du projet CARBOOCEAN (www.carboocean.org/).

1. RSD (Residual Standard Deviation) = $\sqrt{\frac{\sum_i (Y_i - X_i)^2}{N-2}}$, N est le nombre de points.

- **Zone géographique** : océan Nord Atlantique ; de $10,5^\circ N$ - $75,5^\circ O$ à $75,5^\circ N$ - $9,5^\circ E$.

- **Année(s)** : de 2004 à 2006.

- **Paramètre(s) utilisé(s)** : SST, MLD (Mixed layer depth ; la profondeur de la couche de mélange océanique), et Chl-*a* satellitaire.

- **Résultat(s) obtenu(s)** :

- * Pour tout l'ensemble de données, un RMSE global de $11,6 \mu atm$ a été obtenu lors de l'étape de construction du modèle.

- * En comparant avec des données indépendantes, un RMSE allant de $6,3 \mu atm$ à $19,3 \mu atm$ a été obtenu (la valeur change en fonction des cinq sous-régions biogéochimiques de la zone d'étude).

- **Problème(s) soulevé(s) et/ou suggestion(s)** :

- * Le manque de données satellitaires de Chl-*a*, limite les performance du modèle.

- * Le modèle à tendance à lisser (*smooth*) les valeurs maximales et minimales, dans la zone Est des subtropiques.

- * L'introduction des données de SSS est suggérée pour améliorer les résultats du modèle.

4. Landschützer *et al.* [2013] : "*A neural network-based estimate of the seasonal to inter-annual variability of the Atlantic Ocean carbon sink*"

- **Objectif(s) de l'étude** : création de cartes mensuelles de pCO_2 de surface, avec une résolution spatiale de $1^\circ \times 1^\circ$. L'approche SOM est utilisée pour partitionner la zone d'étude en 16 régions biogéochimiques, et l'approche FNN (Feedforward Neural Network) pour trouver le modèle d'interpolation de fCO_2 pour chaque région. À partir de ces cartes de fCO_2 , des cartes de flux air-mer de CO_2 ont été construites. Les 20 003 données *in situ* utilisées sont issues de la BD SOCAT v.1.5 [Sabine *et al.*, 2013].

- **Zone géographique** : océan Atlantique entre $44^\circ S$ et $79^\circ N$ (à l'Ouest de $30^\circ E$).

- **Année(s)** : de 1998 à 2007.

- **Paramètre(s) utilisé(s)** : SST, Chl-*a*, MLD, SSS, et $xCO_{2,atm}$ (fraction molaire du CO_2 atmosphérique).

- **Résultat(s) obtenu(s)** :

- * Un RMSE de $10 \mu atm$ a été obtenu lors de l'étape de construction des modèles.

- * Lors d'une comparaison avec des données indépendantes à celle de la construction des modèles, un RMSE de $22,8 \mu atm$ a été obtenu pour tout l'océan Atlantique, et un RMSE de $16,3 \mu atm$ pour les points situées au Sud de $40^\circ N$.

- **Problème(s) soulevé(s) et/ou suggestion(s)** :

- * La Chl-*a* ne peut être négligée lorsque l'on étudie la fCO_2 sur une longue couverture temporelle.

- * Plus de données *in situ* sont nécessaires pour améliorer la performance du modèle.

5. Nakaoka *et al.* [2013] : "*Estimating temporal and spatial variation of ocean surface pCO₂ in the North Pacific using a self-organizing map neural network technique*"

- **Objectif(s) de l'étude** : Production de cartes mensuelles de pCO₂ en utilisant l'approche SOM, avec une résolution spatiales est de 0,25° × 0,25°. Les 73 284 données *in situ* utilisées, sont fournies par le programme NIES VOS (National Institute of Environmental Studies, Volunteer-Observing-Ship; <http://soop.jp/>).

- **Zone géographique** : océan Nord Pacifique ; entre 10°N-90°O et 60°N-120°E.

- **Année(s)** : entre 2002 et 2008.

- **Paramètre(s) utilisé(s)** : SST, MLD, Chl-*a*, et SSS.

- **Résultat(s) obtenu(s)** :

* Un RMSE de 17,6 μatm a été obtenu lors de la construction du modèle.

* Un RMSE de 20,2 μatm a été obtenu lors de la comparaison avec des données indépendantes de celles utilisées pour construire le modèle.

- **Problème(s) soulevé(s) et/ou suggestion(s)** :

* La SSS améliore la performance du modèle quant à l'estimation de pCO₂.

6. Zeng *et al.* [2014] : "*A Global Surface Ocean fCO₂ Climatology Based on a Feed-Forward Neural Network*"

- **Objectif(s) de l'étude** : utilisation de l'approche FNN pour la création de climatologie mensuelle de fCO₂ de surface, avec une résolution spatiale de 1° × 1°. Ceci à fin de calculer le flux air-mer de CO₂. Les 127 880 données *in situ* utilisées, sont issues de la BD SOCAT v.2 [Sabine *et al.*, 2013].

- **Zone géographique** : l'océan global.

- **Année(s)** : 2000.

- **Paramètre(s) utilisé(s)** : Mois, Latitude, Longitude, SST, SSS, et Chl-*a*.

- **Résultat(s) obtenu(s)** :

* Un RSD global de 15,8 μatm a été obtenu lors de la construction du modèle.

* Un RSD de 16 μatm a été obtenu pour des données indépendantes à celles utilisées lors de l'étape de construction du modèle.

Les résultats de ces études s'avèrent très importants, et doivent être pris en considération pour décider de la démarche à suivre.

La première remarque qu'on puisse faire, est que sur les six études présentées ci-dessus, dans quatre d'entre elles, l'approche SOM a été adoptée pour l'interpolation et la cartographie de fCO₂ (ou pCO₂). Les deux autres études utilisent l'approche FNN. En effet, la méthode SOM est, après la MLR, la méthode la plus utilisée pour l'interpolation de fCO₂. Bien que celle-ci est à la base, une méthode qui est utilisée pour résoudre les problèmes tels que ; la quantification vectorielle², la robotique, la classification . . . , son utilisation pour l'approximation de fonction s'est beaucoup répandue ces 15 dernières années [Barreto et Araújo, 2002; Barreto, 2007]. Néanmoins, bien que FNN n'a pas été beaucoup utilisée pour l'interpolation de fCO₂, elle reste tout de même l'approche la plus répandue est la plus efficace pour les problèmes de régressions non linéaires [Samarasinghe,

2. Une technique de compression de données.

2006]. Ainsi, j'ai choisi d'utiliser la méthode FNN, ceci afin d'explorer les avantages de cette méthodes, et essayer de développer un modèle d'interpolation de $f\text{CO}_2$, qui soit performant et simple d'utilisation, pour enfin pouvoir construire des cartes de $f\text{CO}_2$ de bonne qualité, à partir de données satellitaires.

La deuxième remarque concerne le choix des paramètres en entrée du modèle, utilisés pour l'estimation de $f\text{CO}_2$. Vu l'importance de la SST et de la Chl-*a*, et leurs grandes influences sur la $f\text{CO}_2$, on peut constater que celles-ci se retrouvent comme entrées du modèle, dans cinq sur six des études présentées. La SSS quant-à-elle, est utilisée dans les trois dernières études (les plus récentes), mais son importance pour l'estimation de $f\text{CO}_2$ est souvent évoquée dans les études [Telszewski *et al.*, 2009; Nakaoka *et al.*, 2013].

On remarque aussi l'utilisation de la MLD dans trois de ces études. Selon Zeng *et al.* [2014], celle-ci a le moins d'effet sur la $f\text{CO}_2$ et, en outre, les données de MLD sont plus rares comparées aux autres paramètres. Ainsi, l'introduction de ce paramètre à la BD de construction du modèle, entrainera une forte diminution de la taille de cette base, et donc, une perte d'informations sur la $f\text{CO}_2$. Le meilleur choix serait donc d'exclure ce paramètre, et de combiner les données de SST, de Chl-*a*, et de SSS pour construire le modèle d'interpolation de $f\text{CO}_2$.

4.2 Les réseaux de neurones artificiels

Cette approche fait partie du domaine de l'intelligence artificielle, qui représente tous les aspects où on essaye d'imiter les processus biologiques qui émanent de la nature, afin de rendre les machines intelligentes. Le cerveau, le système génétique, le système immunitaire, ou même le comportement de certains animaux ou insectes, ont été une source d'inspiration pour plusieurs chercheurs, qui ont su observé et analysé ces mécanismes, pour ensuite développé des algorithmes performants, qui reproduisent le fonctionnement de ces mécanismes.

L'histoire du développement des réseaux de neurones (RN) artificiels, a été documentée par plusieurs auteurs, on peut cité comme exemple; le livre *Neurocomputing : Foundations of Research*, édité en 1988 par Anderson et Rosenfeld [Anderson et Rosenfeld, 1988]. Ce livre contient un ensemble de 43 des études les plus pertinentes, et qui ont révolutionné le monde des RNs artificiels.

Inspirés du fonctionnement du système nerveux, un RN artificiel consiste en un ensemble de neurones formels (ou unités de calcul) interconnectés sous forme de "*couche*" (voir **Figure 4.1**), et visant à calculer une sortie à partir d'une ou plusieurs entrées, en passant par ce qu'on appelle "*apprentissage*".

Le neurone formel (voir **Section 4.2.1.1**) représente l'élément de base dans l'élaboration d'un RN. Il a été inspiré du modèle proposé par les deux physiciens Warren McCulloch et Walter Pitts en 1943, qui ont montré qu'un neurone formel, doit pouvoir calculer n'importe qu'elle fonction arithmétique ou logique, et ont inventé le neurone dit "*binnaire*", qui, en appliquant une fonction à seuil à la somme pondérée de ses entrées, donne un résultat de 0 ou de 1 [McCulloch et Pitts, 1943].

Dans les années 1950s, la recherche dans ce domaine s'est rapidement développée, et plusieurs types de RN ont vu le jour. On peut citer :

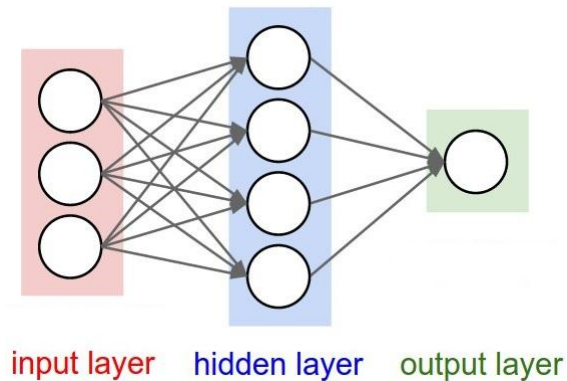


FIGURE 4.1 – Illustration d'un réseau de neurones avec : 1 couche d'entrée (input layer), 1 couche cachée (hidden layer), et 1 couche de sortie (output layer).

- Le psychologue Donald Hebb, a introduit en 1949 la notion d'*apprentissage* dans les neurones biologiques [Hebb, 1949].
- La première application pratique des RNs artificiels, a été initiée en 1958 par Frank Rosenblatt qui a inventé le type dit "*perceptron*" et sa règle d'apprentissage. Il a appliqué ce type pour la reconnaissance des formes [Rosenblatt, 1958].
- En 1960, vient l'invention de l'ADALINE (Adaptive Linear neural network) par Bernard Widrow et Marcian Hoff [Widrow et Hoff, 1960; Widrow, 1960] qui ont introduit l'apprentissage par l'algorithme du gradient, appliqué sur un RN linéaire d'une seule couche, dans lequel la performance du RN est estimée par l'"*erreur quadratique moyenne*" (MSE; *Mean Square Error*). Ceci fut l'ancêtre de l'algorithme de "*rétropropagation*" (voir **Section 4.2.2.1**).
- En 1969, Marvin Minsky et Seymour Papert ont publié un livre intitulé "*Perceptrons : an introduction to computational geometry*" [Minsky et Papert, 1969], dans lequel ils ont présenté les limites des RNs proposés par Rosenblatt [1958] et Widrow et Hoff [1960].
L'influence du livre de Minsky et Papert [1969], et les limitations-en terme de puissance de calcul-des ordinateurs durant cette époque, ont fait que plusieurs chercheurs ne croyait plus en les performances des RNs, et ont quitté cette discipline. Seulement quelques importantes études ont continué durant les années 1970, notamment, Kohonen [1972] qui avait initié la création du RN de type SOM [Kohonen, 1982].
- Ce n'est qu'à partir des années 1980, et avec l'avènement de nouvelles technologies qui ont rendu les ordinateurs plus puissants, que de nouveaux concepts et idées de RN ont commencé à apparaître. En effet, on doit ce renouveau d'intérêt au physicien John Hopfield, qui a introduit le modèle récurrent. Ce dernier consiste en un RN où les neurones sont complètement connectés [Hopfield, 1982].
- L'autre grande découverte des années 1980, était l'algorithme dit de "*rétropropagation*" (*Backpropagation*, voir **Section 4.2.2.1**), pour l'apprentissage du "*perceptron multi-couches*" (MLP; *Multi Layer Perceptron*, voir **Section 4.2.2**). On doit cette invention à David Rumelhart et James McClelland [Rumelhart et McClelland, 1986].

À partir de cette décennie, plusieurs chercheurs ont manifesté un intérêt grandissant pour les RNs. Ainsi, des centaines de travaux scientifiques ont été réalisés, traitant différents problèmes dans différents domaines de recherche, et ceci continu jusqu'à aujourd'hui.

Les application des RNs aux problèmes qui concernent l'environnement ont commencé à prendre de l'ampleur dès les années 2000. Les RNs ont été utilisés pour les études atmosphériques [ex. : Cavazos, 2000; Niang *et al.*, 2006, ...], météorologiques [ex. : Ali *et al.*, 2007, ...], et océanographiques [ex. : Richardson *et al.*, 2003; Lefèvre *et al.*, 2005; Liu *et al.*, 2006; Reusch *et al.*, 2007; Friedrich et Oschlies, 2009; Telszewski *et al.*, 2009; Landschützer *et al.*, 2013; Nakaoka *et al.*, 2013; Zeng *et al.*, 2014, ...].

4.2.1 Principe de fonctionnement

Le fonctionnement du RN repose sur l'élément de base qui le constitue, à savoir ; le neurone formel.

4.2.1.1 Neurone formel

Comme illustré dans la **Figure 4.2**, les composants d'un neurone formel (ou artificiel) sont les suivants :

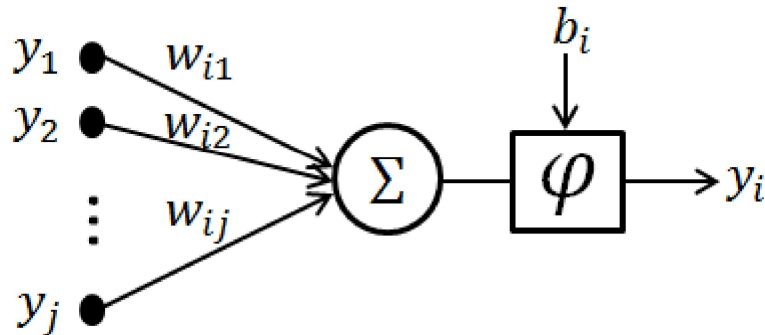


FIGURE 4.2 – Représentation basique d'un neurone artificiel.

- Les entrées y_j : représentent les sorties de neurones antérieurs, ou bien les entrées du RN.
- Les poids synaptiques w_{ij} : permettent la transmission de l'information provenant de l'entrée y_j au neurone i .
- Sommation de la jonction Σ ou net_i : à ce niveau là, la somme pondérée des entrées est effectuée suivant l'équation :

$$net_i = \sum_j w_{ij} y_j \quad (4.1)$$

- La fonction de transfert φ : appelée aussi *fonction d'activation*, elle permet de définir la sortie y_i du neurone, et peut être linéaire ou non linéaire.

- *Le biais b_i* : appliqué à la fonction de transfert φ , permet au RN d'être plus performant. Son intérêt peut être constaté en comparant un neurone biaisé avec un autre non biaisé, en effet, la sortie du neurone non biaisé sera toujours égale à 0 dans le cas où les entrées sont égales à 0.
- *La sortie y_i* : elle est calculée comme suit :

$$y_i = \varphi(\text{net}_i + b_i) \quad (4.2)$$

Un neurone artificiel peut donc être considéré comme une fonction mathématique, qui a pour variables les entrées de ce neurone ($y_1 \dots y_j$). La valeur numérique de cette fonction (y_i) dépend des poids synaptiques du neurone ($w_{i1} \dots w_{ij}$) ainsi que de la fonction d'activation φ . Cette dernière peut être choisie suivant le problème qu'on veut résoudre. Plusieurs formes possibles de cette fonction existent. Les trois fonctions les plus utilisées sont présentées dans le **Tableau 4.1**.





Fonction	Forme	Graphe
<i>Hard limit</i>	$\varphi(\text{net}_i + b_i) = \begin{cases} 0 & \text{si } (\text{net}_i + b_i) < 0 \\ 1 & \text{sinon.} \end{cases}$	
<i>Linear</i>	$\varphi(\text{net}_i + b_i) = \text{net}_i + b_i$	
<i>Log-sigmoid</i>	$\varphi(\text{net}_i + b_i) = \frac{1}{1 + e^{-(\text{net}_i + b_i)}}$	
<i>Tan-sigmoid*</i>	$\varphi(\text{net}_i + b_i) = \frac{2}{1 + e^{-2 \times (\text{net}_i + b_i)}} - 1$	

TABLEAU 4.1 – Présentation des trois fonctions d'activations les plus utilisées. **tangente-sigmoïde*

4.2.1.2 Caractéristiques d'un RN artificiel

Tout RN artificiel est défini par deux critères principaux :

- Son architecture : concerne principalement le nombre de couches que contient le RN, ainsi que le type de connexions entre les neurones qui décrit sa topologie. On peut distinguer : des connexions directes (non bouclées ; "*feedforward*"), et des connexions récurrentes (bouclées ; "*feedback*") (voir **Figure 4.3**).
- Son algorithme d'apprentissage : c'est la propriété la plus intéressante des RNs. Il consiste en plusieurs étapes qui se déroulent en plusieurs itérations, tout en modifiant certaines propriétés du RN, jusqu'à trouver la meilleure solution. Les deux types d'apprentissage les plus utilisés sont :

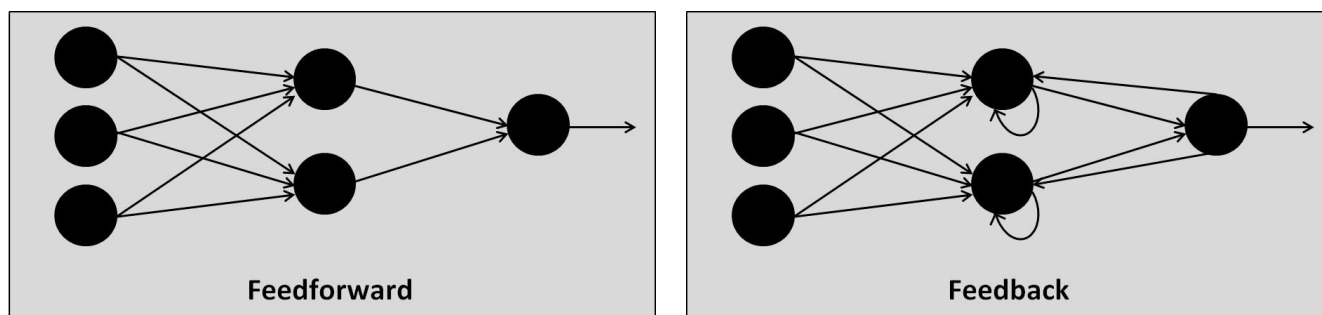


FIGURE 4.3 – Deux types d’architecture de réseau de neurones.

* Apprentissage supervisé : ce type d’apprentissage consiste à modifier le RN en comparant le résultat calculé avec le résultat désiré, jusqu’à trouver la bonne sortie.

* Apprentissage non supervisé : ce type est beaucoup plus appliqué pour la classification. Il consiste à modifier le RN en se basant seulement sur les régularités statistiques des entrées, pour ensuite établir deux ou plusieurs catégories de ces entrées.

4.2.2 Le perceptron multi-couches : MLP

Le MLP (*Multilayer perceptron*) est sans doute l’architecture la plus connue et la plus utilisée des RNs. Il consiste en un réseau contenant des neurones organisés en couches ; une couche d’entrée, une ou plusieurs couches cachées, et une couche de sortie (**Figure 4.4**). Les fonctions de transfert attribuées à ces couches, peuvent être différentes les unes des autres. Dans un MLP, chaque neurone est complètement connecté aux neurones de la couche suivante, et les neurones d’une couche deviennent les entrées de la couche suivante (**Figure 4.4**). L’information dans un MLP circule en un seul sens ; de la couche d’entrée vers la couche de sortie (*feedforward* ; **Figure 4.3**).

L’apprentissage dans un MLP est contrôlé par l’algorithme supervisé cité auparavant ; rétropropagation (*backpropagation*).

4.2.2.1 *Backpropagation* : Algorithme de rétropropagation du gradient

Dans une première étape un RN est dit "ignorant", c.à.d. tant que l’apprentissage n’a pas commencé, le RN est dépourvu de toute expérience : si on lui présente n’importe quel couple (\vec{x}, t) , où x représente un vecteur d’entrée au RN, et t la sortie désirée, la sortie qu’il donnera (y_i) sera très différente de la sortie désirée (t). Ceci est dû au fait que les poids initiaux (w_{ij}) ne sont pas encore adaptés aux couples (\vec{x}, t) .

Le terme rétropropagation, peut être expliqué comme suit :

- Passe avant : c’est l’étape où les données d’apprentissage sont introduites au RN ; c’est la **propagation**.
- Passe arrière : c’est l’étape où l’algorithme se charge de modifier les poids des connexions (w_{ij}), tout en essayant de minimiser l’erreur MSE ; c’est la **rétropro-**

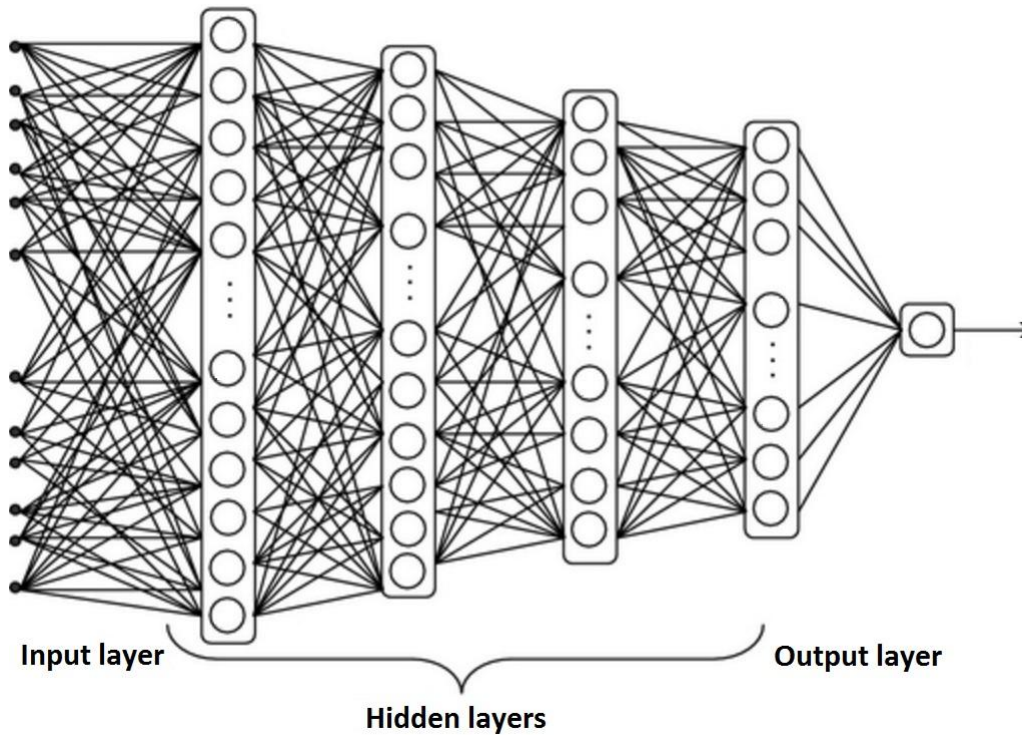


FIGURE 4.4 – Architecture d'un RN de type MLP. [Bohlouli *et al.*, 2012]

propagation.

Les notations suivantes vont être utilisées dans ce qui suit :

- C : l'ensemble des neurones de sortie.
- A_i : l'ensemble des neurones j antérieurs au neurone i .
- E : la MSE de tout le RN :

$$E = \frac{1}{2} \sum_j (e_o)^2 \quad (4.3)$$

où e_o est l'erreur du neurone en sortie (o) ; $e_o = t_o - y_o$.

1. **Propagation** : les sorties des neurones de la couche d'entrée, sont déterminées par les entrées du RN. Pour tous les autres neurones, les sorties sont calculées en fonction des sorties des neurones antérieurs ($\in A_i$), ainsi :

$$y_i = \varphi_i \left(\sum_{j \in A_i} (net_i + b_i) \right) \quad (4.4)$$

2. **Rétropropagation de l'erreur** : L'adaptation des poids se fait sur tous les neurones, en commençant par la couche de sortie.

Couche de sortie : le gradient du poids est calculé ainsi :

$$\Delta w_{ij} = -\eta \frac{\partial E}{\partial w_{ij}} \quad (4.5)$$

où η représente le taux d'apprentissage.

$$\Delta w_{ij} = \eta \delta_i y_j \quad (4.6)$$

avec

$$\delta_i = e_i y_i [1 - y_i] \quad (4.7)$$

qui correspond au gradient du neurone i

Couche cachée : le gradient du poids est calculé ainsi :

$$\Delta w_{ij} = -\eta \frac{\partial E}{\partial w_{ij}} \quad (4.8)$$

$$\Delta w_{ij} = \eta \delta_i y_j \quad (4.9)$$

avec

$$\delta_i = y_i [1 - y_i] \sum_{k \in C} \delta_k w_{ki} \quad (4.10)$$

Pour notre cas, on a adopté un RN avec une seule couche cachée. Les entrées du RN correspondent aux variables SST, SSS, et Chl- a , les sorties désirées correspondent à la variable fCO₂. La fonction de transfert utilisée est la tangente-sigmoïde :

$$\varphi(\text{net}_i + b_i) = \frac{2}{(1 + e^{-2(\text{net}_i + b_i)}) - 1} \quad (4.11)$$

4.3 Le partitionnement de données

La classification ou le partitionnement de données (*clustering*) relève des méthodes statistiques d'analyse de données, et a pour objectif d'organiser un ensemble de données en plusieurs sous-ensembles ou classes (*clusters*), de sorte que les composants de chaque classe aient une grande ressemblance (minimisation de l'inertie³ intra-classe), et que les composants appartenant à différentes classes, aient un grand degré de dissemblance (maximisation de l'inertie inter-classe; **Figure 4.5**). Plusieurs algorithmes de classification existent, citons comme exemples : la classification hiérarchique [*hierarchical clustering*; Johnson, 1967], l'algorithme EM [*Expectation-Maximisation*; Dempster et al., 1977], les cartes auto-adaptatives de Kohonen [*self organizing maps*, SOM; Kohonen, 1982], et l'algorithme K -means utilisé dans notre étude.

4.3.1 K -means

K -means est l'une des méthodes les plus utilisées pour les problèmes de partitionnement de données, en raison de sa simplicité en terme de programmation informatique, et aussi de son économie en terme de calculs [MacQueen, 1967]. Bock [2007] a très bien détaillé l'historique de cette méthode, en reprenant toutes les évolutions qu'a connues cette méthode. Bien que le nom " K -means" a été utilisé pour la première fois par MacQueen [1967], le premier à avoir formulé le problème de partitionnement en K -means

3. La somme pondérée des carrés des distances des points au centre de gravité.

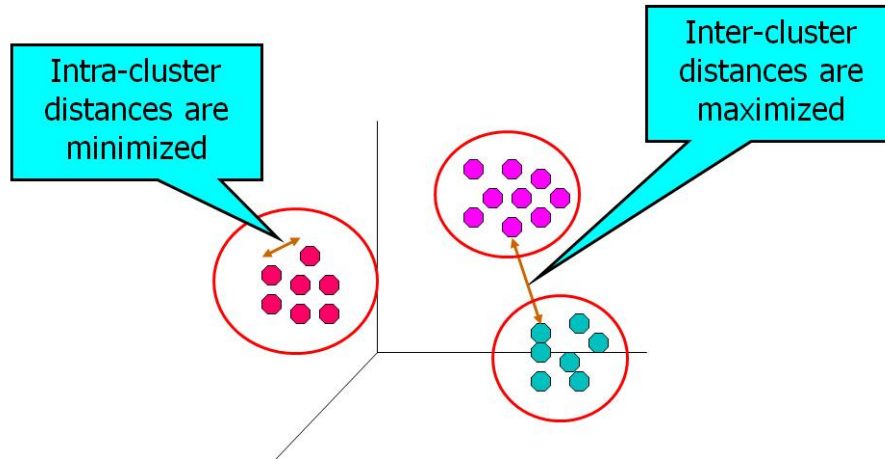


FIGURE 4.5 – Partitionnement de données en 3 classes. [Source : apandre.wordpress.com]

(K -moyennes) était Steinhaus [1956]. Un premier algorithme pour cette approche a été proposé par Lloyd [1957], après cette date, plusieurs études ont été réalisées afin d'améliorer cette méthode [Forgy, 1965; Jancey, 1966; MacQueen, 1967; Hartigan, 1975; Hartigan et Wong, 1979].

La comparaison de cette méthode avec d'autres méthodes de classification, a été réalisée par plusieurs chercheurs. Pour les problèmes de classification, la méthode K -means donne des résultats presque similaires à ceux donnés par les RNs de type SOM [*Self Organizing Maps*; Bação *et al.*, 2005], ou même meilleurs dans certains cas [moins de points mal classés; Balakrishnan *et al.*, 1994].

K -means est une méthode représentée par un algorithme itératif, qui a pour objectif de partitionner un ensemble de n observations (x_1, x_2, \dots, x_n) en k ($k \leq n$, doit être défini avant de commencer l'algorithme) classes $C^{(t)} = C_1^{(t)}, C_2^{(t)}, \dots, C_k^{(t)}$, chaque classe est définie par son centroïde (moyenne) $m_i^{(t)}$. Chaque observation est un vecteur de dimension d , qui est affecté à une classe en fonction du centroïde le plus proche. L'algorithme consiste en les étapes suivantes :

1. Initialisation aléatoire de k moyennes : $m_1^{(1)}, m_2^{(1)}, \dots, m_k^{(1)}$,
2. Affectation de chaque observation à la classe la plus proche, en calculant une distance choisie :

$$C_i^{(t)} = \left\{ x_p : \|x_p - m_i^{(t)}\|^2 \leq \|x_p - m_j^{(t)}\|^2 \forall j, 1 \leq j \leq k \right\} \quad (4.12)$$

3. Mettre à jour les moyennes des classes ;

$$m_i^{(t+1)} = \frac{1}{|C_i^{(t)}|} \sum_{x_j \in C_i^{(t)}} x_j \quad (4.13)$$

4. Répéter l'étape 2 et l'étape 3 jusqu'à ce que :
 - Deux affectations successives soient identiques ($\pm \epsilon$),
 - ou
 - Une autre condition soit vérifiée (ex. un nombre maximal d'itération).

La **Figure 4.6** illustre un exemple de déroulement de cet algorithme, pour un ensemble de données qu'on veut répartir en 3 classes. Le résultat est défini par les 3 centroïdes.

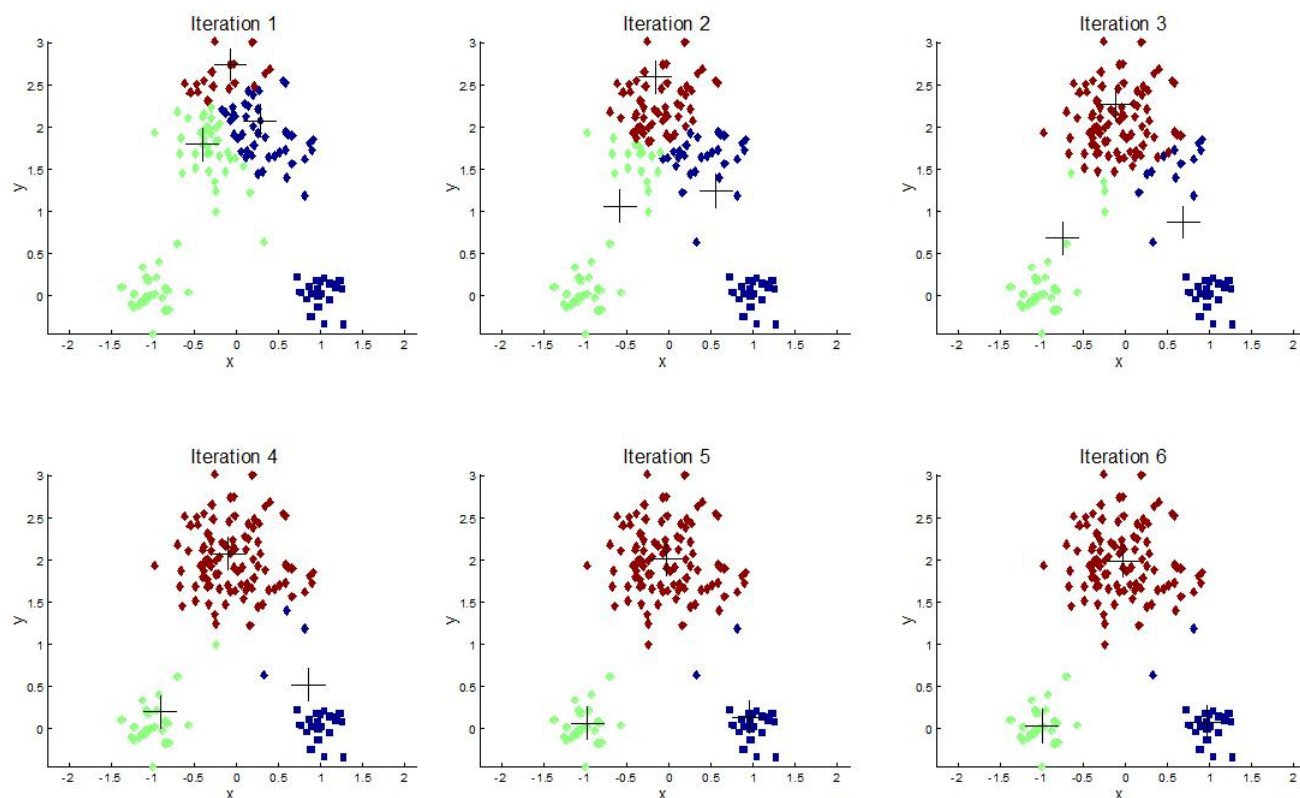


FIGURE 4.6 – Déroulement de l'algorithme K -means en 6 itérations, pour une répartition en 3 classes. Les croix représentent les centroïdes des classes [Source : apandre.wordpress.com]

4.4 langage de programmation : MATLAB®

Le langage que j'ai utilisé pour l'implémentation des programmes, est le langage le plus utilisé par les scientifiques et les ingénieurs, à savoir ; MATLAB®.

MATLAB® est une abréviation de "*matrix laboratory*". Ce langage haut-niveau de quatrième génération⁴, est développé par "Mathworks" (<http://fr.mathworks.com/>), et est utilisé dans plusieurs domaines ; apprentissage machine, traitement de signal, traitement d'image, finance, robotique, et plusieurs d'autres disciplines [The MathWorks, Inc., 2016]. Il est destiné principalement pour le calcul numérique, et est optimisé pour opérer sur les matrices et les vecteurs de façon très rapides, ce qui rend son utilisation très efficace pour l'analyse des grandes BDs, et pour la conception de modèles mathématiques et informatiques.

Tous les programmes MATLAB implémentés pour cette étude, sont présents dans l'**Annexe A**.

4. Un type apparu en 1980, basé sur de simples commandes proches du langage naturel.

Chapitre 5

PRÉSENTATION DES DONNÉES

Ce chapitre contient la présentation de toutes les données utilisées dans cette étude, ainsi que l'étape de construction des bases de données (BDs) pour la phase de l'interpolation de la $f\text{CO}_2$ (**Chapitre 6**).

Avant de passer aux données, on présente dans la **Figure 5.1** l'algorithme général de l'approche proposée dans cette étude pour l'interpolation de $f\text{CO}_2$. Il se compose des étapes suivantes :

- (a) Étant donné qu'on travaille sur de larges ensembles de données, cette première étape est considérée comme étant très importante, car elle permet de mieux gérer les données dans la suite du travail.
- (b) Cette deuxième étape consiste à utiliser les données satellitaires de SST, afin de partitionner la zone en trois classes présentées sous forme d'intervalles.
- (c) Les intervalles de SST retrouvés lors de l'étape (b) sont utilisés pour classifier les données *in situ*.
- (d) Pour chaque classe produite en (c), un ensemble de 20 RNs est construit en passant par trois étapes : apprentissage et validation du RN (en utilisant des données *in situ*), et ensuite le test du RN (en utilisant des données satellitaires et climatologiques).
- (e) Pour chaque classe, choisir le RN optimal (parmi l'ensemble de 20 RNs) en fonction du RMSE minimal retrouvé lors de l'étape de test.
- (f) Pour chaque saison, utiliser les trois RNs construits, pour produire des cartes mensuelles de $f\text{CO}_2$ à partir des données satellitaires.

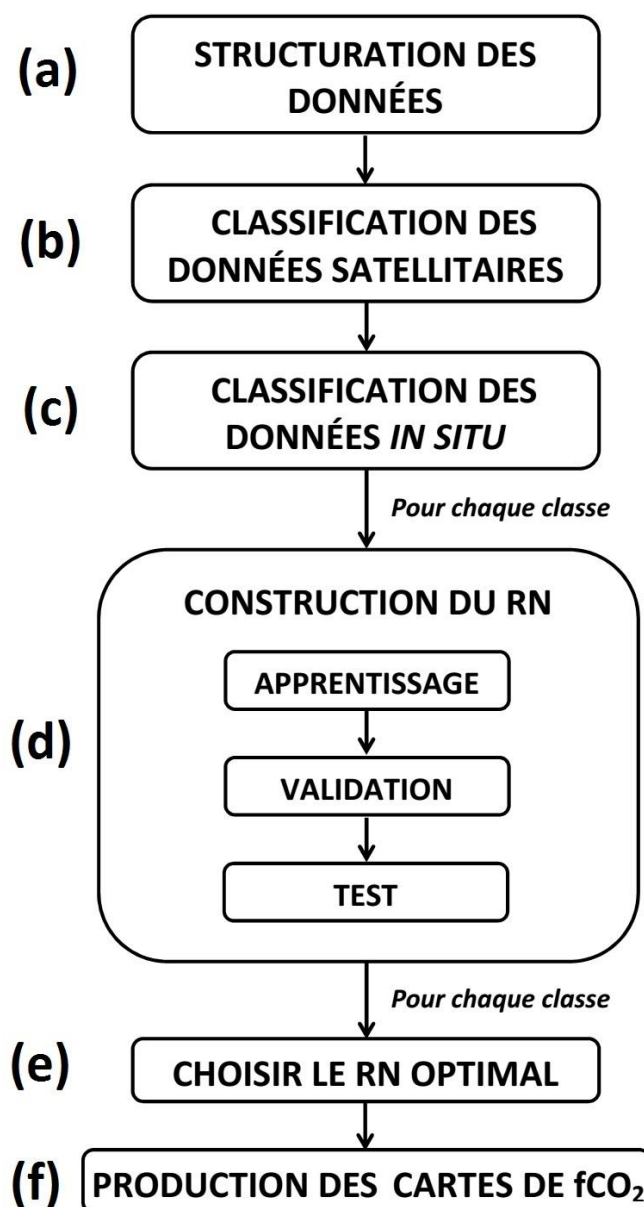


FIGURE 5.1 – Algorithme général résumant les principales étapes de cette étude.

Trois types de données ont été utilisés dans cette étude ; *in situ*, satellitaires, et assimilées :

1. Données *in situ* : utilisées pour la créations des RNs (SST, SSS, et fCO₂).
2. Données satellitaires : utilisées dans plusieurs étapes :
 - pour la classification : SST de MODIS-A (4 km × 4 km),
 - pour la construction des RNs : puisque la BD *in situ* ne contient pas des données de Chl-*a*, on a utilisé les données satellitaires ; SeaWifs (9 km × 9 km) et MODIS-A (4 km × 4 km),
 - pour le test des RNs : SST et Chl-*a* de MODIS-A (4 km × 4 km), et SSS de SMOS (25 km × 25 km),

- pour la création des cartes de $f\text{CO}_2$: SST et Chl-*a* de MODIS-A (4 km × 4 km), et SSS de SMOS (25 km × 25 km).

3. Données assimilées : puisque les données de SSS SMOS ne sont pas disponibles entre 2002 et 2009, les données de SSS SODA (Simple Ocean Data Assimilation, 50 km × 50 km [Carton et Giese, 2008]), ont été utilisées pour le test des RNs.

5.1 Données *in situ*

La BD SOCAT V.3 utilisée (brièvement présentée dans la **Section 2.8**) est de type "*Subset of synthesis data set (ii)*" [Bakker *et al.*, 2014, 2016] téléchargée en utilisant l'outil interactif en ligne : "*SOCAT Data Viewer*" (http://ferret.pmel.noaa.gov/SOCAT_Data_Viewer/). Avant de récupérer les données, cet outil permet de choisir la région d'étude ainsi que les paramètres, afin d'éviter d'avoir un très grand nombre de données. Dans notre cas, les limites de la zone d'étude sont : 23,5°S-70°O à 23,5°N-15°E, et les paramètres choisis sont : **année - mois - jour - heure - minute - longitude - latitude - bathymétrie - distance à la masse terrestre - SST - SSS - $f\text{CO}_2$** . Le fichier télécharger est de type "*NetCDF*"¹ d'une taille de 236 Mo.

La première étape consiste à lire ce fichier (suivant le code MATLAB présenté dans l'**Annexe A.2**), pour ensuite extraire les données qu'il contient. Une première étape de structuration de ces données est donc appliquée, pour pouvoir les enregistrer sous format MATLAB (`.mat`), afin de rendre leur utilisation facile (le code MATLAB est présenté dans l'**Annexe A.3**). La structure (`.mat`) choisie consiste en un tableau de type "*Cell*"² contenant 3 vecteurs : le premier contient les années, le deuxième contient les mois, et le troisième contient les données de ces mois. Le nombre global de données (avant tout prétraitement) était de 632778 points, réparties entre 2000 et 2014.

5.1.1 Prétraitement des données *in situ*

C'est une étape de "filtrage des données *in situ*" afin de ne garder que les données utilisables. En d'autres termes, la BD téléchargée contient des données réparties sur toute la surface de l'océan Atlantique tropical, néanmoins, étant donné que l'océan se compose de deux parties : une dite "côtière" (*coastal ocean*, profondeur < 500 m), et l'autre dite "grand large" (*open ocean*, profondeur > 500 m), et que ces deux zones sont caractérisées par des propriétés chimiques, biologiques, et physiques très différentes les unes des autres, ceci donne lieu à un fort gradient entre ces deux zones. Comme vu dans la **Figure 5.2**, lorsqu'on est dans les zones dites côtières, les valeurs de $f\text{CO}_2$ augmentent et/ou diminuent brusquement, particulièrement pour les deux saisons printemps et été.

Par ailleurs, l'étude présentée ici ne concerne que la partie dites *open ocean*. Ainsi, les données *in situ* ont été structurées par saison (suivant le code MATLAB dans l'**Annexe A.3.1.1**), après cela, un premier filtre a été appliqué afin d'éliminer les points situés dans

1. Network Common Data Form : "un ensemble de bibliothèques logicielles et d'un format de données auto-descriptives (contient des informations sur les données qu'il contient), permettant la création, l'accès, et le partage des données scientifiques" [UNIDATA, 2016].

2. Un type MATLAB permettant de rassembler dans un seul tableau, des données de différents types, et de différentes tailles.

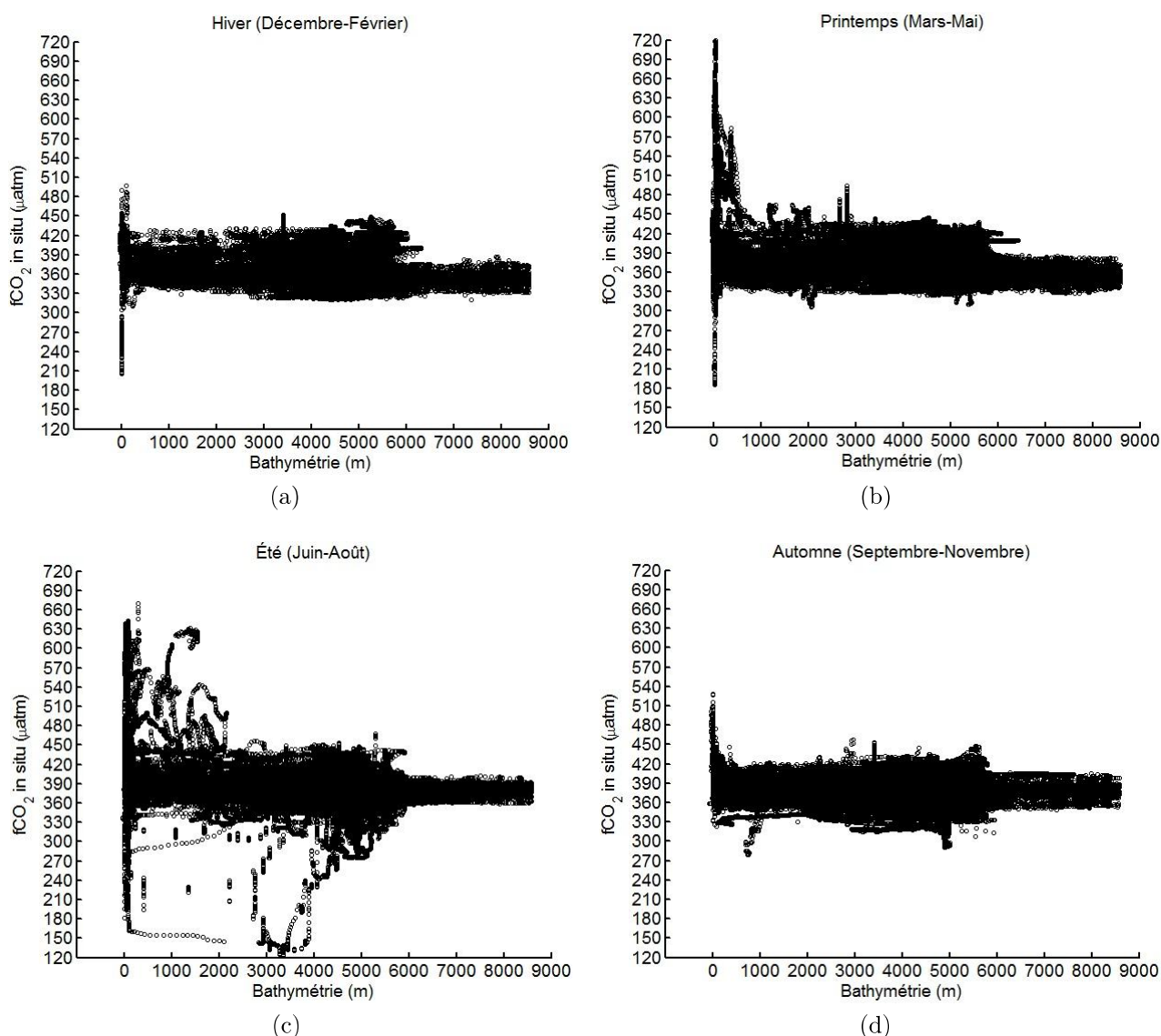


FIGURE 5.2 – Variation de la $f\text{CO}_2$ en fonction de la bathymétrie pendant les quatre saisons.

les zones côtières. Pour ceci, j'ai utilisé le paramètre "**bathymétrie**" donnant l'information sur la profondeur de l'océan en chaque point de la BD. Un seuil de 500 m a été appliqué sur ce paramètre afin d'éliminer tous les points où la profondeur de l'océan est inférieure 500 m.

En deuxième lieu, comme il y a des points situés à l'embouchure des deux fleuves Amazone et Orénoque, il y a des valeurs de SSS qui descendent jusqu'à 16 (voir **Figure 5.3**). De ce fait, un seuil de 32 a été appliqué sur les valeurs de SSS, afin d'éliminer ces points.

Le code MATLAB utilisé pour l'application des deux filtres précédents est présenté dans l'**Annexe A.4.1**.

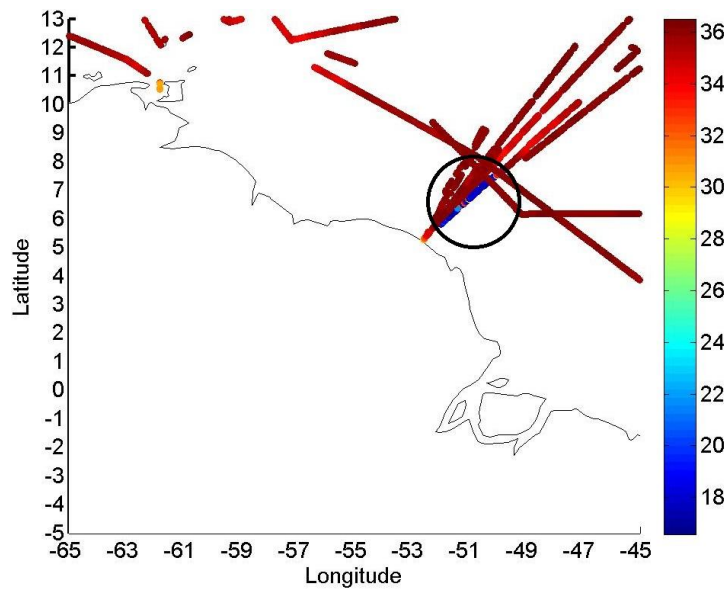


FIGURE 5.3 – Extrait de la BD *in situ* où l'on peut voir de très faibles valeurs de SSS (cercle noir).

La **Figure 5.4** montre que l'application de ces deux filtres a permis d'éliminer les points qui présentent un très fort gradient sur les valeurs de $f\text{CO}_2$. On peut constater que, pour la saison de l'été, les points ayant de faibles valeurs de $f\text{CO}_2$ et situés dans des zones où la bathymétrie est inférieure à 500 m, ont été éliminés, parce que ces points étaient associés à de faibles valeurs de salinités (< 32), et donc leur élimination a été faite indirectement avec l'application du seuil sur la SSS. Cependant, pour cette même saison (été), on remarque que même après l'application des filtres, il reste des points où les valeurs de $f\text{CO}_2$ sont très fortes par rapport à la tendance générale pendant cette saison, et d'autres points où la $f\text{CO}_2$ est faible.

Afin d'éliminer ces points, on a procédé à quelques étapes de vérification. La première est la localisation de ces points dans la zone d'étude. Comme on voit dans la **Figure 5.5a**, les points présentant de fortes valeurs de $f\text{CO}_2$ (cercle noir) sont situés près de la côte Ouest-Africaine, qui est connue par les upwellings permanents (**Section 2.3**). C'est fortes $f\text{CO}_2$ sont associées à de faibles SST, et à de fortes Chl-*a* (**Figure 5.6**).

La **Figure 5.5b** montre les points où la $f\text{CO}_2$ est faible (cercle noir). L'eau dans cette zone est mélangée avec celle de l'Amazonie, ce qui explique ces faibles valeurs de $f\text{CO}_2$, associées à de fortes valeurs de Chl-*a* (**Figure 5.7**).

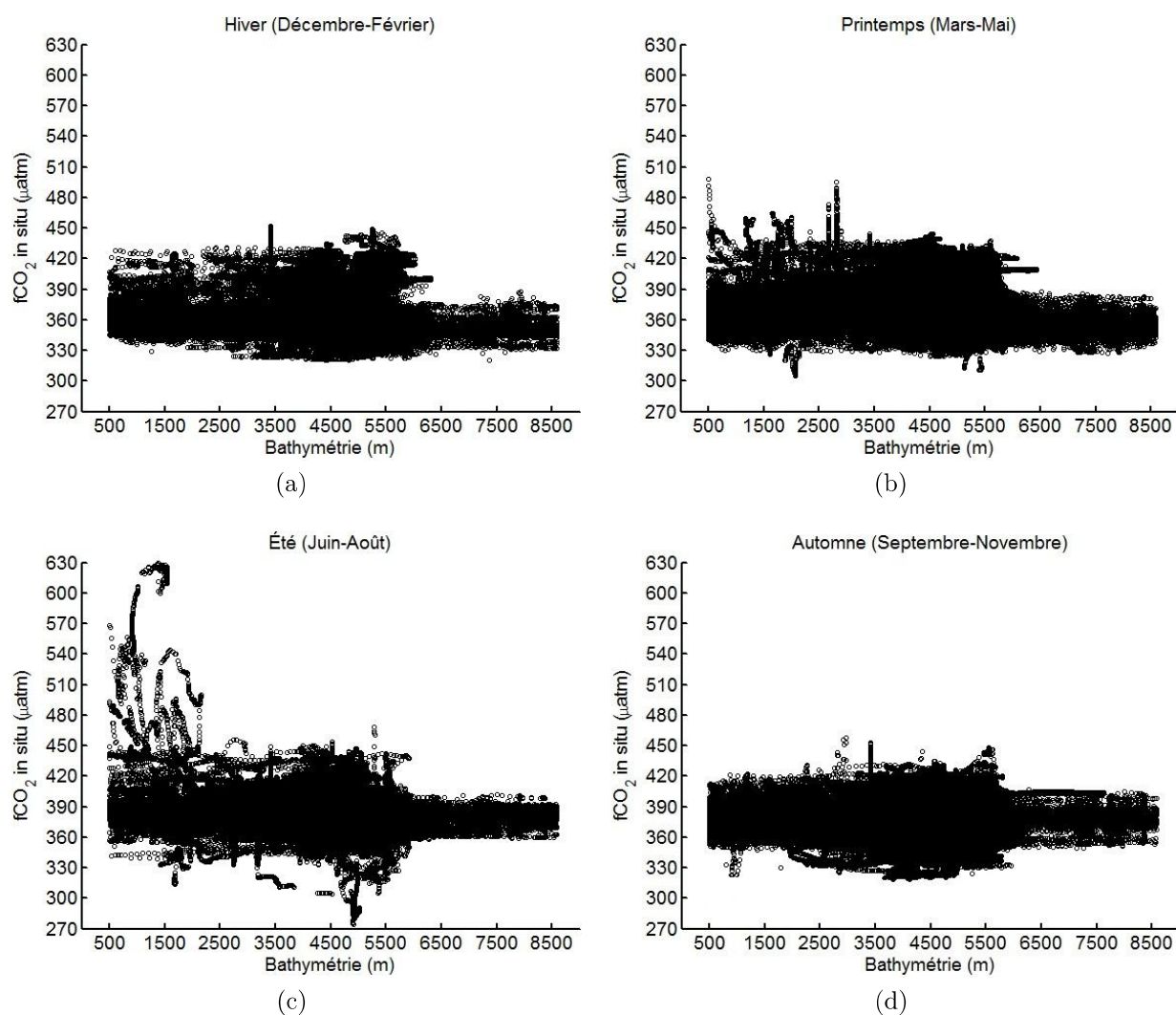


FIGURE 5.4 – Variation de la $f\text{CO}_2$ en fonction de la bathymétrie pendant les quatre saisons, après l'application des deux premiers filtres.

En affichant la $f\text{CO}_2$ en fonction de la distance à la masse terrestre dans la **Figure 5.8a**, on remarque que les $f\text{CO}_2$ élevées sont situées à une distance inférieure à 140 km par rapport à la terre. Ainsi, pour écarter ces valeurs, on a calculé la boîte de dispersion présentée dans la **Figure 5.8b**. Puisque, statistiquement, les valeurs de $f\text{CO}_2$ (de l'ensemble des données de l'été) considérées comme aberrantes, commencent à partir de 425 μatm , et comme, par ailleurs, il y a des endroits dans la zone d'étude où la $f\text{CO}_2$ atteint 450 μatm pendant la saison de l'été (**Figure 5.5**), afin d'écarter les points où la $f\text{CO}_2$ est très forte, une condition regroupant deux critères a été appliquée sur ces données. Ces deux critères sont : une distance à la masse terrestre inférieure à 140 km, associée à des valeurs de $f\text{CO}_2$ dépassant 450 μatm .

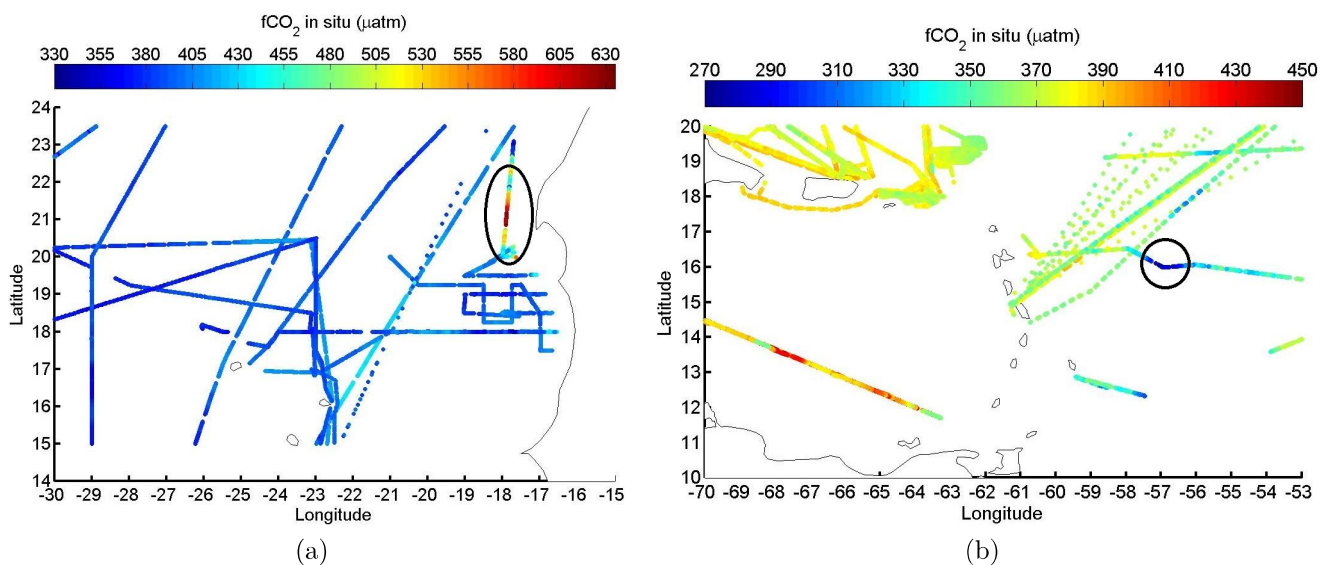


FIGURE 5.5 – Variabilité de $f\text{CO}_2$ pendant la saison de l'été (Juin-Août) : les zones contenant de fortes valeurs de $f\text{CO}_2$ dans (a), et de faibles valeurs de $f\text{CO}_2$ dans (b).

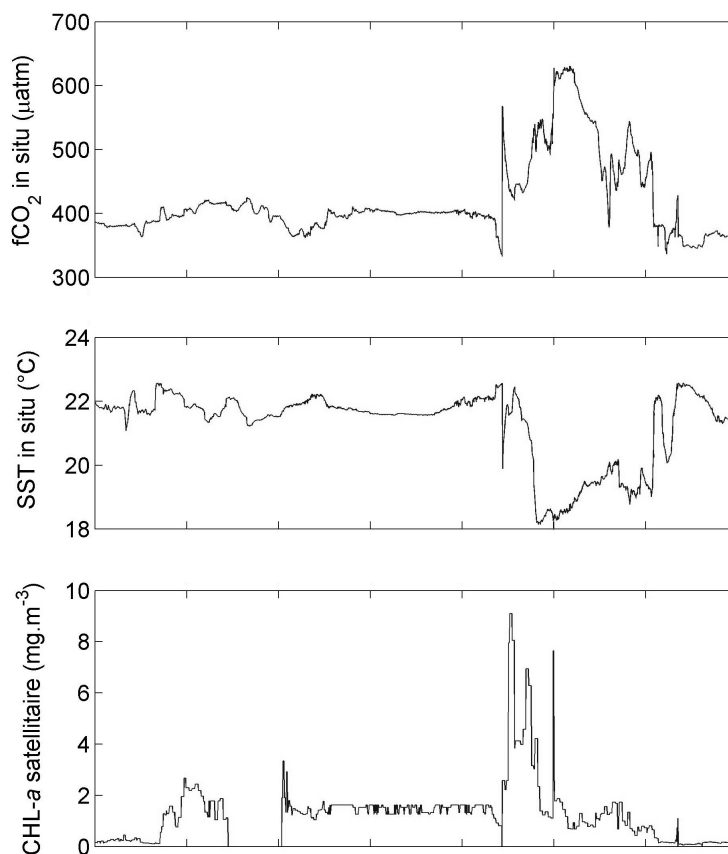


FIGURE 5.6 – Variabilité de $f\text{CO}_2$, SST, et Chl-a pendant la saison de l'été, dans la zone entourée par un cercle noir dans la **Figure 5.5a**.

Pour les faibles valeurs de $f\text{CO}_2$, comme il n'y a pas de relation avec la distance à

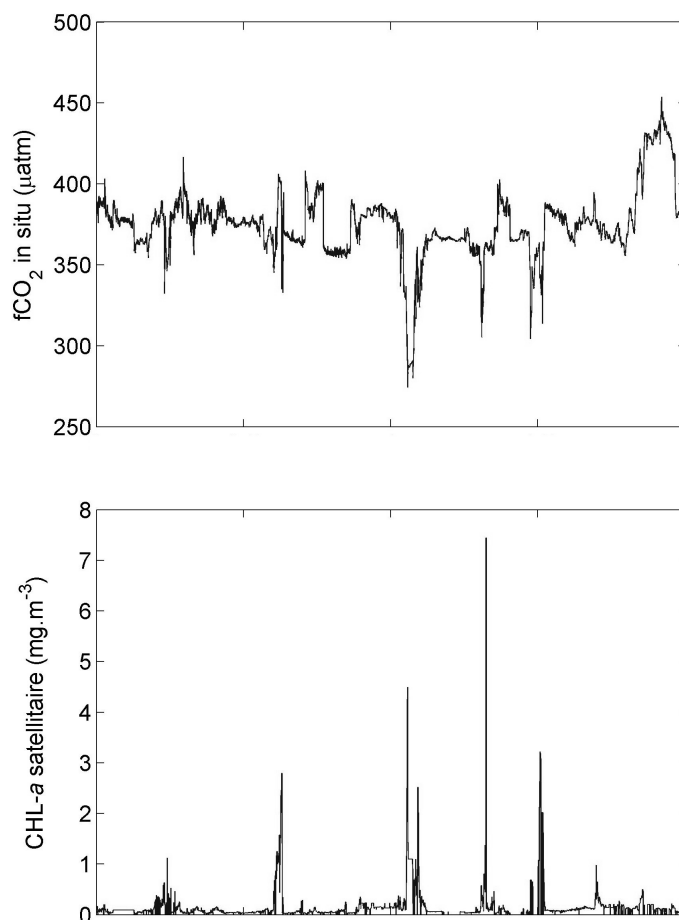


FIGURE 5.7 – Variabilité de $f\text{CO}_2$ et de $\text{Chl-}a$ pendant la saison de l'été, dans la zone entourée par un cercle noir dans la **Figure 5.5b**.

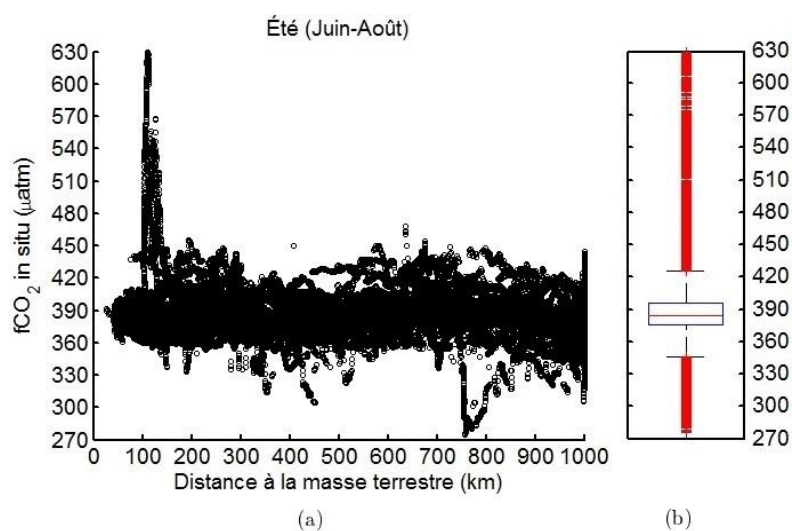


FIGURE 5.8 – $f\text{CO}_2$ pendant la saison de l'été, après le premier prétraitement ; (a) : variabilité en fonction de la distance à la masse terrestre. (b) : boîte à dispersion de $f\text{CO}_2$.

la masse terrestre (**Figure 5.8a**), seulement la boîte à dispersion a été utilisée (**Figure 5.8b**). La limite inférieure donnée par celle-ci est de 340 μatm , cependant, des valeurs de $f\text{CO}_2$ avoisinant 340 μatm sont fréquents dans l'ensemble de données, ainsi, un seuil de 320 μatm a été appliqué sur les valeurs de $f\text{CO}_2$ de la saison de l'été.

Pour ces deux filtres cités ci-dessus, le code MATLAB utilisé est présenté dans l'**Annexe A.4.2**).

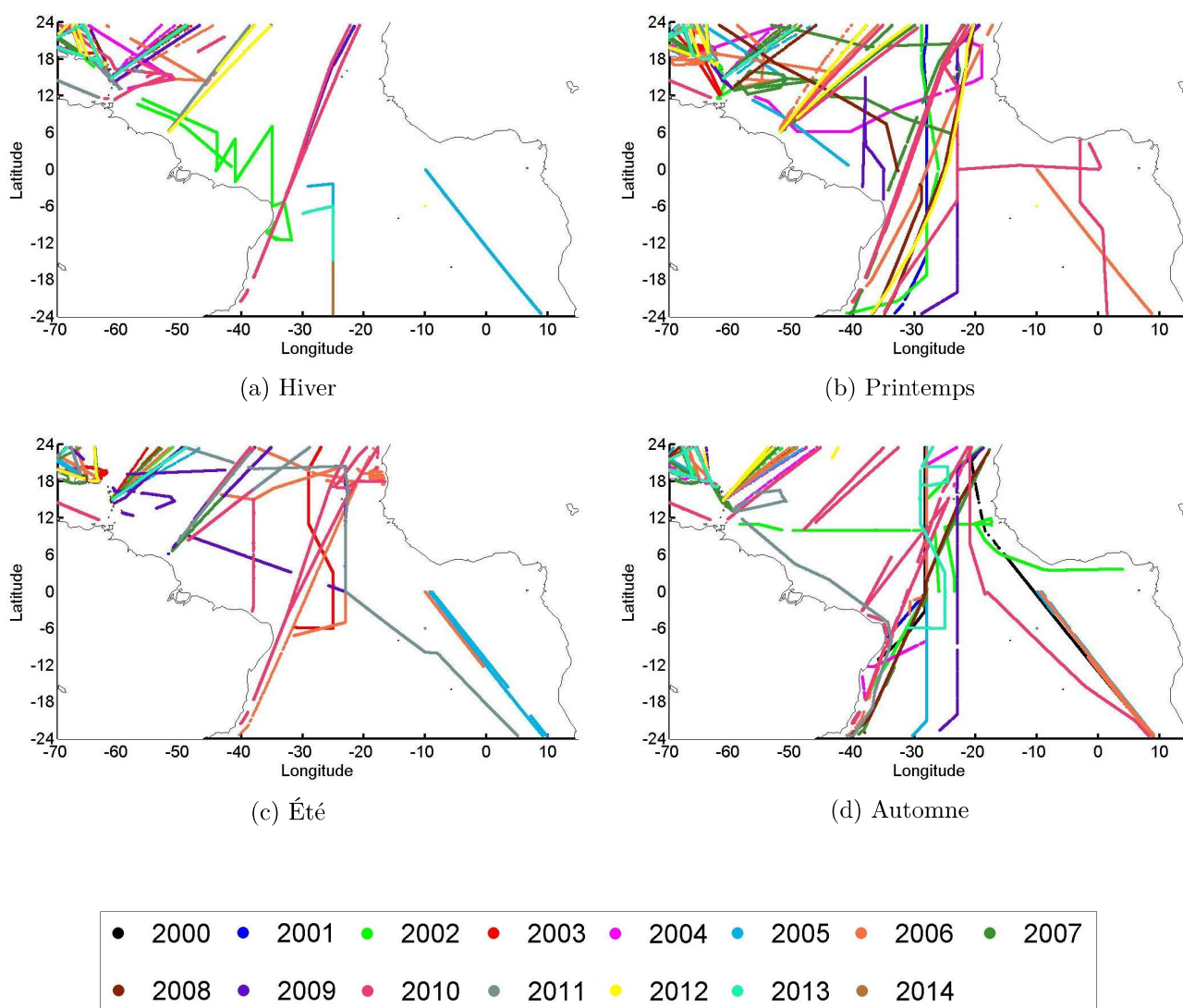
Après ces étapes de prétraitement, le nombre total de données restantes est devenu 549968 (le nombre de données pour chaque saison est présenté dans le **Tableau 5.1**).

Saisons	Hiver			Printemps		
Mois	DEC	JAN	FEV	MAR	AVR	MAI
NDM	35457	23060	33830	43976	65179	57559
NDS	92347			166714		
%*	16,79			30,31		
Saisons	Été			Automne		
Mois	JUN	JUI	AOU	SEP	OCT	NOV
NDM	40500	44958	47557	25932	51899	80061
NDS	133015			157892		
%*	24,19			28,71		

TABLEAU 5.1 – Nombre de données de $f\text{CO}_2$ de surface (extraites de la BD SOCAT V.3, de 23,5°S-70°O à 23,5°N-15°E) après l'étape de prétraitement, par mois et par saison. NDM : Nombre de données par mois, NDS : Nombre de données par saison, * Le pourcentage de données par rapport au nombre total de données (549968)

La distribution spatiale des données est illustrée dans la **Figure 5.9**. Il est très visible que la partie Sud-Est est la moins échantillonnée. D'après cette figure et le **Tableau 5.1**, on note que la saison du printemps et celle contenant le plus de données, tandis que la saison de l'hiver est celle qui en contient le moins.

Vu que pour les quatre saisons de l'année, on est passé par les même étapes pour le traitement de données ainsi que pour la construction des RNs, on présentera dans ce qui va suivre de ce chapitre et dans le chapitre suivant (**Chapitre 6**), la variabilité des paramètres (*in situ* et satellitaires) des deux saisons printemps et hiver, ainsi que les résultats de la construction des RNs pour ces mêmes saisons dans le **Chapitre 6**. La présentation des paramètres ainsi que les résultats obtenus pour les deux autres saisons (été et automne) seront présentés en annexe (**B** et **C**).

FIGURE 5.9 – Distribution spatiale des données de $f\text{CO}_2$ de surface.

5.1.1.1 Variabilité des paramètres *in situ* : Printemps

Pour la saison du printemps (mars, avril, et mai), les données sont disponibles de 2001 à 2013, avec 166714 points.

SST : Entre 2001 et 2013, les moyennes de SST varient de 25,89 à 27,52 °C (**Tableau 5.2**), autour d'une moyenne globale de l'ordre de 26,68 °C (**Figure 5.10a**). Comme vu dans la **Figure 5.10b**, vu que la zone d'étude est très grande, les données sont disponibles dans des régions différentes d'une année à une autre et même dans une même année, ce qui induit à une forte variance dans une même année, et à des écarts-types élevés (**Figure 5.10a** et **Tableau 5.2**).

La **Figure 5.10b** présente la distribution spatiale de SST pendant cette saison, on remarque que le Nord-Est et le Sud-Est sont les zones où la SST est la plus faible (elle atteint une valeur minimale de 18,2 °C en 2005, **Tableau 5.2**). Les valeurs de SST

Année	Minimum (°C)	Maximum (°C)	Moyenne (°C)	Écart-type (°C)
2001	21,48	28,88	26,34	2,33
2002	21,7	28,86	26,4	0,76
2003	25,06	28,36	27,01	0,54
2004	19,77	27,96	25,93	1,81
2005	18,2	28,96	23,74	3,45
2006	18,41	29,1	26,64	1,91
2007	21,38	29,64	26,35	1,83
2008	21,04	29,68	27,03	1,82
2009	20,07	29,91	26,05	2,42
2010	22,17	30,76	27,52	1,93
2011	24,96	28,76	26,57	0,88
2012	19,71	28,42	25,89	2,04
2013	24,94	28,47	26,6	0,97

TABLEAU 5.2 – Paramètres statistiques de SST de toutes les années pendant la saison du printemps.

maximales se trouvent, quant-à-elles, dans la partie équatoriale de la zone (la SST atteint une valeur maximale de 30,76 °C en 2010, **Tableau 5.2**). À partir de cette figure (**5.10b**) et les deux figures **5.10c** et **5.10d**, on constate que la SST est globalement beaucoup plus corrélée avec la latitude qu'avec la longitude, et que la SST augmente lorsqu'on se rapproche de l'équateur (latitude 0°). Comme on l'a mentionné ci-dessus, puisque la zone est trop grande, on retrouve des trajets très différents les uns des autres, et donc, en calculant le coefficient de corrélation (R)³ global entre la SST et la latitude, on retrouve une valeur de -0,44. Cependant, cette corrélation est très forte pour les années où l'on a un seul trajet (ou plusieurs trajets similaires), par exemple pour l'année 2001 où l'on a des données sur un seul trajet allant de 23,5°S à 23,5°N, le coefficient R entre la SST et la latitude est de -0,85.

Pour la longitude, sa corrélation avec la SST reste très faible (R = -0,02), et même si l'on prend année par année, le coefficient R reste faible (par ex. pour 2001, elle est de -0,22), Néanmoins, on peut voir dans les deux **Figures 5.10b** et **5.10d**, que pour les années qui contiennent des données au Nord et au Sud de la zone, en allant d'Ouest en Est, la SST a tendance à diminuer.

3. Mesure de la dépendance linéaire entre deux variables X et Y : $Cor(X, Y) = \frac{Cov(X, Y)}{\sigma_X \cdot \sigma_Y}$, où $Cov(X, Y)$ est la covariance de X et Y , et σ_X et σ_Y leurs écarts-types.

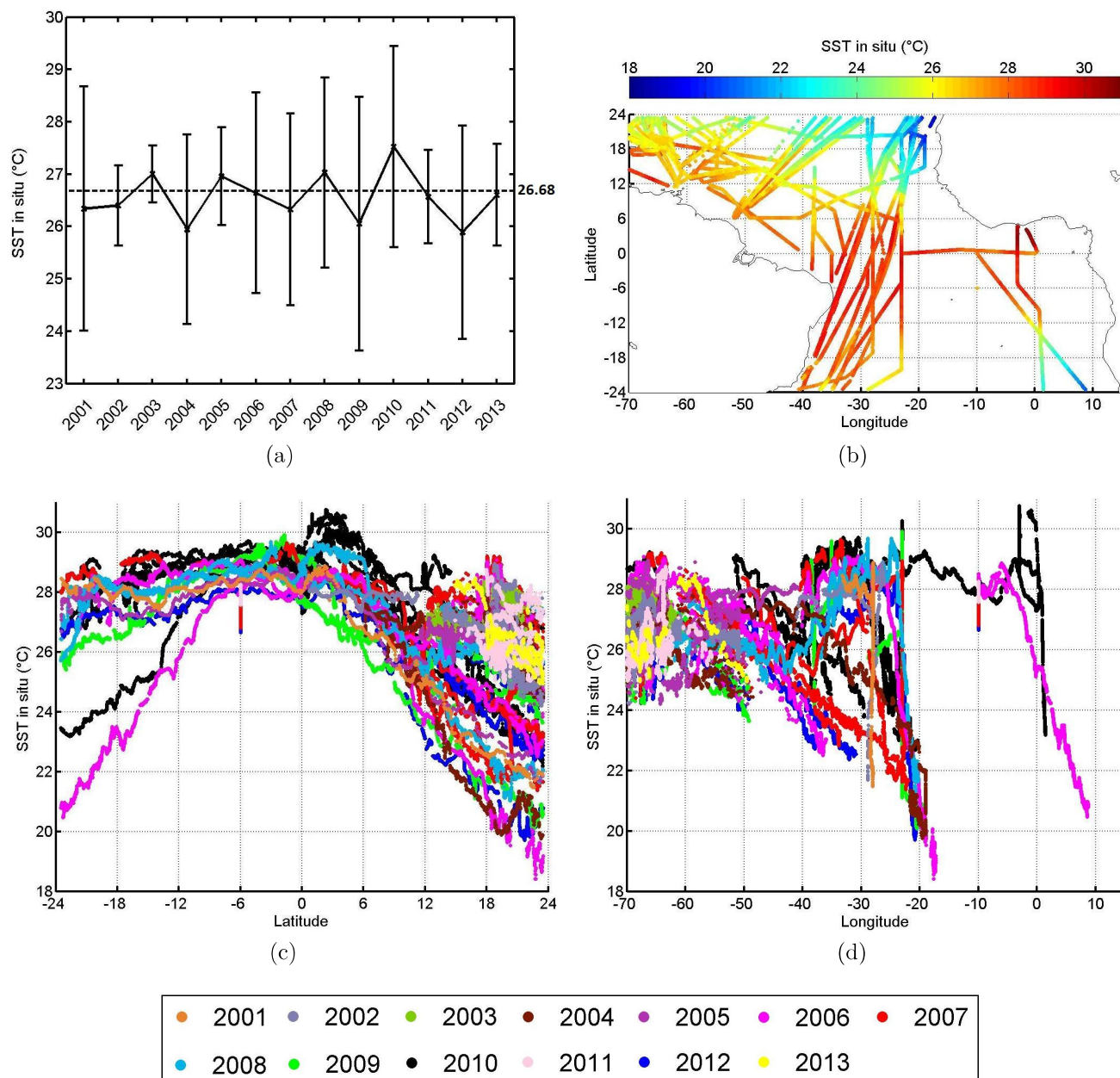


FIGURE 5.10 – Représentation de la variabilité de SST *in situ* pendant la saison du printemps. (a) Moyenne et écarts-type de SST de chaque année, (b) Distribution spatiale de SST, (c) Variation de SST en fonction de la latitude, (d) Variation de SST en fonction de la longitude

SSS : Entre 2001 et 2013, les moyennes de SSS varient de 36,01 à 36,50 (**Tableau 5.3**), autour d'une moyenne globale de l'ordre de 36,20 (**Figure 5.11a**). Comme pour la SST, les écarts-types de SSS sont importants (**Figure 5.11a** et **Tableau 5.3**) à cause de l'étendue de la zone d'étude.

D'après la distribution spatiale de SSS illustrée dans la **Figure 5.11b**, ce paramètre est plus élevé dans le Nord et le Sud-Ouest de la zone (la valeur maximale atteinte est de 37,63 en 2001, **Tableau 5.3**). La SSS présente de faibles valeurs dans : la partie

Année	Minimum	Maximum	Moyenne	Écart-type
2001	34,306	37,633	36,496	0,601
2002	34,57	37,625	36,067	0,351
2003	34,71	36,913	36,008	0,311
2004	34,008	37,331	36,309	0,398
2005	33,8	37,324	36,083	0,546
2006	33,46	37,457	36,033	0,55
2007	32,025	37,576	36,247	0,648
2008	35,03	37,588	36,359	0,61
2009	33,79	37,365	36,209	0,65
2010	32,085	37,601	36,174	0,714
2011	32,05	37,244	36,029	0,441
2012	34,508	37,475	36,247	0,554
2013	35,033	37,422	36,359	0,632

TABLEAU 5.3 – Paramètres statistiques de SSS de toutes les années pendant la saison du printemps.

équatoriale, une partie de la zone Nord-Ouest, et une partie de la zone Sud-Est (la valeur minimale atteinte est de 32,03 en 2007, **Tableau 5.3**)).

La corrélation entre la SSS et la latitude est moins bonne que celle de la SST ; le coefficient R global est de -0,14. Cependant pour certaines années cette corrélation est assez bonne ; par exemple pour l'année 2001, elle est de -0,6. Comme on voit dans la **Figures 5.11b**, la SSS diminue quand on se rapproche de l'équateur (latitude 0°). Pour les années 2006 et 2010 elle diminue aussi en dessous de 12°S. Ce qu'on peut remarquer aussi, c'est que la SSS chute dans la zone située entre 6-14 °N et 50-60°O.

La corrélation avec la longitude reste aussi faible pour la SSS (R = -0.07, **Figure 5.11d**).

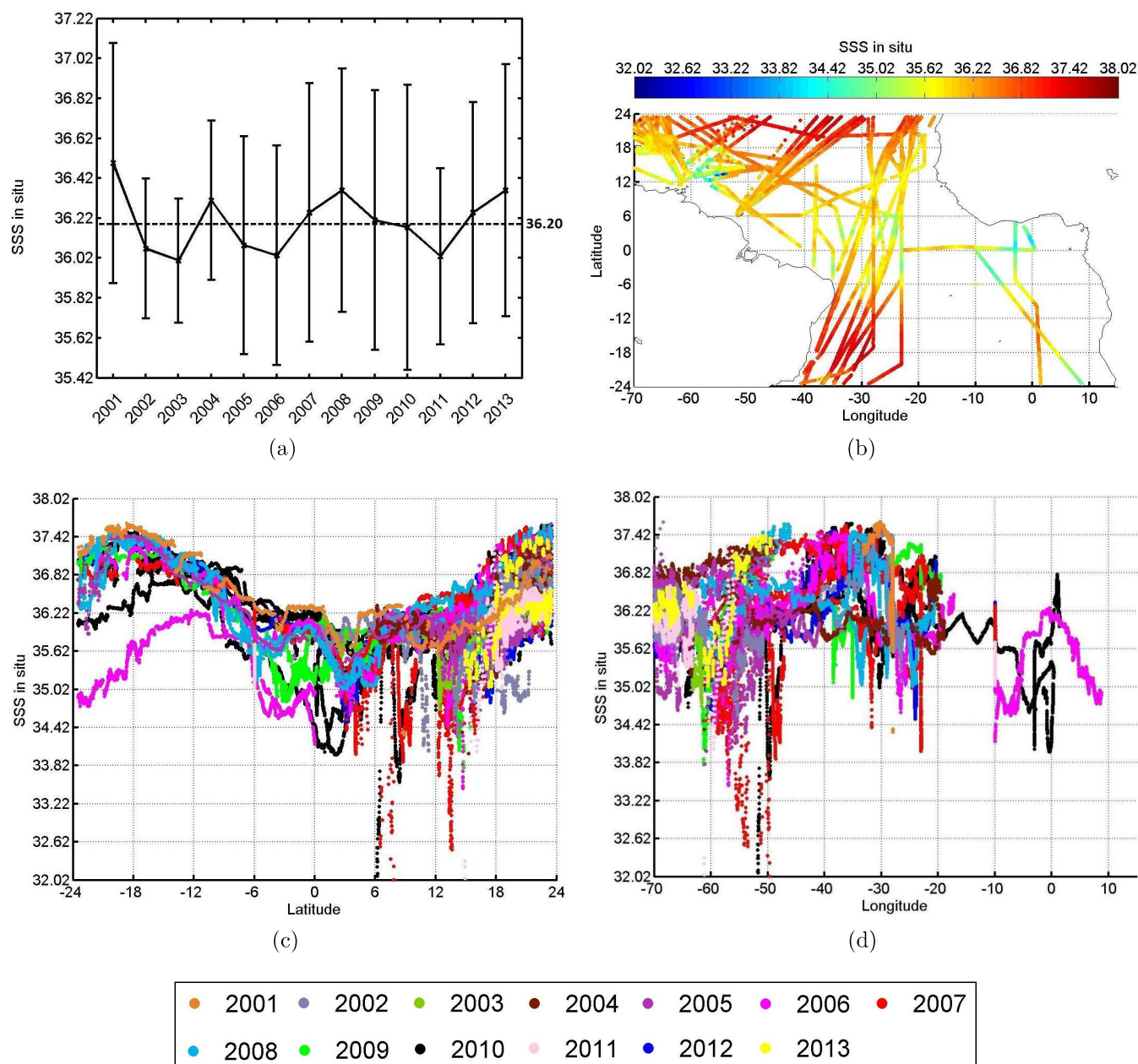


FIGURE 5.11 – Représentation de la variabilité de SSS *in situ* pendant la saison du printemps. (a) Moyenne et écarts-type de SSS de chaque année, (b) Distribution spatiale de SSS, (c) Variation de SSS en fonction de la latitude, (d) Variation de SSS en fonction de la longitude

$f\text{CO}_2$: Entre 2001 et 2013, les moyennes de $f\text{CO}_2$ ont une tendance générale ascendante, elles varient de 348,77 à 394,01 μatm (Tableau 5.4), autour d'une moyenne globale de l'ordre de 377.17 μatm (Figure 5.12a). Pour la même raison que pour la SST et la SSS (la grande étendue de la zone d'étude), les écarts-types de $f\text{CO}_2$ de toutes les années varient entre 9 et 23 μatm (Tableau 5.4).

La distribution spatiale de $f\text{CO}_2$ (Figure 5.12b) montre que ce paramètre est plus élevé dans la partie Sud de la zone (la $f\text{CO}_2$ atteint un maximum de 444.28 μatm en

2010, **Tableau 5.4**), et est plus faible dans la partie Nord (la $f\text{CO}_2$ atteint un minimum de 310,83 μatm en 2009, **Tableau 5.4**).

Contrairement à la SST et la SSS, la corrélation globale entre la $f\text{CO}_2$ et la latitude est assez bonne; $R = -0,67$, et elle aussi aussi bien corrélée avec la longitude; $R = 0,62$ (voir **Figures 5.12c** et **5.12d**). La **Figure 5.12c** montre que la $f\text{CO}_2$ augmente lorsqu'on se rapproche de 6°S (sauf pour quelques points de l'année 2006, où la $f\text{CO}_2$ diminue au niveau de cette latitude) et la **Figure 5.12d** montre que la $f\text{CO}_2$ augmente d'Ouest en Est.

Année	Minimum (μatm)	Maximum (μatm)	Moyenne (μatm)	Écart-type (μatm)
2001	334,66	405,81	364,13	21,29
2002	324,39	406,91	348,77	15,44
2003	328,35	388,26	358,51	10,03
2004	340,59	393,9	361,36	9,02
2005	332,36	408,14	356,84	16,63
2006	324,07	428,13	368,71	19,6
2007	334,25	443,9	372,15	21,98
2008	341,73	425,97	379,73	17,8
2009	310,83	437,54	384,34	23,36
2010	336,86	444,28	394,01	19,79
2011	329,1	392	363,68	12
2012	336,85	431,9	375,06	22,33
2013	350,44	401,41	372,02	14,35

TABLEAU 5.4 – Paramètres statistiques de $f\text{CO}_2$ de toutes les années pendant la saison du printemps.

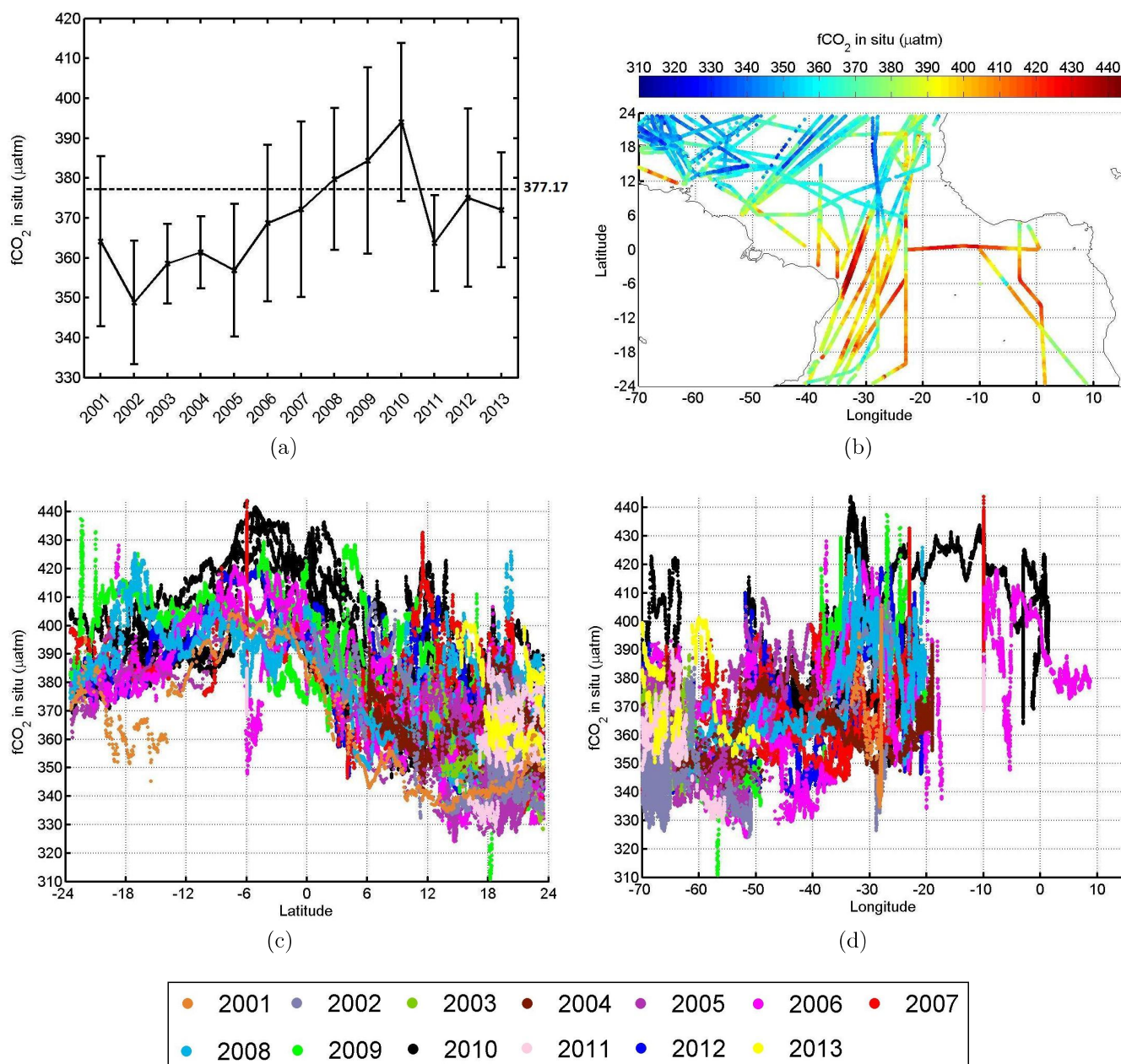


FIGURE 5.12 – Représentation de la variabilité de fCO_2 *in situ* pendant la saison du printemps. (a) Moyenne et écarts-type de fCO_2 de chaque année, (b) Distribution spatiale de fCO_2 , (c) Variation de fCO_2 en fonction de la latitude, (d) Variation de fCO_2 en fonction de la longitude

5.1.1.2 Variabilité des paramètres *in situ* : Hiver

Pour la saison de l'hiver (décembre, janvier, et février), les données sont disponibles de 2002 à 2014, avec 92347 points.

SST : Entre 2002 et 2014, les moyennes de SST varient de 25,1 à 27,7 °C (Tableau 5.5), autour d'une moyenne globale de l'ordre de 26,46 °C (Figure 5.13a). Comme vu

dans la **Figure 5.13a** et le **Tableau 5.5**, les écarts-types de SST sont moins importants que ceux de la saison du printemps. Le plus grand écart-type est celui de l'année 2005, car cette année est la seule qui contient la plus grande gamme de températures, qui va de 19,18 à 28,63 °C.

Année	Minimum (°C)	Maximum (°C)	Moyenne (°C)	Écart-type (°C)
2002	25,91	28,9	27,7	0,5
2003	25,14	28,45	26,87	0,46
2004	24,44	27,95	26,3	0,5
2005	19,18	28,63	25,1	2,66
2006	23,5	28,75	26,1	0,86
2007	24,42	27,89	26,74	0,38
2008	24,29	28,43	26,66	0,88
2009	23,05	28,78	26,45	1,44
2010	22,35	29,25	27,11	1,26
2011	24,46	28,68	26,62	0,99
2012	23,1	27,44	25,61	0,86
2013	23,74	28,13	26,21	0,85
2014	25,21	27,1	25,78	0,41

TABLEAU 5.5 – Paramètres statistiques de SST de toutes les années pendant la saison de l'hiver.

La **Figure 5.13b** présente la distribution spatiale de SST pendant cette saison, on remarque que ; comme pour la saison du printemps, les deux zones les plus froides sont le Nord-Est et le Sud-Est ; la SST atteint une valeur minimale de 19,18 °C en 2005 (**Tableau 5.5**). La SST est plus élevée dans la partie centrale de la zone d'étude, elle atteint une valeur maximale de 29,25 °C en 2010, **Tableau 5.5**).

Pendant cette saison, le coefficient R global entre la SST et la latitude est de 0,05. Cependant, comme on peut constater dans les deux **Figures 5.13b** et **5.13c**, la SST a tendance à augmenter lorsqu'on se rapproche de la latitude 0°. Comme dit dans la saison du printemps, les localisations différentes des trajets affectent la valeur du coefficient R, ainsi, on prenant l'exemple d'un seul trajet allant de 17°S à 23°N de l'année 2009, on retrouve un coefficient R de -0,83.

La corrélation entre la SST et la longitude est assez bonne, globalement, elle est de -0,36. En prenant le même trajet de 2009, le coefficient R est de -0,85. Comme on peut voir dans les deux **Figures 5.13b** et **5.13d**, la SST diminue en allant d'Ouest en Est.

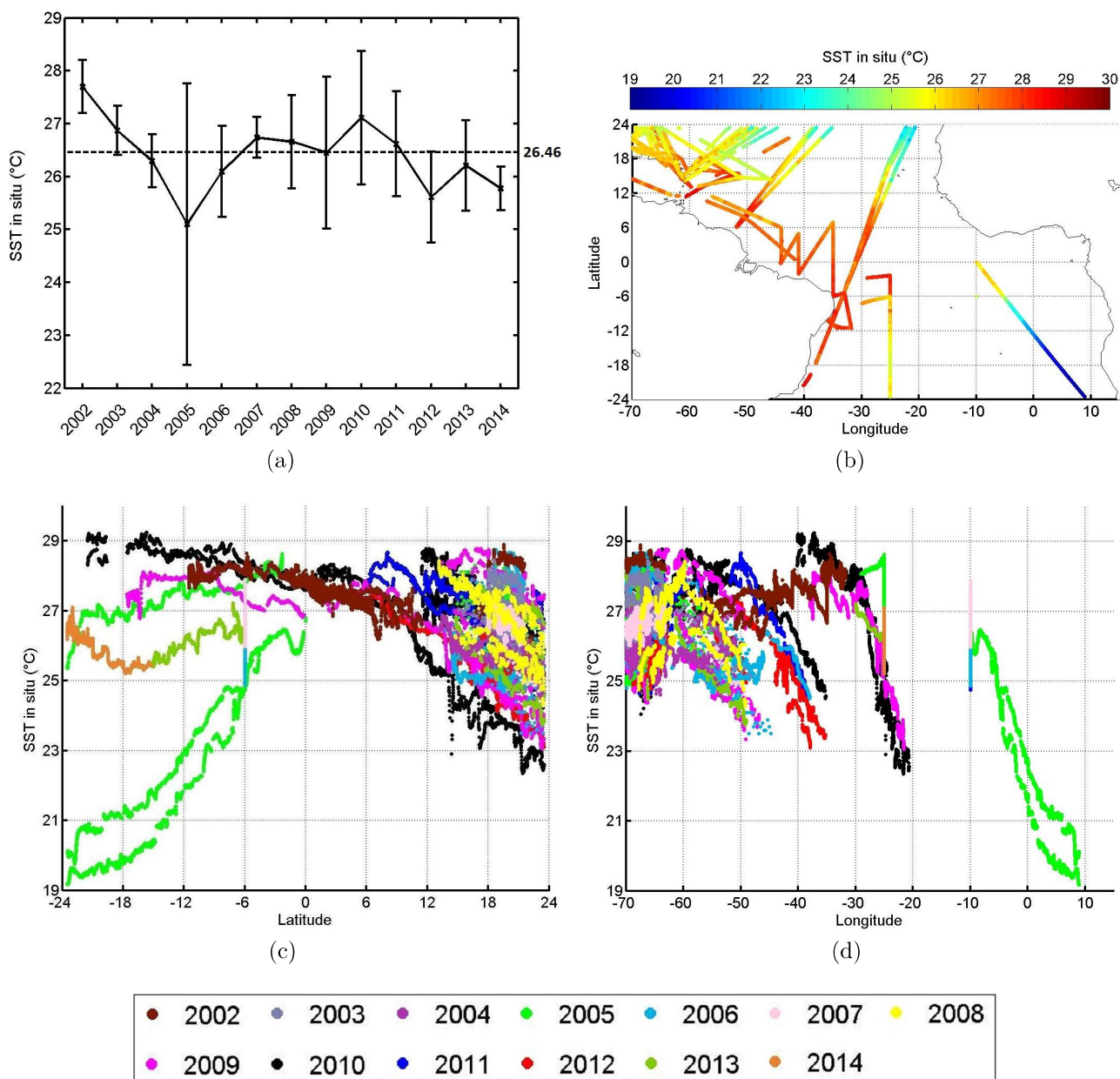


FIGURE 5.13 – Représentation de la variabilité de SST *in situ* pendant la saison de l’hiver. (a) Moyenne et écarts-type de SST de chaque année, (b) Distribution spatiale de SST, (c) Variation de SST en fonction de la latitude, (d) Variation de SST en fonction de la longitude

SSS : Entre 2002 et 2014, les moyennes de SSS varient de 35,86 à 37,24 (**Tableau 5.6**), autour d’une moyenne globale de l’ordre de 36,13 (**Figure 5.14a**). En 2014, la moyenne de SSS est trop élevée par rapport aux autres années, ceci s’explique par le fait que pour cette année, les données sont situées au Sud-Ouest de la zone ; de 23,5°S à 15°S-25°O (**Figure 5.14b**), et que cette région est caractérisée par de fortes valeurs de SSS.

Comme vu dans la **Figure 5.14b**, pareillement à la saison du printemps, la SSS est plus élevée dans le Nord et le Sud-Ouest de la zone, elle atteint une valeur maximale

Année	Minimum	Maximum	Moyenne	Écart-type
2002	33,83	37,336	36,043	0,518
2003	34,731	37,45	36,027	0,363
2004	34,305	37,373	36,102	0,455
2005	34,51	37,4	36,122	0,605
2006	34,986	37,298	36,082	0,455
2007	35,374	37,07	35,901	0,169
2008	34,627	37,319	36,165	0,643
2009	33,69	38,09	36,34	0,753
2010	32,544	37,572	36,063	0,637
2011	32,082	37,506	35,859	0,694
2012	32,81	37,42	36,148	0,555
2013	34,61	37,496	36,43	0,592
2014	36,896	37,392	37,244	0,12

TABLEAU 5.6 – Paramètres statistiques de SSS de toutes les années pendant la saison de l’hiver.

de 38,09 en 2009 (**Tableau 5.6**). La SSS présente de faibles valeurs dans : la partie équatoriale, une partie de la zone Nord-Ouest, et une partie de la zone Sud-Est. La valeur minimale atteinte est de 32,08 en 2011 (**Tableau 5.6**).

La corrélation globale entre la SSS et la latitude est de -0,2. Mais comme on l’a déjà mentionné, puisque les trajets sont très différents d’une année à une autre (et parfois dans une même année), cette valeur n’est pas très significative. En prenant l’exemple d’un trajet de l’année 2010 allant de 23,5°S à 23,5°N, le coefficient R calculé est de -0,50.

La corrélation entre la SSS et la latitude reste moins bonne que celle de la SST, cependant, on peut tout de même constater (**Figures 5.14c**) que la SSS diminue quand on se rapproche de la latitude 12°N, et qu’au Sud, elle commence à diminuer en dessous de 12°S ; pour les points de l’année 2005 situés dans la partie Sud-Est (**Figure 5.14c** et **5.14d**), et en dessous de 18°S ; pour les points de 2005 situés dans la longitude 25°O, et aussi pour l’année 2014 (**Figure 5.14c** et **5.14d**).

Ce qu’on peut remarquer aussi dans ces figures, c’est que la SSS chute dans la zone située entre 6-12 °N et 50-70°O.

On voit (**Figure 5.14d**) que la corrélation entre la SSS et la longitude est très faible. Le coefficient R global est de 0,24, et est de -0.5 pour le trajet de 2010 (entre 23,5°S et 23,5°N).

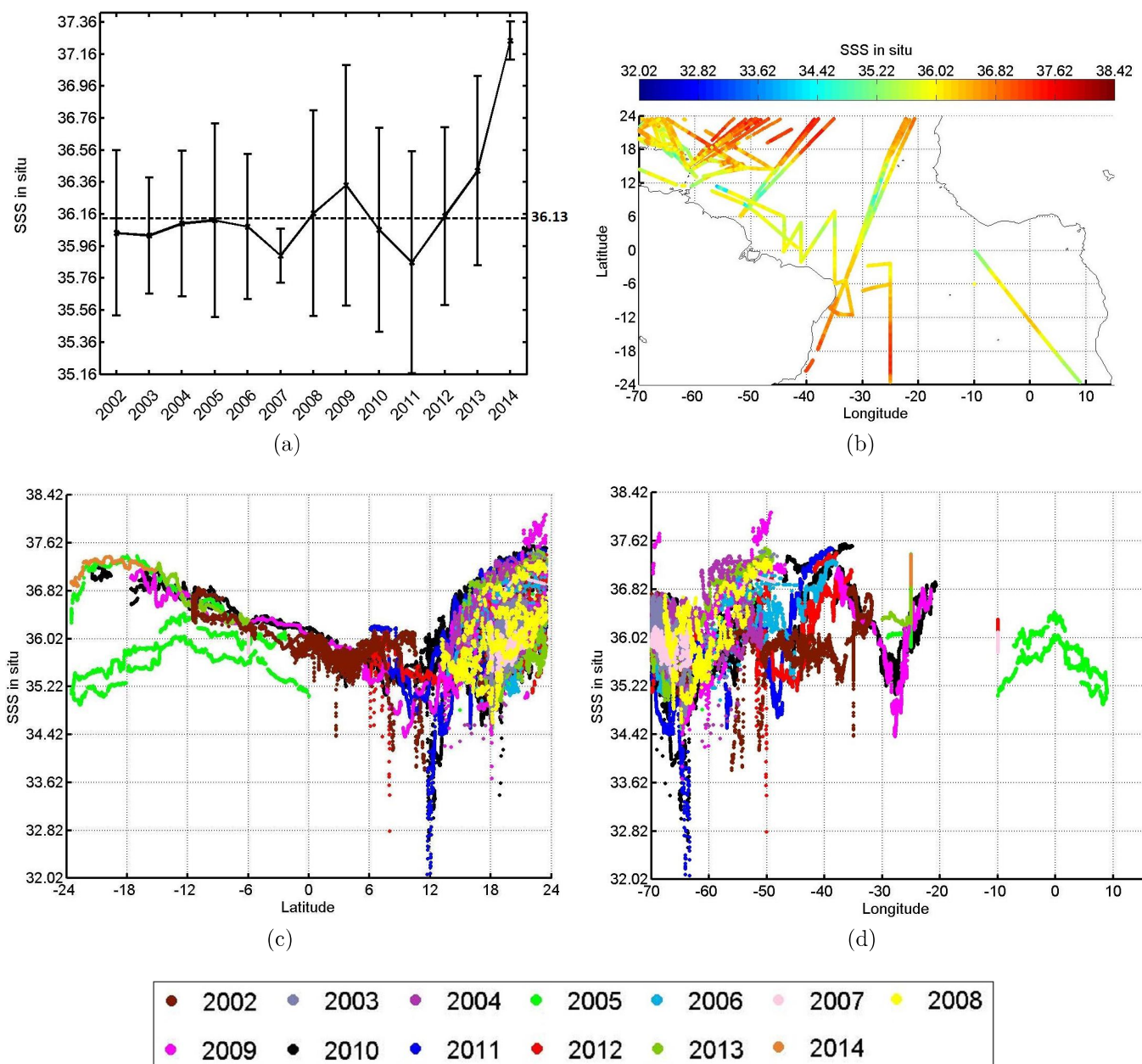


FIGURE 5.14 – Représentation de la variabilité de SSS *in situ* pendant la saison de l’hiver. (a) Moyenne et écarts-type de SSS de chaque année, (b) Distribution spatiale de SSS, (c) Variation de SSS en fonction de la latitude, (d) Variation de SSS en fonction de la longitude

fCO₂ : Entre 2002 et 2014, la fCO₂ présente beaucoup de variabilité d’une année à une autre (Figure 5.15a). Les moyennes de celle-ci varie de 348,76 à 415,16 μatm (Tableau 5.7). La tendance de la variabilité de fCO₂ reste la même que pour la saison du printemps. Comme illustrée dans la Figure 5.15b, la fCO₂ est plus élevée dans la partie Sud de la zone, où elle atteint un maximum de 449 μatm en 2010, et est plus faible dans la partie Nord, où la fCO₂ atteint une valeur de 320,34 μatm en 2010 (Tableau 5.7).

Année	Minimum (μatm)	Maximum (μatm)	Moyenne (μatm)	Écart-type (μatm)
2002	331,75	425,08	381,85	14,52
2003	340,74	415,27	356,17	10,32
2004	337,59	377,59	348,76	5,56
2005	328,67	432,74	379,17	27,19
2006	320,67	451,9	354,96	30,14
2007	341,69	448,1	379,72	38,22
2008	335,49	373,49	351,71	6,62
2009	326,49	418,68	362	21,34
2010	320,34	449	369,29	24,15
2011	330,89	446,1	369,62	26,33
2012	330,14	416,16	362,57	23,25
2013	343,48	428,71	387,64	22,53
2014	388,14	448,8	415,16	14,09

TABLEAU 5.7 – Paramètres statistiques de $f\text{CO}_2$ de toutes les années pendant la saison de l'hiver.

Contrairement à la SST et la SSS, la corrélation globale de $f\text{CO}_2$ est élevée à la fois avec la latitude ($R = -0,78$) et la longitude ($R = 0,68$). La **Figure 5.15c** montre que la $f\text{CO}_2$ augmente lorsqu'on se rapproche de 6°S , et la **Figure 5.15d** montre que la $f\text{CO}_2$ augmente d'Ouest en Est.

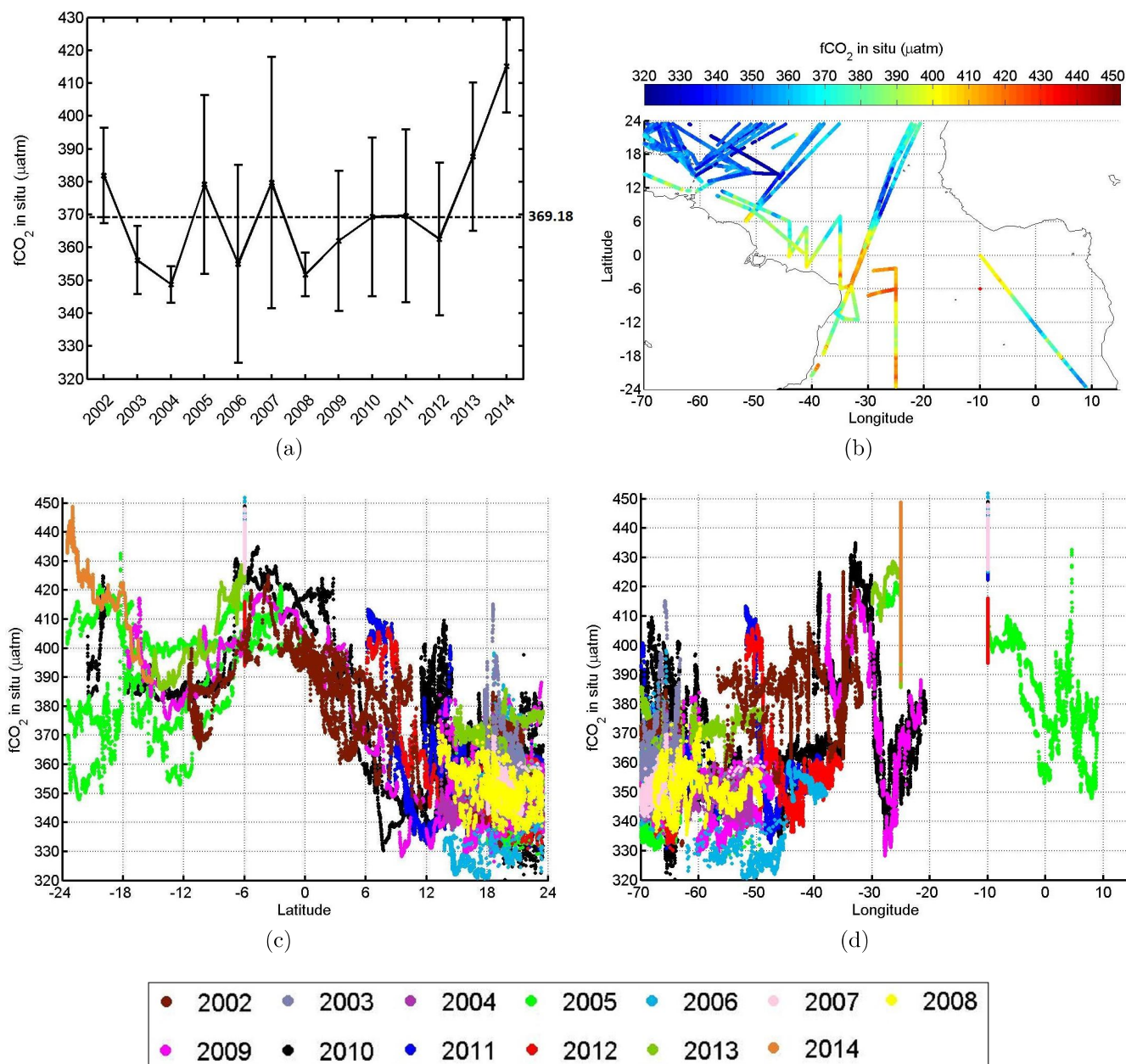


FIGURE 5.15 – Représentation de la variabilité de fCO_2 *in situ* pendant la saison de l'hiver. (a) Moyenne et écarts-type de fCO_2 de chaque année, (b) Distribution spatiale de fCO_2 , (c) Variation de fCO_2 en fonction de la latitude, (d) Variation de fCO_2 en fonction de la longitude

5.2 Données satellitaires : Chl-*a*, SST, et SSS

Les données satellitaires utilisées dans cette étude concernent les trois paramètres ; Chl-*a*, SST, et SSS. Les données de Chl-*a*, sont utilisées dans les étapes d'apprentissage, de validation, et de test des RNs, ainsi que pour la création des cartes de fCO_2 . Les données de SST et de SSS sont utilisées à la fois dans la partie test des RNs, et dans la création des cartes de fCO_2 .

Afin d'utiliser ces données satellitaires en complément à la BD *in situ* (pour la construction des RNs), en raison du manque de points contenant des données pour les résolutions temporelles réduites (1 jour, 3 jours, 8 jours ...), on a choisi d'utiliser les données mensuelles.

5.2.1 Chl-*a*

Comme déjà mentionné dans le **Chapitre 1 (Section 1.5.2)**, les mécanismes biologiques ont une grande influence sur la variabilité de $f\text{CO}_2$, d'où l'importance d'intégrer le paramètre Chl-*a* aux modèles. Cependant, étant donné que la BD *in situ* ne comporte pas de données de Chl-*a*, on a choisi de compléter cette BD par les données satellitaires de MODIS-A (avec une résolution spatiale de $4 \text{ km} \times 4 \text{ km}$; <http://oceandata.sci.gsfc.nasa.gov/MODIS-Aqua>, **Section 3.5.1**) en ne prenant que les points dont les positions géographiques correspondent à celles de la BD *in situ*. Puisque le capteur MODIS-A offre des données à partir de juillet 2002, pour les mois antérieurs à cette date, on a utilisé les données du capteur SeaWiFS (avec une résolution de $9 \text{ km} \times 9 \text{ km}$; <http://oceandata.sci.gsfc.nasa.gov/SeaWiFS>, **Section 3.5.1**).

Un extrait de Chl-*a* des quatre saisons de l'année 2012/2013 est illustré dans la **Figure 5.16**. On remarque dans cette figure que les régions où la Chl-*a* augmente concernent principalement : 1) les zones côtières ; où l'on trouve les décharges des trois plus grands fleuves dans le monde : l'Amazone et l'Orénoque à l'Est, et le fleuve du Congo à l'Ouest, on voit bien dans la figure que durant toute l'année, la Chl-*a* est forte dans ces régions. 2) les zones d'upwelling ; l'upwelling côtier au Nord et au Sud de la côte Africaine, et l'upwelling équatorial qui est assez remarquable dès le mois de juin dans les **Figures 5.16g, 5.16h, 5.16i, 5.16j, 5.16k, et 5.16l** (voir **Section 2.6** pour plus de détail sur la variabilité de la Chl-*a*).

La Chl-*a* a la particularité d'avoir une distribution asymétrique, où les données sont beaucoup plus concentrées dans les petites valeurs (voir **Figure 5.17a**). Étant donné que ceci peut affecter l'apprentissage des RNs, dans le sens où le RN résultant sera beaucoup plus optimisé pour les faibles Chl-*a*, et aura donc une faible performance quant aux larges Chl-*a*, une transformation Log_{10} a été appliquée sur les données de Chl-*a* avant de passer à l'étape de construction des RNs (voir **Figure 5.17b**).

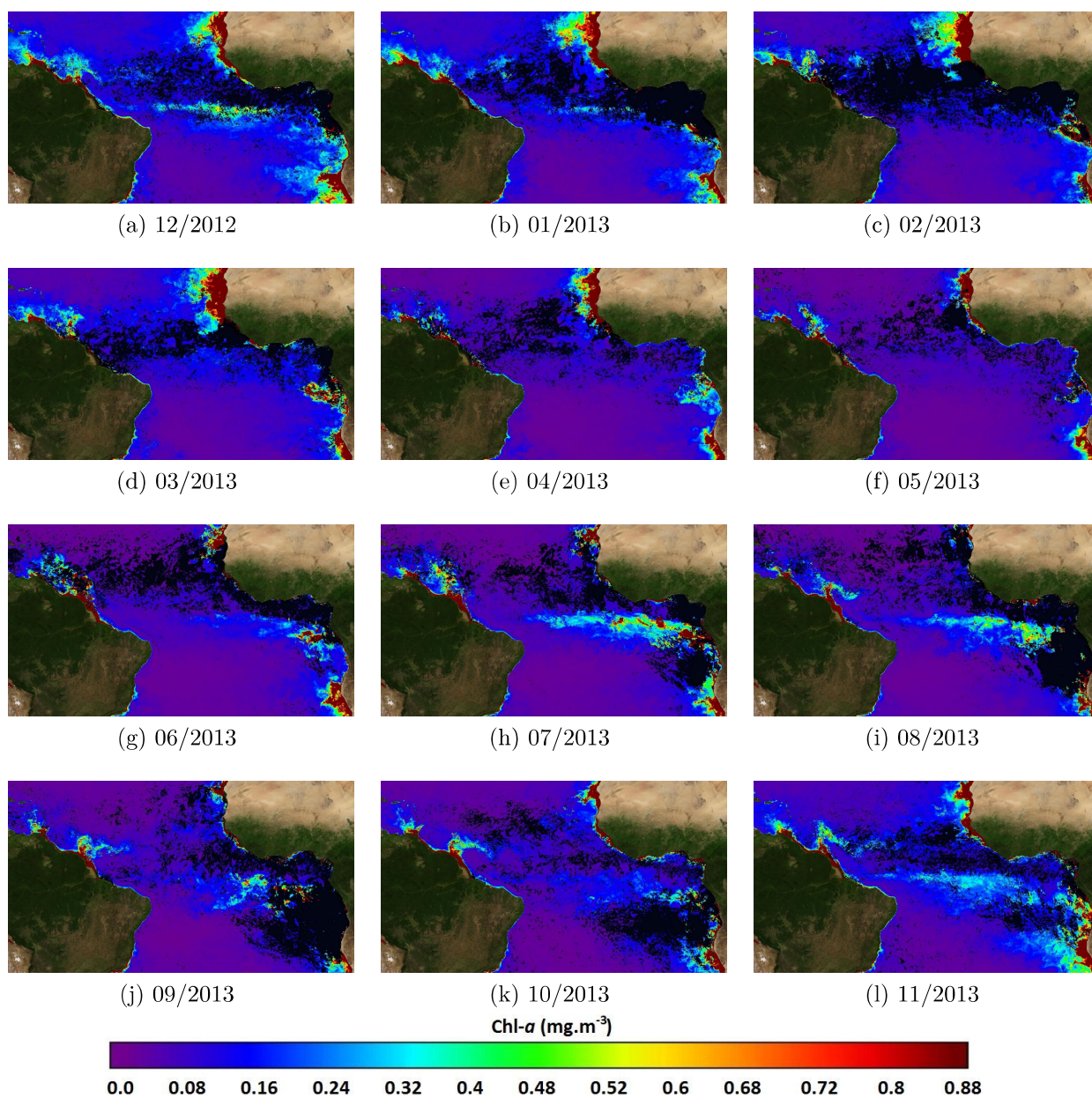


FIGURE 5.16 – Données satellitaires de Chl-*a* mensuelles, fournies par le capteur **MODIS-A**, dans la zone 23,5°S-70°O à 23,5°N-15°E, entre décembre 2012 et novembre 2013, avec une résolution spatiale de 4 km × 4 km.

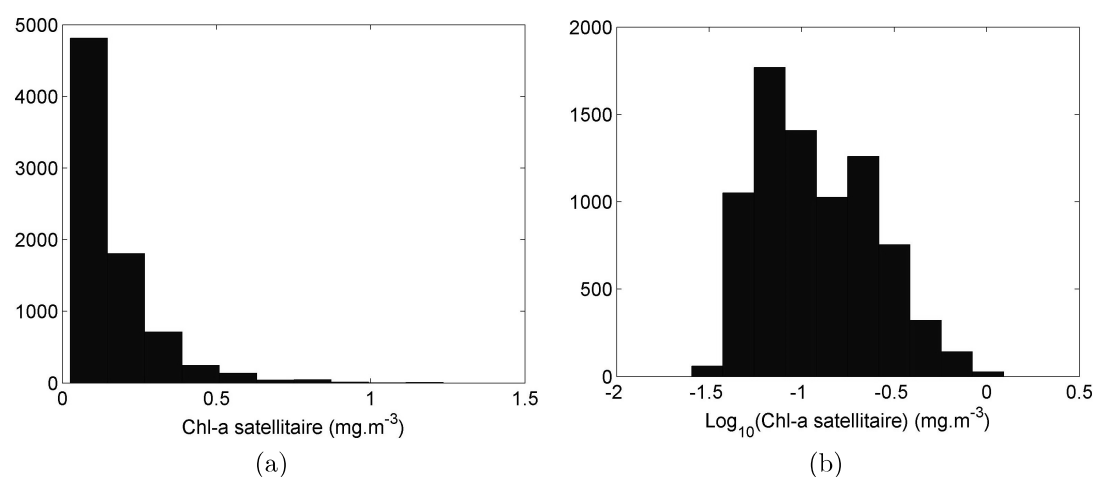


FIGURE 5.17 – Histogramme des données de Chl-*a* extraites de la BD de la saison du printemps. (a) Avant application de la transformation 'Log10', (b) Après application de la transformation 'Log10'

5.2.1.1 Variabilité de Chl-*a*

Printemps : Entre 2001 et 2013, les moyennes de Chl-*a* (en Log₁₀) varient de -1,28 à -1,09 (Tableau 5.8), autour d'une moyenne globale de l'ordre de -1,13 (Figure 5.18a).

Année	Minimum (mg.m ⁻³)	Maximum (mg.m ⁻³)	Moyenne (mg.m ⁻³)	Écart-type (mg.m ⁻³)
2001	-1,75	-0,56	-1,11	0,27
2002	-2,02	-0,45	-1,13	0,2
2003	-2,13	-0,53	-1,18	0,25
2004	-1,95	0,95	-1,09	0,37
2005	-2,46	-0,26	-1,25	0,19
2006	-2,08	0,38	-1,1	0,26
2007	-1,89	-0,21	-1,13	0,25
2008	-1,74	-0,08	-1,11	0,28
2009	-2,43	-0,02	-1,11	0,33
2010	-1,83	0,14	-1,12	0,2
2011	-1,73	-0,43	-1,21	0,19
2012	-2,21	-0,19	-1,14	0,29
2013	-1,88	-0,72	-1,28	0,21

TABEAU 5.8 – Paramètres statistiques de Log₁₀(Chl-*a* satellitaire) de toutes les années pendant la saison du printemps.

Le coefficient R global entre Chl-*a* et la latitude est de 0,04, et est de 0,33 avec la longitude. Cependant, comme pour les paramètres *in situ*, la faible corrélation avec la latitude est due à la différence qui se trouve entre les trajets de toutes les années. En

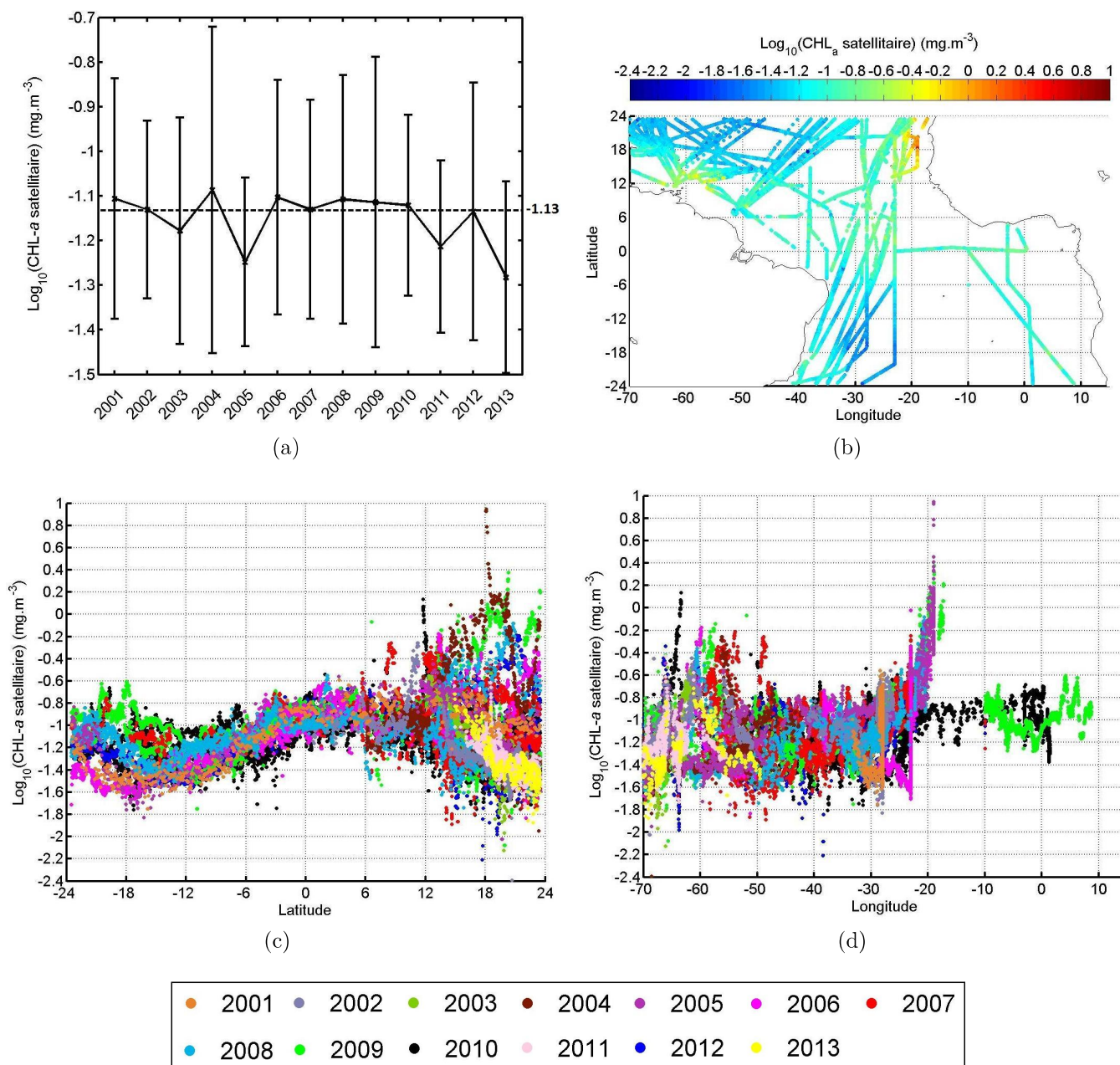


FIGURE 5.18 – Représentation de la variabilité de $\text{Log}_{10}(\text{Chl-a satellite})$ pendant la saison du printemps. (a) Moyenne et écarts-type de Chl-a de chaque année, (b) Distribution spatiale de Chl-a, (c) Variation de Chl-a en fonction de la latitude, (d) Variation de Chl-a en fonction de la longitude

prenant le cas de l'année 2001, le coefficient R est de 0,76 avec la latitude, mais reste un peu faible avec la longitude ; 0,41.

La relation entre la Chl-a et la latitude est bien visible dans la **Figure 5.18c**, où l'on remarque que la Chl-a augmente lorsqu'on se rapproche de la latitude 0°. On remarque aussi des pics de Chl-a au Nord de $\sim 10^\circ\text{N}$, dans les zones qui correspondent à la longitude 20°O et à l'ouest de la longitude 50°O (**Figure 5.18d**).

Les paramètres statistiques de Chl-a sans l'application du Log_{10} sont présentés dans

le **Tableau 5.9**.

Année	Minimum (mg.m ⁻³)	Maximum (mg.m ⁻³)	Moyenne (mg.m ⁻³)	Écart-type (mg.m ⁻³)
2001	0,02	0,28	0,09	0,05
2002	0,01	0,36	0,08	0,05
2003	0,01	0,30	0,08	0,07
2004	0,01	8,86	0,15	0,36
2005	0,003	0,56	0,06	0,05
2006	0,01	2,37	0,11	0,15
2007	0,01	0,61	0,09	0,07
2008	0,02	0,83	0,10	0,10
2009	0,004	0,95	0,10	0,09
2010	0,02	1,37	0,09	0,05
2011	0,02	0,37	0,07	0,04
2012	0,01	0,65	0,10	0,09
2013	0,01	0,19	0,06	0,03

TABLEAU 5.9 – Paramètres statistiques de Chl-*a* satellitaire de toutes les années pendant la saison du printemps.

Hiver : Entre 2002 et 2014, les moyennes de Chl-*a* (en Log₁₀) varient de -1,59 à -0,98 (**Tableau 5.10**), autour d'une moyenne globale de l'ordre de -1,14 (**Figure 5.19a**).

Année	Minimum (mg.m ⁻³)	Maximum (mg.m ⁻³)	Moyenne (mg.m ⁻³)	Écart-type (mg.m ⁻³)
2002	-1,74	-0,26	-0,98	0,21
2003	-1,48	-0,54	-1,17	0,11
2004	-1,4	-0,33	-1,18	0,14
2005	-1,87	-0,32	-1,13	0,33
2006	-1,4	-0,56	-1,11	0,1
2007	-1,41	-0,51	-1,19	0,12
2008	-1,41	-0,67	-1,16	0,11
2009	-1,74	-0,52	-1,17	0,17
2010	-1,99	0,06	-1,07	0,21
2011	-1,59	0,73	-1,05	0,24
2012	-1,47	0,49	-1,16	0,13
2013	-1,81	-0,45	-1,29	0,2
2014	-1,84	-1,39	-1,59	0,09

TABLEAU 5.10 – Paramètres statistiques de Log₁₀(Chl-*a* satellitaire) de toutes les années pendant la saison de l'hiver.

Le coefficient R global entre Chl-*a* et la latitude est de 0.01, et est de 0.003 avec la longitude. De même que pour les paramètres *in situ* décrits auparavant, en prenant le cas d'un trajet de l'année 2009 allant de 17°S à 23.5°N, le coefficient R est de 0.60 avec la latitude ainsi qu'avec la longitude.

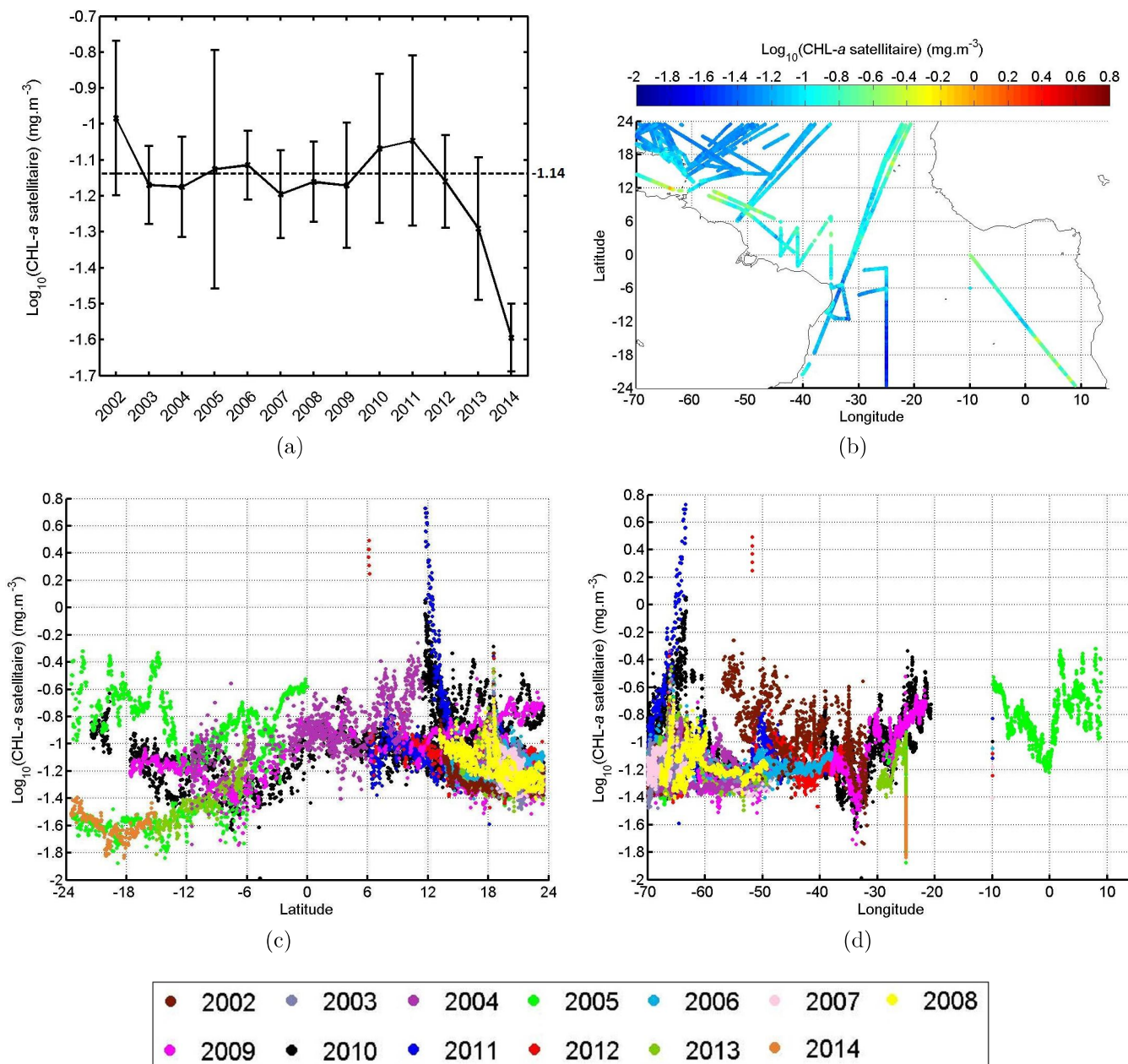


FIGURE 5.19 – Représentation de la variabilité de $\text{Log}_{10}(\text{Chl-}a \text{ satellitaire})$ pendant la saison de l'hiver. (a) Moyenne et écarts-type de Chl-*a* de chaque année, (b) Distribution spatiale de Chl-*a*, (c) Variation de Chl-*a* en fonction de la latitude, (d) Variation de Chl-*a* en fonction de la longitude

D'après les trois **Figures 5.19b**, **5.19c** et **5.19d**, on remarque que la Chl-*a* augmente à l'Ouest, au niveau de la latitude 6°N et 12°N, ce qui correspond à une zone où l'eau se mélange avec celle du fleuve Amazone.

Le **Tableau 5.11** contient les paramètres statistiques de Chl-*a* sans l'application de la transformation Log_{10} .

Année	Minimum (mg.m^{-3})	Maximum (mg.m^{-3})	Moyenne (mg.m^{-3})	Écart-type (mg.m^{-3})
2002	0,02	0,55	0,12	0,06
2003	0,03	0,29	0,07	0,02
2004	0,04	0,46	0,07	0,03
2005	0,01	0,48	0,1	0,08
2006	0,04	0,27	0,08	0,02
2007	0,04	0,31	0,07	0,02
2008	0,04	0,21	0,07	0,02
2009	0,02	0,3	0,07	0,04
2010	0,01	1,14	0,1	0,08
2011	0,03	5,35	0,13	0,31
2012	0,03	3,12	0,08	0,11
2013	0,02	0,35	0,06	0,03
2014	0,01	0,04	0,03	0,01

TABLEAU 5.11 – Paramètres statistiques de Chl-*a* satellitaire de toutes les années pendant la saison de l'hiver.

5.2.2 SST

Les données de SST utilisées sont aussi de type mensuel (pour avoir un maximum de points) fournies par le capteur MODIS-A (<http://oceandata.sci.gsfc.nasa.gov/MODIS-Aqua>), et sont disponibles depuis juillet 2002 (jusqu'à aujourd'hui), avec une résolution de $4 \text{ km} \times 4 \text{ km}$ (**Section 3.5.2**). Un extrait des quatre saisons de l'année 2012/2013 est illustré dans la **Figure 5.20**. On remarque dans cette figure que durant toute l'année, les deux parties les plus froides sont le Nord-Est et le Sud-Est. On peut constater qu'il y a une contrebalance entre ces deux zones; le Nord-Est est plus froid entre l'hiver et le printemps, tandis que le Sud-Est est plus froid entre l'été et l'automne. Pendant la saison du printemps, la partie Sud est beaucoup plus chaude que la partie Nord, et commence à se refroidir en été, où une langue d'eau froide commence à se former à l'Est du bassin dès le mois de juin (**Figures 5.20g, 5.20h, et 5.20i**; pour plus de détail sur cette langue d'eau froide, voir **Section 2.4.1**). Pendant cette saison (été), il y a un réchauffement de la partie Nord du bassin. Ce comportement est quasiment le même pour toutes les autres années.

Le code MATLAB utilisé pour la lecture et la structuration des fichiers satellitaires de Chl-*a* et de SST (qui sont de type HDF; **Section 3.4**) est présenté dans l'**Annexe A.5.1**.

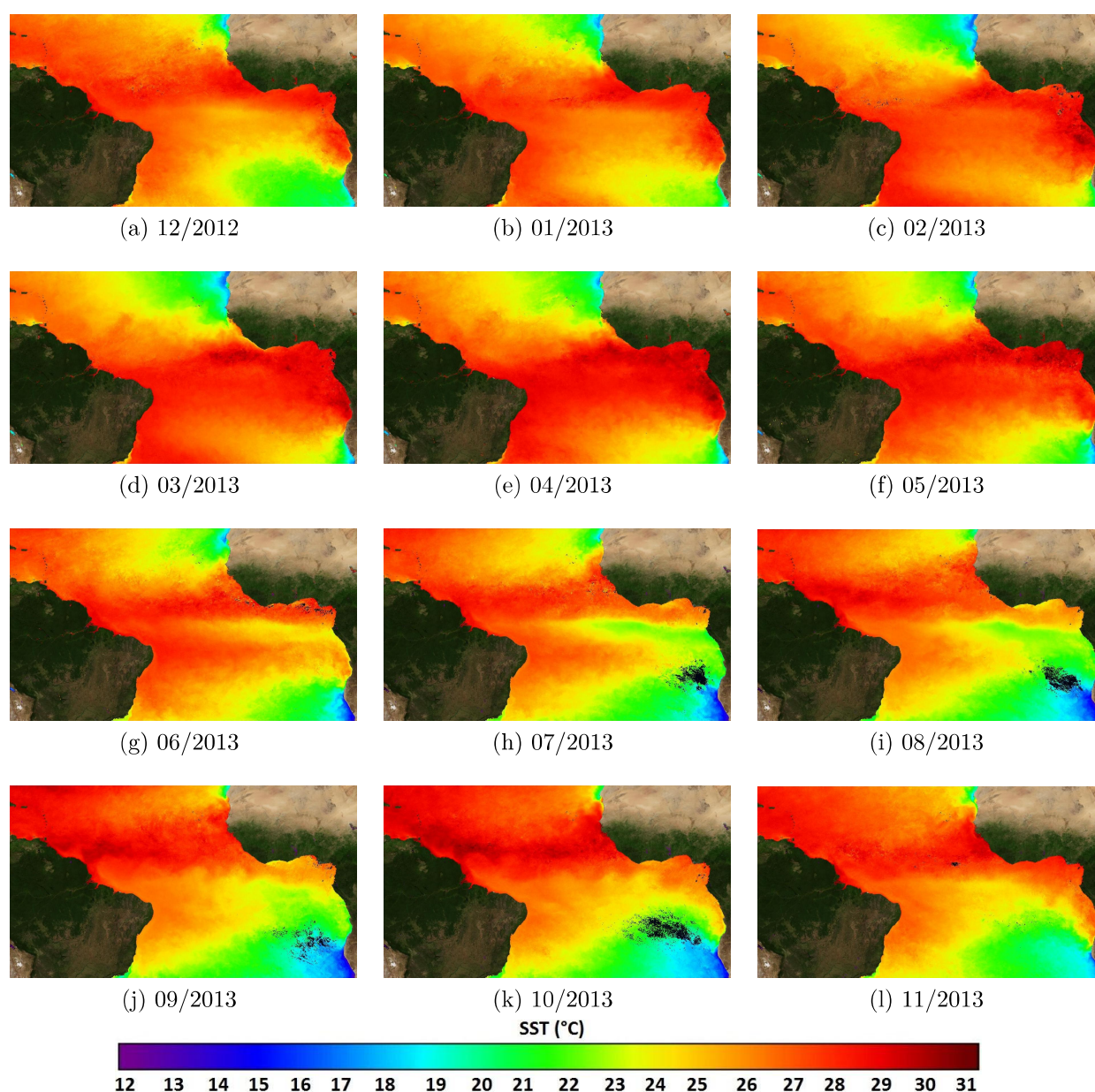


FIGURE 5.20 – Données satellitaires de SST mensuelles, fournies par le capteur **MODIS-A**, dans la zone $23,5^{\circ}\text{S}$ - 70°E à $23,5^{\circ}\text{N}$ - 15°E , entre décembre 2012 et novembre 2013, avec une résolution spatiale de $4\text{ km} \times 4\text{ km}$.

5.2.3 SSS

Les données satellitaires de SSS utilisées sont de types mensuelles, fournies par le capteur SMOS (<http://www.smos-bec.icm.csic.es/>) et disponibles depuis janvier 2010 jusqu'à décembre 2013 sur le site-web : <http://cp34-bec.cmima.csic.es/>. Un extrait des quatre saisons de l'année 2012/2013 est présenté dans la **Figure 5.21**. On constate que tout au long de l'année, la SSS est plus forte au Nord et au Sud de la zone d'étude, et est plus faible dans la zone centrale et les zone côtières.

Les données SSS de toutes les années ont été téléchargées en un seul fichier de type

NetCDF, d'une taille de 759 Mo. L'extraction ainsi que la structuration des données de ce fichier à été faite suivant le code MATLAB de l'Annexe A.7).

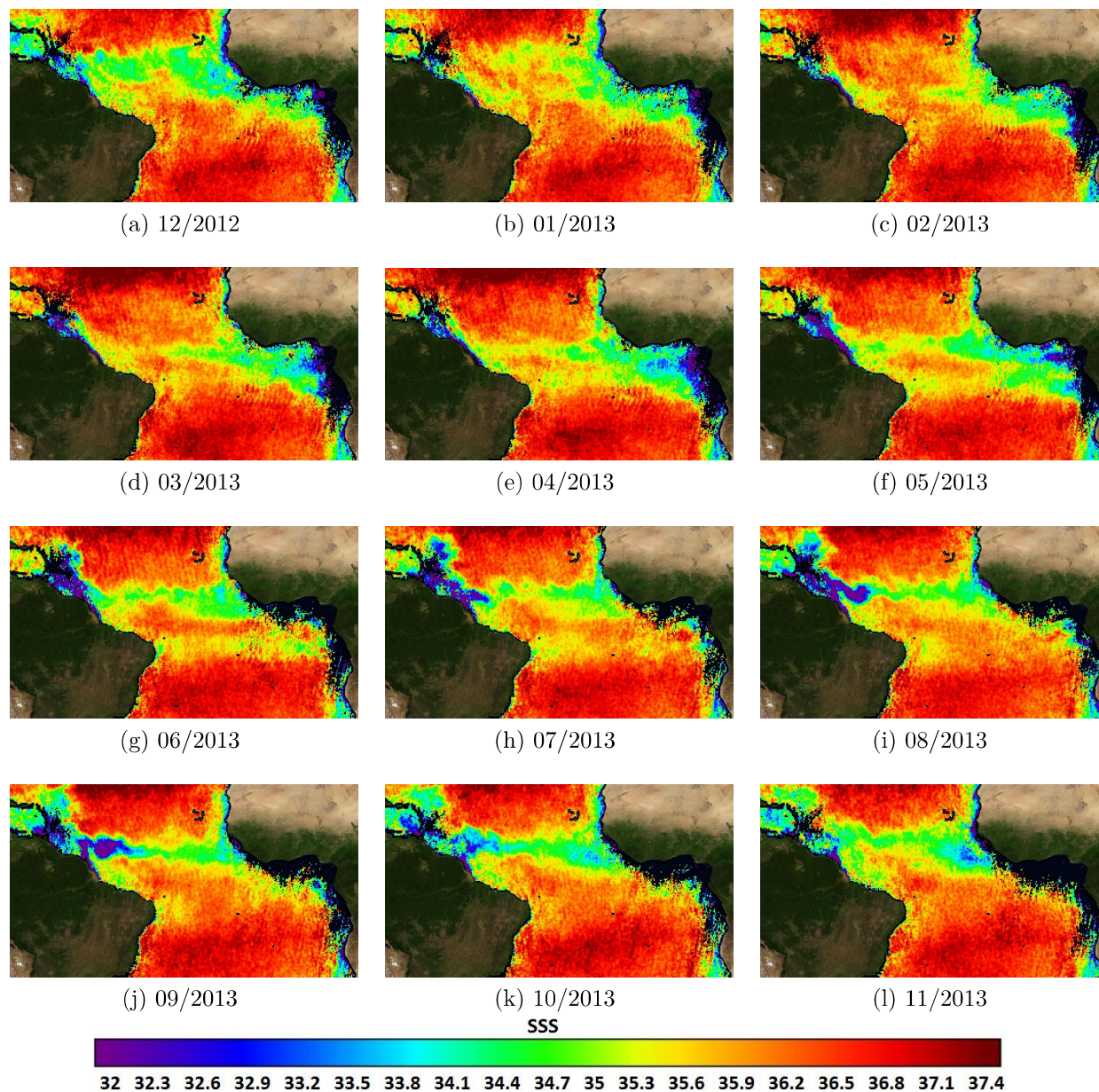


FIGURE 5.21 – Données satellitaires de SSS mensuelles, fournies par le capteur **SMOS**, dans la zone $23,5^{\circ}\text{S}$ - 70°O à $23,5^{\circ}\text{N}$ - 15°E , entre décembre 2012 et novembre 2013, avec une résolution spatiale de $25\text{ km} \times 25\text{ km}$.

5.3 Données assimilées : SSS

Puisque les données satellitaires de SSS SMOS ne sont disponibles qu'à partir de 2010, et que, par ailleurs, les données *in situ* sont disponibles à partir de l'année 2000,

afin de permettre le test des RNs (qui est censé être réalisé par des données satellitaires), pour les années entre 2002 (car les données de SST et Chl-*a* satellitaires sont disponibles à partir de 2002) et 2009 on a utilisé des données de SSS qui se rapprochent le plus des données satellitaires, en terme de résolution spatiotemporelle et d'incertitude. Ces dernières sont des données de SODA V.2.2.4 (Simple Ocean Data Assimilation, [Carton et Giese, 2008]) offrant une résolution spatiale de 50 km × 50 km, et une incertitude d'environ 0,4. Ces données sont mises en lignes pour téléchargement via le serveur <http://sodaserver.tamu.edu/> (les fichiers sont de types NetCDF).

Le projet SODA a été initié pendant les années 1990 [Carton et Hackert, 1989, 1990; Raghunath et Carton, 1990], suite à une collaboration entre le "Département des sciences atmosphériques et océaniques" de l'"Université de Maryland", et le "Département d'océanographie" de l'"Université de Texas A&M". Depuis sa création, plusieurs versions ont été réalisées, basées sur différents modèles océanographiques (voir <http://www.atmos.umd.edu/~ocean/history.html> pour plus de détail sur l'historique de la création des versions de SODA).

Le code MATLAB utilisé pour structurer ces données est présenté dans l'**Annexe A.8**.

5.4 Regroupement des données

Après le prétraitement des données *in situ*, le téléchargement et la structuration des données satellitaires et des données assimilées, une matrice MATLAB globale (de type "Cell") a été créée afin de regrouper toutes ces données. Cette matrice représente la "BD de construction des RNs".

Pour ce faire, un seul principe a été utilisé pour tous les types de données ; les coordonnées géographiques *in situ* ont été utilisées comme références, ainsi, pour chaque point *in situ*, on cherche si le point satellitaire (ou le point de la donnée assimilée) correspondant existe, en calculant la distance euclidienne entre les deux points. Si ce point existe, on combine les deux points dans cette nouvelle matrice.

Pour les données satellitaires de Chl-*a* et de SST, ces dernières ont d'abord été combinées ensembles en utilisant le code MATLAB présenté dans l'**Annexe A.6**, après cela, le code MATLAB de l' **Annexe A.9** a été utilisé pour intégrer ces données à la matrice. Un extrait de cette matrice est illustré dans la **Figure 5.22**.

Le code MATLAB utilisé pour les données de SSS SODA est présenté dans l'**Annexe A.10**, et celui utilisé pour les données de SSS SMOS est présenté dans l'**Annexe A.11**.

Puisque les données satellitaires de Chl-*a* sont utilisées dans l'étape de construction des RNs, et vu qu'il y a plusieurs endroits dans cette zone d'étude où les données de Chl-*a* sont inexistantes (points noirs dans la **Figure 5.16**), l'intégration de ces données entraîne une perte de certains points *in situ* de la BD de construction des RNs. En effet, comme on peut voir dans le **Tableau 5.12**, un taux de 14% a été perdu sur l'ensemble des données *in situ*.

148x3 cell

	1	2	3
1	'year'	'month'	'data'
2	'2000'	'10'	3093x22 double
3	'2000'	'11'	2204x22 double
4	'2001'	'03'	753x22 double
5	'2001'	'04'	381x22 double
6	'2001'	'11'	1833x22 double
7	'2002'	'01'	813x22 double
8	'2002'	'02'	4269x22 double
9	'2002'	'03'	4270x22 double
10	'2002'	'04'	1541x22 double
11	'2002'	'05'	650x22 double
12	'2002'	'06'	1771x22 double
13	'2002'	'07'	1085x22 double
14	'2002'	'08'	418x22 double

(a)

Données *in situ*

Année	Mois	Jour	Heure	Minute	Bathy.	Latitude	Longitude	Dist. mass.	SSS	SST	fCO ₂

...

Chl- <i>a</i> /SST sat				SSS SODA			SSS SMOS		
Latitude	Longitude	Chl- <i>a</i>	SST	Latitude	Longitude	SSS	Longitude	Latitude	SSS

...

(b)

FIGURE 5.22 – Extrait de la matrice MATLAB représentant la BD de construction des RNs. (a) matrice globale, (b) détail de chaque cellule de la colonne 'data' dans (a). Bathy. : Bathymétrie, Dist. mass. : Distance à la masse terrestre

Proportion de données perdues (%)	
Hiver	2,9
Printemps	12,6
Été	24,6
Automne	13,3
14	

TABLEAU 5.12 – Les proportions des données *in situ* perdues après l'utilisation des données satellitaires de Chl-*a* pour la construction des RNs.

5.5 Classification des données

Le choix de passer par une étape de classification de la zone d'étude, a été fait en raison de la très grande surface qu'occupe cette zone, et donc (comme vu dans les deux **Sections 5.1.1.1** et **5.1.1.2**) des fortes variabilités des paramètres, particulièrement la SST et la $f\text{CO}_2$. Ainsi, cette zone a été classifiée en se basant sur les données satellitaires de SST, ceci afin de limiter la variabilité dans chaque classe.

Le choix du paramètre SST a été fait pour deux raisons : 1) la SST est le paramètre le plus influent sur la $f\text{CO}_2$ dans l'océan Atlantique tropical, et 2) les données satellitaires de SST sont disponibles sur une grande échelle spatiotemporelle. La description des données utilisées a été faite dans la **Section 5.2.2**.

La description de la méthode utilisée pour la classification (*K-means*) est présentée dans la **Section 4.3.1**.

5.5.1 Données satellitaire : SST

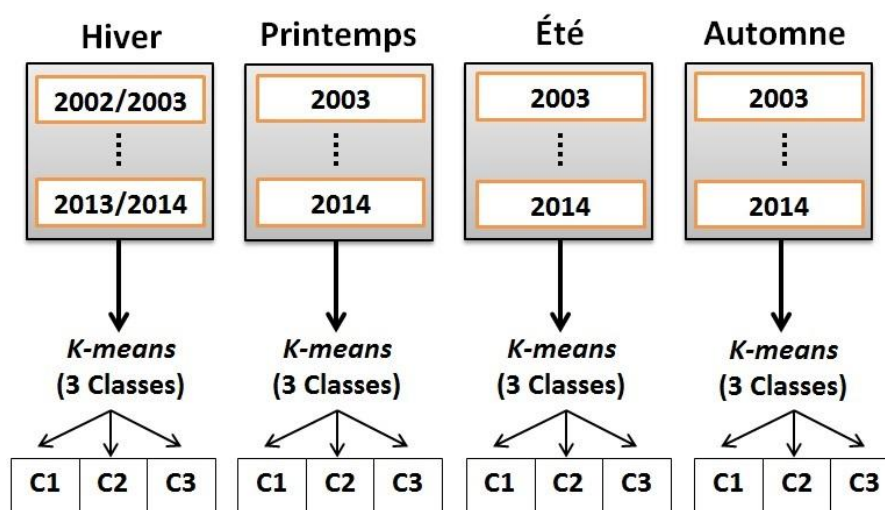
Les données satellitaires utilisées sont comprises entre l'hiver 2002/2003 et l'automne 2014. La lecture et la structuration des fichiers satellitaires (de type HDF) en utilisant le code MATLAB de l'**Annexe A.5.1**, produit un nombre de matrices égale à 145, comportant un nombre total de 204937791 points de données, réparties sur quatre ensembles ; hiver : 51338705 points, printemps : 51297346 points, été : 51137508 points, et automne : 51164232 points. Le **Tableau 5.13** contient ; pour chaque saison, le nombre de points global ainsi que la moyenne et l'écart-type de SST. On remarque dans ce tableau que les écarts-types de la SST satellitaire augmentent d'hiver en automne, et les moyennes ne changent pas beaucoup d'une saison à une autre.

Saison	Nombre de points	Moyenne (°C)	Écart-type (°C)
Hiver	51338705	25,99	2,01
Printemps	51297346	26,43	2,21
Été	51137508	25,45	2,60
Automne	51164232	25,70	3,03

TABLEAU 5.13 – Moyennes et écarts-types de la SST satellitaire (de toutes les années ; 2002/2003 -> 2014) par saison.

La méthode *K-means* est ainsi appliquée séparément sur chaque ensemble de données (voir **Figure 5.23**). Le code MATLAB utilisé se trouve dans l'**Annexe A.12**. Pour chaque saison, le résultat obtenu de la classification, consiste en un ensemble de valeurs, représentant les centroïdes des trois classes, ainsi que l'intervalle de SST pour chaque classe (comme expliqué dans l'algorithme de cette méthode ; **Section 4.3.1**). Ces résultats sont présentés dans le **Tableau 5.14**.

Comparés aux écarts-types de SST avant la classification (**Tableau 5.13**), on remarque dans le **Tableau 5.14** que le choix de faire une classification en trois classes, fait qu'au sein d'une même classe, l'écart-type ait diminué. Néanmoins, on constate que pour les quatre saisons, l'écart-type de la classe 1 (C1) est plus élevé que les autres. Ceci s'explique par le fait que ces classes (C1) contiennent des points où la SST est très faible

FIGURE 5.23 – L'application de la méthode *K-means* sur chaque ensemble de données.

Saison	Classes de SST (°C)	NP*	Centroïde (°C)	Écart-type (°C)
Hiver	C1 : [11,11 - 23,83]	7508680	22,29	1,33
	C2 : [23,84 - 26,47]	19007434	25,39	0,74
	C3 : [26,48 - 34,65]	24822591	27,57	0,73
Printemps	C1 : [10,01 - 24,19]	8032093	22,59	1,48
	C2 : [24,20 - 27,01]	19429478	25,80	0,80
	C3 : [27,02 - 35,00]	23835775	28,24	0,80
Été	C1 : [10,04 - 22,57]	6704414	20,66	1,77
	C2 : [22,56 - 26,03]	20199401	24,49	0,98
	C3 : [26,04 - 34,64]	24233693	27,58	0,94
Automne	C1 : [10,01 - 22,46]	7754913	20,24	1,75
	C2 : [22,47 - 26,39]	18151197	24,69	1,12
	C3 : [26,40 - 34,94]	25258122	28,10	1,01

TABLEAU 5.14 – Récapitulatif des résultats de la classification des données satellitaires. Les écarts-types maximaux sont en rouge. *Nombre de points

(< 17°C) par rapport aux moyennes générales, et donc, les étendues sont plus larges par rapport aux autres classes (C2 et C3). Cependant, comme illustré dans les histogrammes de la **Figure 5.24**, les points où la SST est inférieure à 17°C sont très peu nombreux par rapport aux autres ; en hiver ces points représentent environ 0,03% de la totalité des données, et 0,2% des données de la première classe. Au printemps, ces points représentent 0,09% de la totalité des données, et environ 0,6% des données de la classe 1. En été, ils représentent 0,7% de la totalité des données, et environ 5% de la classe 1. Pendant l'automne, ces points représentent 0,8% de la totalité, et environ 6% de la classe 1.

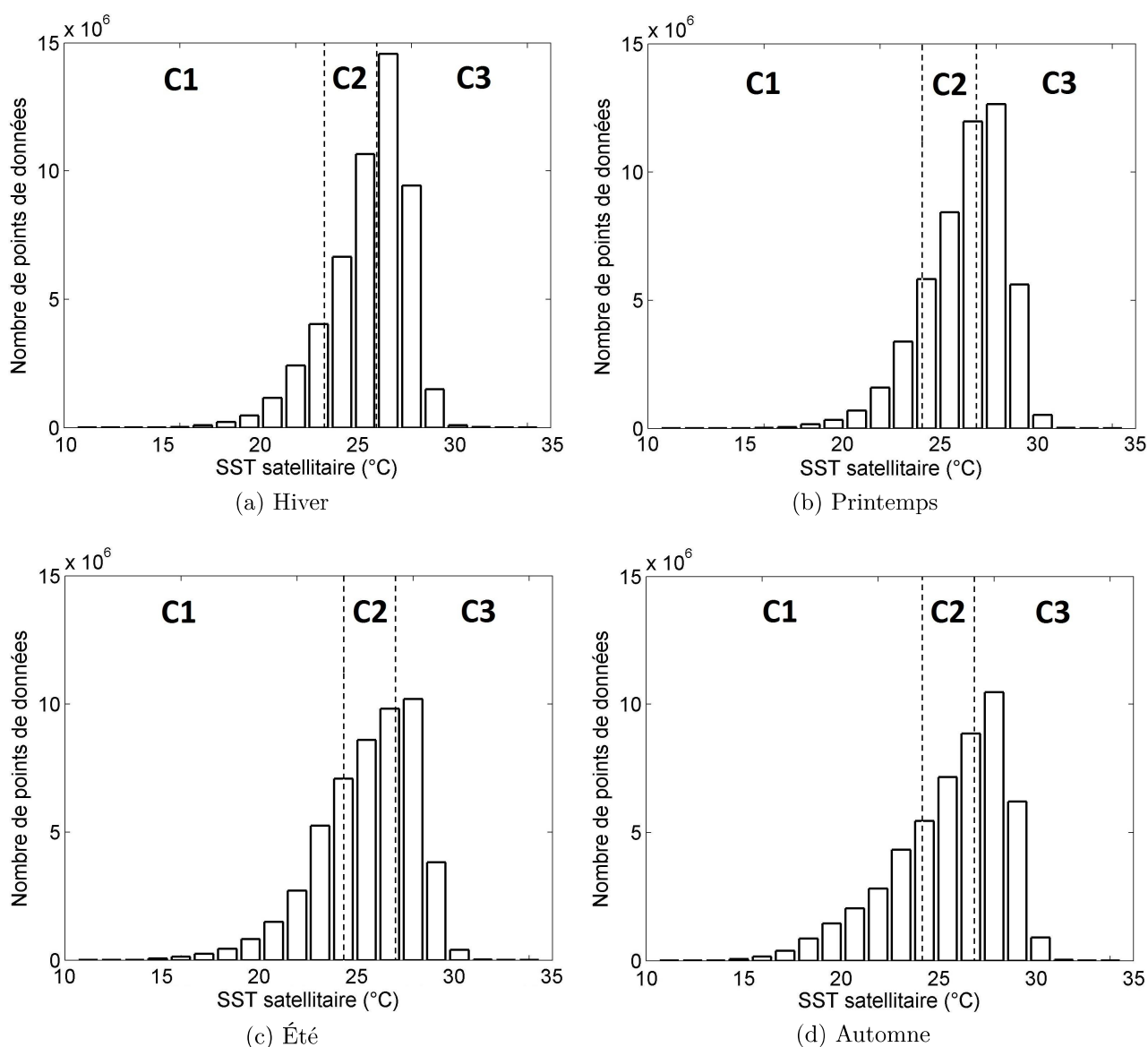


FIGURE 5.24 – Histogrammes de toutes les données de SST satellitaire par saison, avec des séparations entre les 3 classes (C1, C2, et C3).

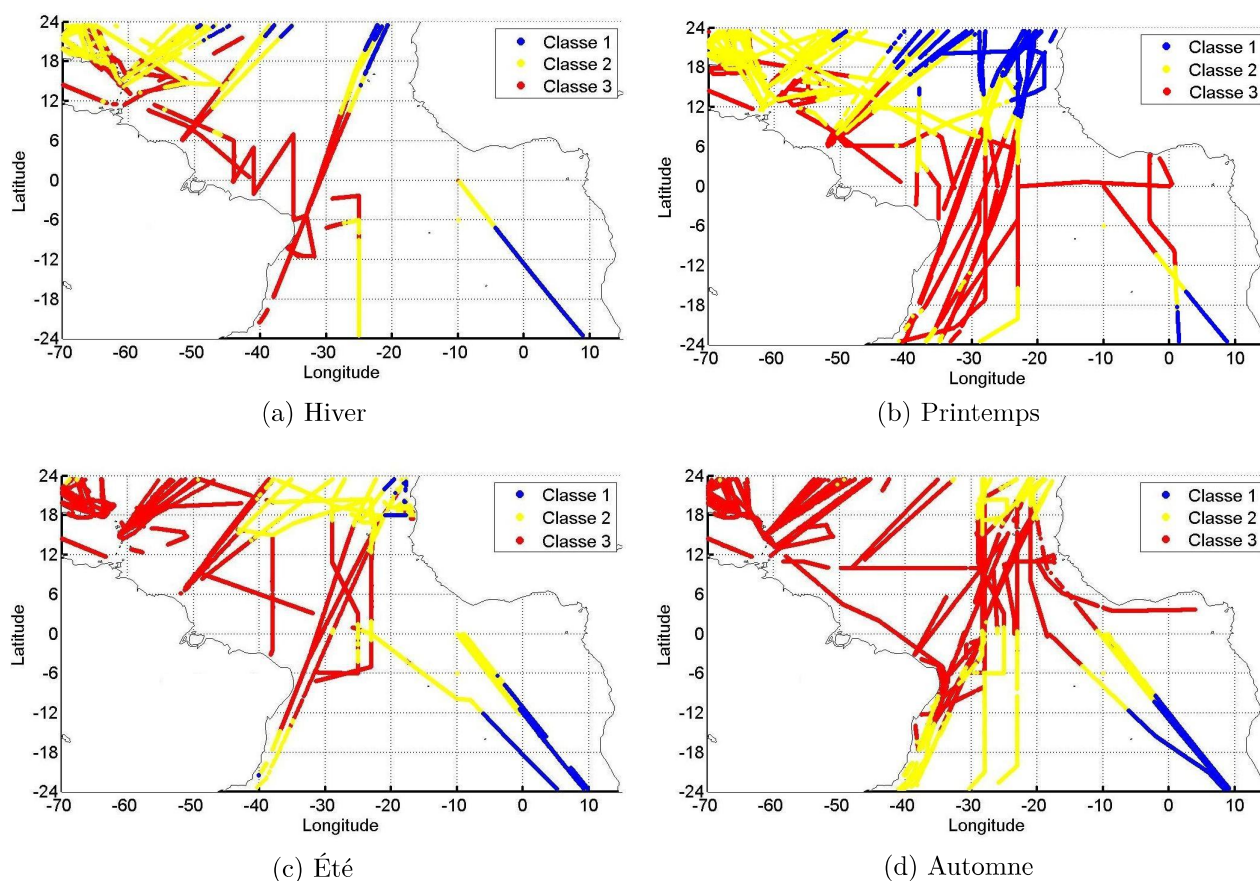
5.5.2 Données *in situ*

Les intervalles retrouvés lors de la classification de la SST satellitaire (**Section 5.5.1**) ont été appliqués pour partitionner les données *in situ* de chaque saison en trois classes (le code MATLAB utilisé se trouve dans l'**Annexe A.13**). Le nombre de données contenues dans chaque classe de chaque saison est présenté dans le **Tableau 5.15**. Ces données sont illustrées dans la **Figure 5.25**.

On remarque (**Tableau 5.15**, **Figure 5.25**) que pour les quatre saisons, la première classe est celle qui contient le moins de données, et la troisième classe est celle qui en contient le plus (comme pour la SST satellitaire, **Tableau 5.14**).

Saison	Classe	Nombre de points
Hiver	C1	3662
	C2	40249
	C3	48436
Printemps	C1	18603
	C2	69239
	C3	78872
Été	C1	6127
	C2	37743
	C3	89145
Automne	C1	8403
	C2	33044
	C3	115911

TABLEAU 5.15 – Nombre de points de données de chaque classe de chaque saison.

FIGURE 5.25 – Représentation spatiale des données *in situ* de chaque saison, divisées en trois classes.

Comme schématisé dans la **Figure 5.26**, à partir de la matrice globale de la BD de construction des RNs, 12 autres matrices ont été créées pour contenir les classes de toutes les saisons.

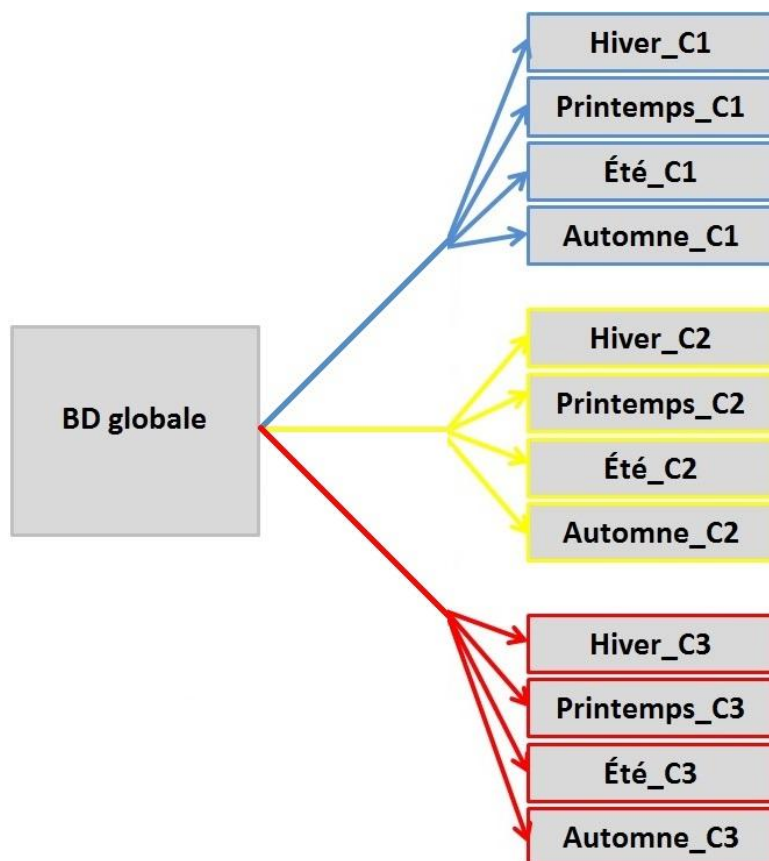


FIGURE 5.26 – L'étape de division de la BD de construction des RNs en 12 matrices.

Ces 12 matrices vont être utilisées par la suite, afin d'exécuter les 3 étapes de construction des RNs ; apprentissage, validation, et test.

Chapitre 6

INTERPOLATION DE LA FUGACITÉ DU CO₂

Ce chapitre contient toute la démarche suivie lors de l'interpolation de la fugacité du CO₂. Il est composé de deux principales sections. Dans la première (**6.1**), on parlera des étapes suivies pour la construction des réseaux de neurones (RNs). La deuxième (**6.2**) contiendra les résultats de l'interpolation de la fCO₂ pour les deux saisons printemps et hiver. Les résultats des deux autres saisons (été et automne) sont présentés dans l'**Annexe C**.

6.1 L'algorithme suivi

Dans le but de réaliser une interpolation de la fCO₂ avec une bonne performance, on a proposé dans cette étude une démarche qui permet de surveiller la capacité de généralisation du RN, au fur et à mesure du déroulement de l'apprentissage.

La **Figure 6.1** illustre le déroulement de l'algorithme d'interpolation de fCO₂ proposé dans cette étude. Celui-ci consiste en un ensemble de 20 itérations (IT.1, IT.2, ..., IT.20 dans la **Figure 6.1**), à chacune d'elles, le nombre de neurones cachés est incrémenté par 1, ce qui donne à la fin un ensemble de 20 RNs contenant un nombre de neurones cachés allant de 1 jusqu'à 20. À chaque itération, les trois étapes suivantes sont appliquées de façon séquentielle :

1. **Apprentissage** : comme expliqué dans la **Section 4.2.1.2**, c'est pendant cette étape que se fait la création du RN. Elle est basée sur les données *in situ*. La matrice utilisée pour l'apprentissage contient trois vecteurs correspondant aux trois paramètres ; SST, SSS, et Chl-*a*. Cependant, puisque pour la majorité des classes *in situ*, les données sont disponibles à la fois au Nord et au Sud de la zone d'étude, et que par ailleurs, on a vu (**Sections 5.1.1.1** et **5.1.1.2**) que la fCO₂ est plus forte au Sud de la zone Atlantique tropical, un quatrième paramètre a été ajouté à la BD d'apprentissage, représentant la position de chaque point *in situ* par rapport à la latitude 0°. Ce paramètre est présenté dans la **Figure 6.1** sous le nom de "**N. bin.**", ce qui signifie "**neurone binaire**" contenant la valeur 1 si le point *in situ* est situé au Nord, et 0 sinon.

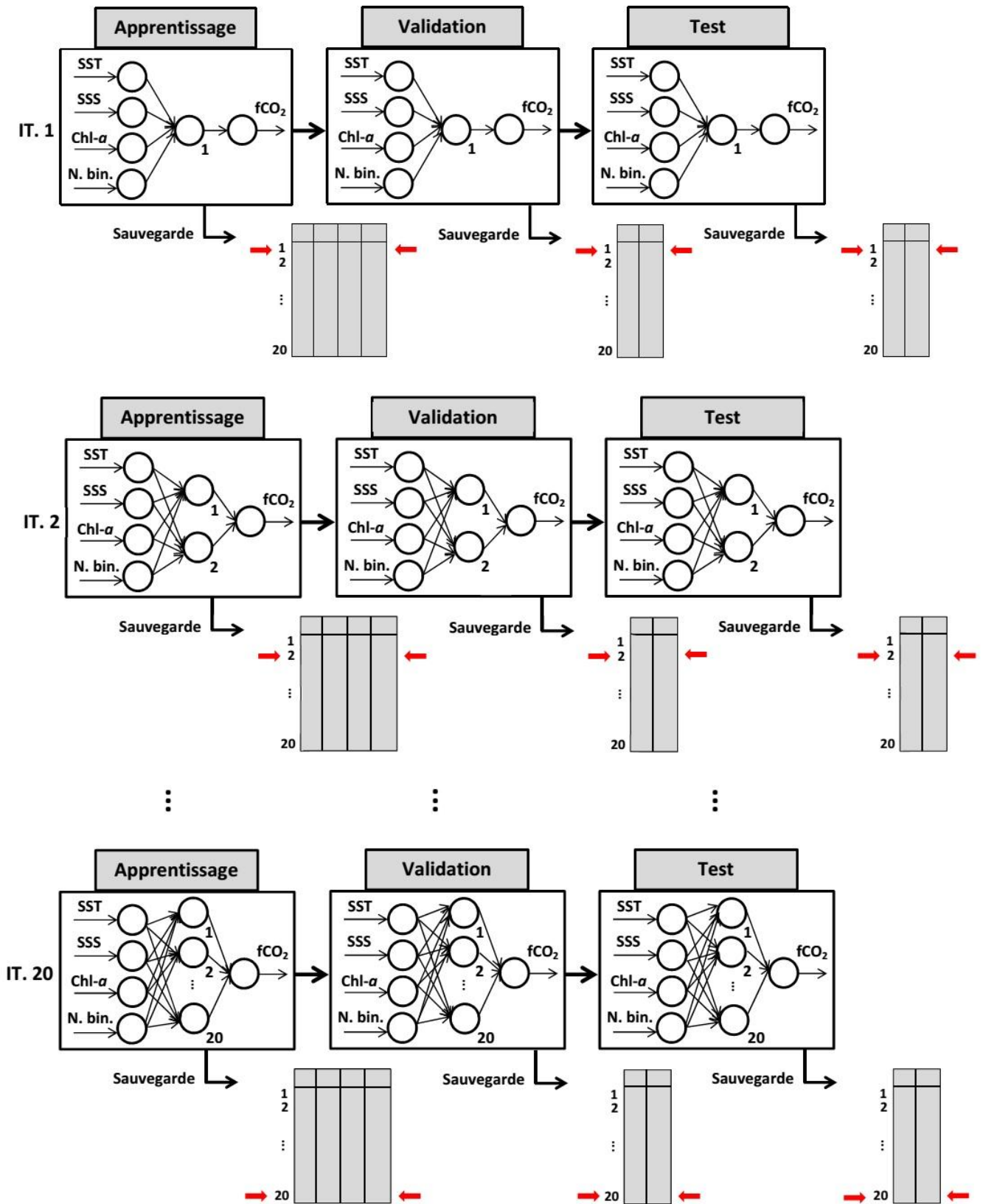


FIGURE 6.1 – Illustration de la 1e, la 2e, et la 20e itération (IT.) du déroulement de l’algorithme proposé pour choisir le RN optimal. N. bin. : neurone binaire

2. **Validation** : c'est la première validation du RN. Dans cette étape le RN est confronté à des données de la même nature que celles qu'il a vu pendant l'étape d'apprentissage, à savoir des données *in situ*.
3. **Test** : c'est considéré comme une deuxième validation. Cette étape est plus importante que la première validation, car elle permet de tester la capacité de généralisation du RN lorsqu'il est confronté à un nouveau set de données, différentes de celles vues pendant l'apprentissage, à savoir des données satellitaires.

Pour permettre le déroulement de ces 3 étapes, pour chaque BD créée lors de l'étape de division de la BD globale (**Section 5.5.2**), la partie contenant les données *in situ* (**Figure 5.22a**) est divisée en 2, de sorte que ~ 80 % des données sera utilisée pour l'apprentissage, et les ~ 20 % restante pour la validation. Pour l'étape du test, toute la partie contenant les données satellitaires et les données assimilées (**Figure 5.22b**) est utilisée. Pour cela, le code MATLAB de l'**Annexe A.14** a été appliqué.

Comme vu dans la **Figure 6.1**, après chaque étape de chaque itération, le résultat est sauvegardé dans un tableau créé au préalable. Ainsi, à la fin des 20 itérations, le résultat est présenté sous forme de 3 tableaux de 20 lignes (un pour l'apprentissage, un pour la validation, est un pour le test). Chacune de ces 20 lignes contient le résultat de l'itération correspondante. Le tableau de l'étape d'apprentissage contient les attributs suivants : le RN créé, toutes les informations correspondantes à cette même étape, la performance du RN (la valeur du RMSE), et enfin le biais que présentent la $f\text{CO}_2$ en sortie du RN. Pour les deux autres étapes, les tableaux contiennent seulement le RMSE ainsi que le biais.

Le code MATLAB permettant le déroulement des trois étapes citées (apprentissage, validation et test) est présenté dans l'**Annexe A.15**.

À la fin de toutes ces étapes, le RN optimal (celui ayant le RMSE minimal pendant le test) est choisi parmi l'ensemble des 20 RNs suivant le code MATLAB de la **Section A.15.1**.

Dans un premier lieu, l'approche expliquée ci-dessus a été validée sur une région limitée de l'océan Atlantique tropical (de 8°N - 61°O à 23°N - 20°O), en utilisant des données de la saison du printemps (mars et avril entre 2001 et 2009) extraites de la BD SOCAT V.1.4 (<http://www.socat.info/SOCATv1/index.html>) et combinées avec des données du projet EC-CARBOOCEAN (<http://dataportal.carboocean.org/>), ce qui a donné une BD *in situ* contenant 7188 points. Pour l'étape de test, on a utilisé des données de SSS interpolées à partir de la SST satellitaire et la latitude, en utilisant une régression linéaire multiple (MLR). Cette approche avait déjà été validée dans notre premier article scientifique "A comparison of multiple non-linear regression and neural network techniques for sea surface salinity estimation in the tropical Atlantic ocean based on satellite data" [Moussa *et al.*, 2015, **Annexe E.1**]. Ainsi, pour l'interpolation de la $f\text{CO}_2$, on a comparé deux méthodes : les RNs et la MLR, et on a montré les RNs donnent un meilleur résultat lors de l'étape de test, avec un RMSE de $8,7 \mu\text{atm}$ et un coefficient de détermination (R^2)¹ de 60% vs. $9,6 \mu\text{atm}$ et 50% pour la MLR. Les résultats de cette étude ont fait l'objet d'un deuxième article scientifique : "Satellite-derived CO_2 fugacity in surface seawater of the

1. Mesure de l'adéquation entre la variable en sortie (Y) et la variable en entrée X . Dans le cas d'une régression linéaire, c'est le carré du coefficient de corrélation R .

tropical Atlantic ocean using feedforward neural network" [Moussa *et al.*, 2016, **Annexe E.2**].

Dans cette étude, cette démarche a été appliquée sur la totalité de la zone de l'océan Atlantique tropical, en utilisant une BD *in situ* plus large que la première, à savoir la BD SOCAT V.3 (<http://www.socat.info/>).

6.2 Construction des réseaux de neurones

6.2.1 Printemps : Classe 1 (C1)

Pour cette première classe, le RN optimal est celui ayant 7 neurones cachés.

6.2.1.1 Apprentissage

Les résultats de l'étape d'apprentissage sont présentés dans le **Tableau 6.1** et la **Figure 6.2**. Un RMSE global de 10,54 μatm a été obtenu, avec un faible biais de 0,17 μatm , une corrélation (R) entre la fCO₂ en sortie du RN et la fCO₂ désirée de 0,80, et un coefficient de détermination (R²) de 64%.

Année	RMSE (μatm)	Biais (μatm)	NP *	PD (%) **
2001	30,94	-29,83	201	1,55
2002	40,13	-39,34	64	0,49
2004	13,08	-7,13	574	4,42
2006	10,49	-4,85	1526	11,75
2007	9,81	2,63	2437	18,77
2008	12,72	5,07	964	7,43
2009	5,72	1,56	1462	11,26
2010	8,61	2,29	2670	20,57
2012	8,89	0,82	3085	23,76
Total	10,54	0,17	12983	/

TABLEAU 6.1 – Apprentissage de la classe C1 de la saison du printemps : RMSE et biais correspondants à chaque année. * nombre de points, ** pourcentage de données par rapport à la BD d'apprentissage

Entre 2001 et 2012, les valeurs de RMSE varient de 5,72 à 40,13 μatm , et les biais (en valeur absolue) de 0,82 à 39,34 μatm . Les points les plus éloignés des valeurs observées, appartiennent aux deux années 2001 et 2002 (RMSE de 30,94 et 40,13 μatm respectivement ; **Tableau 6.1**). Pour ces deux années, la quasi totalité des erreurs d'estimation de fCO₂ sont de l'ordre de 20 à 55 μatm . Effectivement, comme on l'a montré dans l'article Moussa *et al.* [2016], ces larges valeurs de RMSE sont causées par deux facteurs qui sont très liés l'un à l'autre. Le premier étant la variabilité interannuelle de fCO₂, qui fait que ces deux années là (2001 et 2002) sont caractérisées par une faible fCO₂ comparée aux années suivantes.

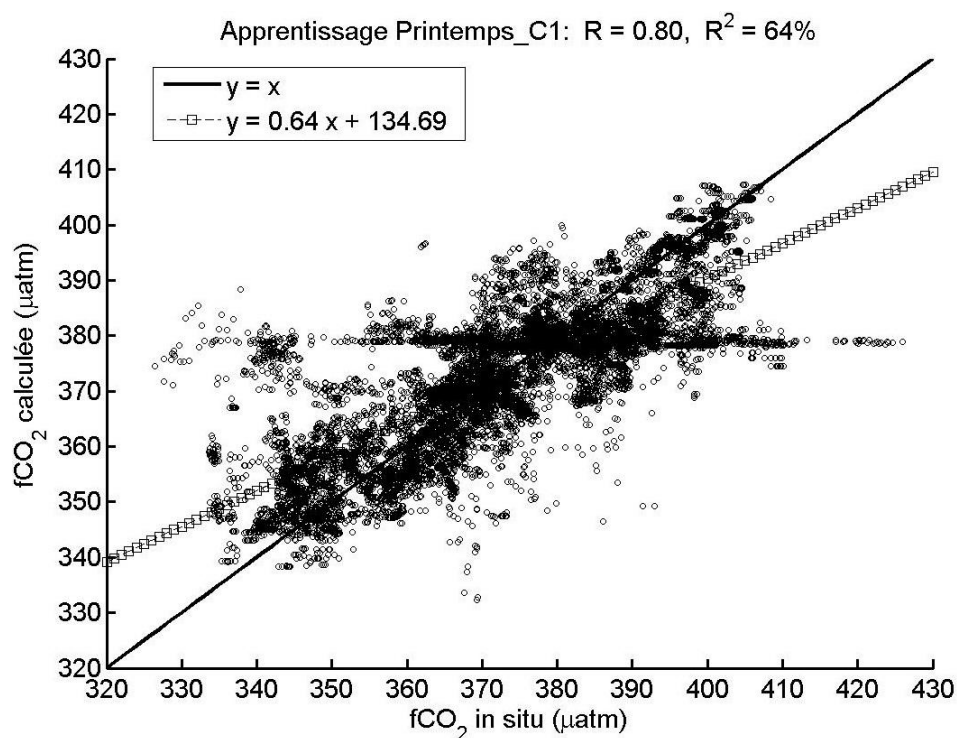


FIGURE 6.2 – Apprentissage de la classe C1 de la saison du printemps : $f\text{CO}_2$ estimée par le RN en fonction de $f\text{CO}_2$ *in situ*.

Pour 2001 la moyenne de $f\text{CO}_2$ est de $344,64 \pm 5,41 \mu\text{atm}$, et est de $336,99 \pm 5,78 \mu\text{atm}$ pour 2002. La différence entre ces deux moyennes et la moyenne globale de $f\text{CO}_2$ de la BD d'apprentissage ($373,63 \pm 17,40 \mu\text{atm}$), est de l'ordre de $\sim 29 \mu\text{atm}$ pour 2001 et $\sim 37 \mu\text{atm}$ pour 2002. Le deuxième facteur concerne le nombre de données qui est très réduit pour 2001 et 2002 ; comme vu dans le **Tableau 6.1**, le pourcentage de données pour 2001 et 2002 (1,55 et 0,49 % respectivement) est très petit par rapport aux autres années. Ces deux facteurs font que le RN construit, soit beaucoup plus adapté aux années postérieures à 2002, ce qui explique les larges RMSE pour 2001 et 2002. En outre, le RMSE global de 10,54 est très affecté par ces deux larges RMSE, en effet, en recalculant ce RMSE global sans prendre en considération ces deux années, on obtient une valeur de $9,5 \mu\text{atm}$.

Si l'on prend en considération la BD d'apprentissage en entier, on constate dans la **Figure 6.3a** et la **Figure 6.3b** que les résidus du modèle augmentent pour les valeurs de $f\text{CO}_2$ caractérisées par une faible densité. En effet, les valeurs situées dans les extrémités de l'intervalle de $f\text{CO}_2$ de cette classe, sont les moins bien apprises par le RN.

Pour les 10 intervalles de $f\text{CO}_2$ de l'histogramme présenté dans la **Figure 6.3a**, on présente dans le **Tableau 6.2** les bornes de chaque intervalle accompagnées du nombre de points de données correspondant, du pourcentage de ces données par rapport à la taille de la BD d'apprentissage, et aussi du RMSE correspondant à chaque intervalle.

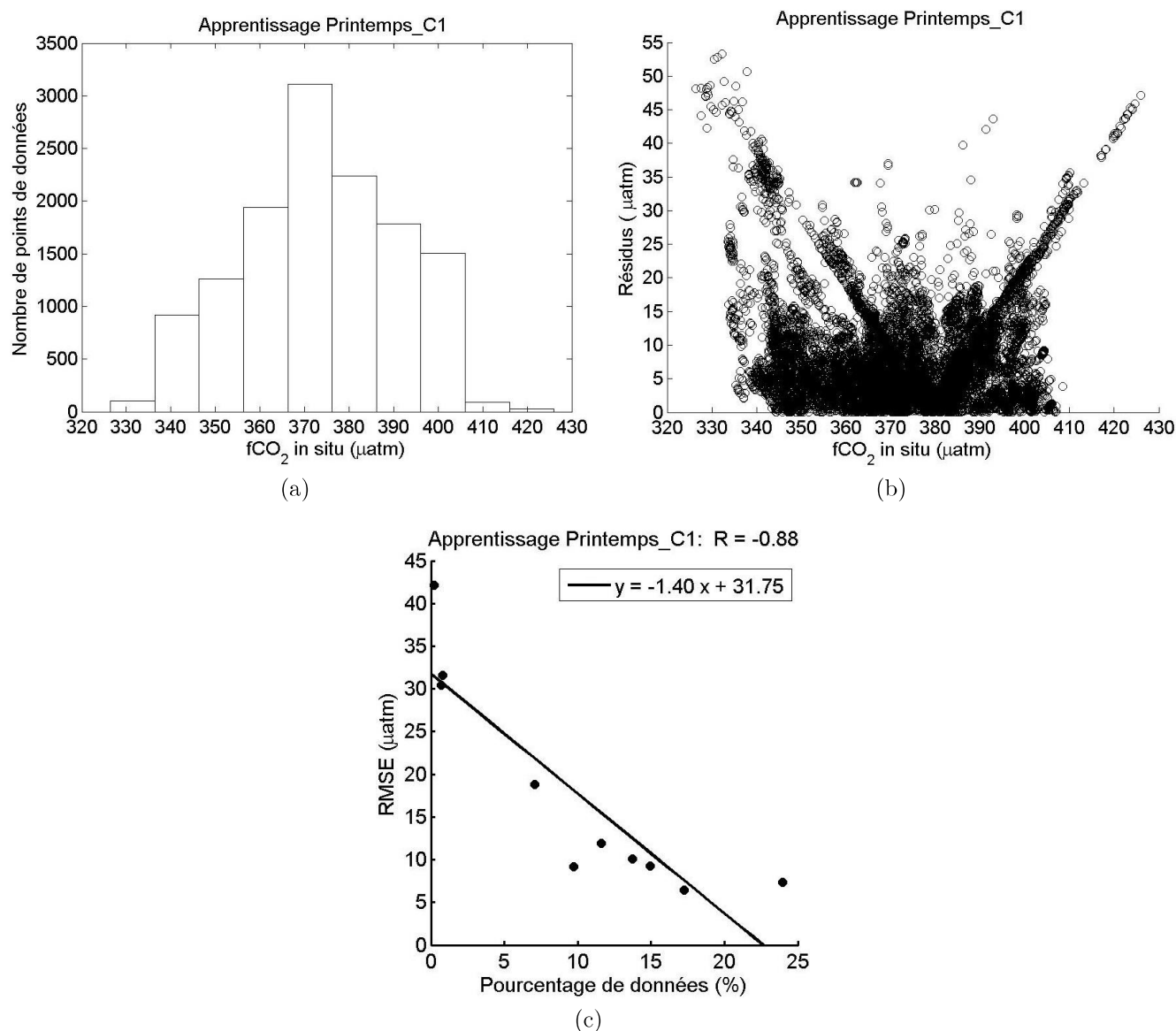


FIGURE 6.3 – Apprentissage de la classe C1 de la saison du printemps : (a) *histogramme de fCO₂*, (b) *relation entre les résidus et les valeurs de fCO₂*, (c) *relation entre le pourcentage de données (par rapport à la BD d'apprentissage) de chaque intervalle de l'histogramme, et les RMSE correspondants.*

Il est clair que la valeur du RMSE augmente pour les intervalles où le pourcentage de données est très petit (ex. N° 1, 9 et 10 ; **Tableau 6.2**), et diminue là où le pourcentage est élevé (ex. N° 4, 5 et 6 ; **Tableau 6.2**). La **Figure 6.3c** contient la présentation graphique de ces valeurs de RMSE. Celle-ci permet de mieux voir l'intensité de la relation entre le RMSE et le pourcentage de données.

N°	Borne inf. (μatm)	Borne sup. (μatm)	NP *	PD (%) **	RMSE (μatm)
1	327,41	335,41	103	0,79	31,57
2	337,37	345,37	921	7,09	18,81
3	347,32	355,32	1263	9,73	9,19
4	357,27	365,27	1940	14,94	9,24
5	367,23	375,23	3110	23,95	7,36
6	377,18	385,18	2241	17,26	6,44
7	387,13	395,13	1784	13,74	10,11
8	397,09	405,09	1505	11,59	11,89
9	407,04	415,04	90	0,69	30,42
10	416,99	424,99	26	0,2	42,14

TABLEAU 6.2 – Apprentissage de la classe C1 de la saison du printemps : relation entre le nombre de points et le RMSE, pour chaque intervalle de l’histogramme de $f\text{CO}_2$. * nombre de points ** pourcentage de données par rapport au nombre total de données de la BD d’apprentissage

6.2.1.2 Validation

Les résultats de l’étape de validation sont présentés dans le **Tableau 6.3** et la **Figure 6.4**. Puisque cette étape consiste à valider le RN avec un ensemble de données de la même nature que celui de l’apprentissage, à savoir des données *in situ*, on remarque bien que le résultat est proche de celui de l’apprentissage (RMSE : 10,56 μatm , biais : 0,15 μatm , R : 0,8, R^2 : 64%). On remarque aussi que les RMSEs élevés sont en 2001 et 2002, et ceci pour les mêmes raisons citées auparavant (dans l’étape d’apprentissage). Ainsi, on peut dire que ce RN est capable de produire la $f\text{CO}_2$ à partir d’un ensemble différent de données *in situ*.

Année	RMSE (μatm)	Biais (μatm)	NP *	PD (%) **
2001	31	-29,99	39	1,5
2002	41,52	-40,81	13	0,5
2004	13,15	-7,5	115	4,42
2006	10,67	-4,99	307	11,8
2007	9,9	2,68	489	18,8
2008	12,42	4,86	194	7,46
2009	5,54	1,56	293	11,26
2010	8,62	2,33	534	20,53
2012	8,9	0,87	617	23,72
Total	10,56	0,15	2601	/

TABLEAU 6.3 – Validation de la classe C1 de la saison du printemps : RMSE est biais correspondant à chaque année. * nombre de points, ** pourcentage de données par rapport à la BD de validation

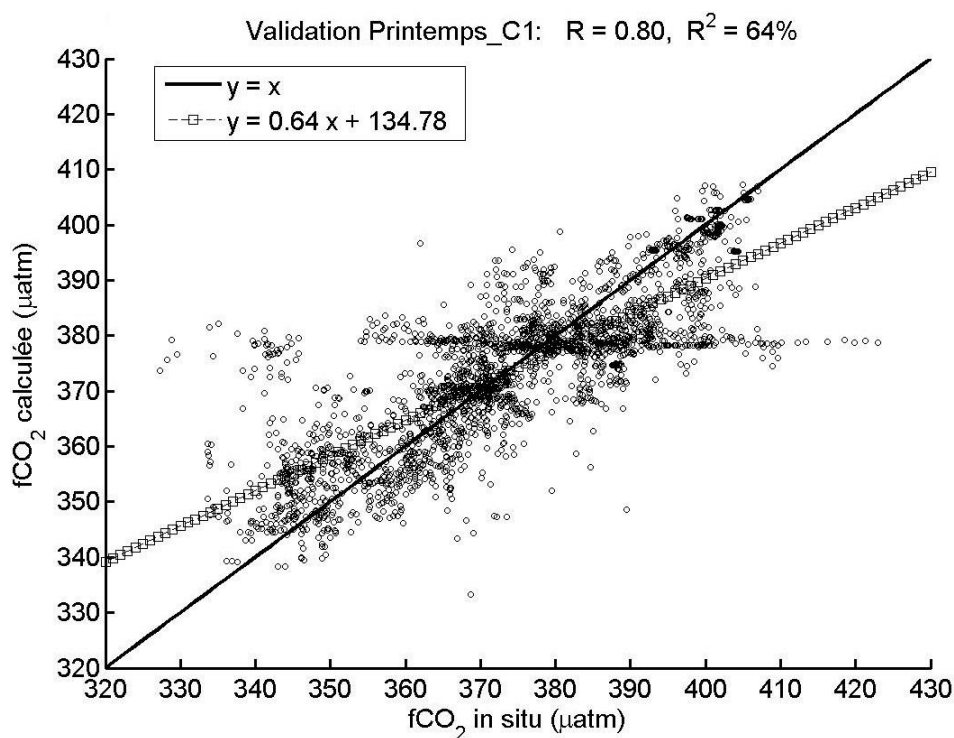


FIGURE 6.4 – Validation de la classe C1 de la saison du printemps : fCO₂ estimée par le RN en fonction de fCO₂ *in situ*.

6.2.1.3 Test

Les résultats de l'étape de test sont présentés dans le **Tableau 6.4** et la **Figure 6.5**. Le RMSE global obtenu est de 11,93 µatm, avec un biais de -0,25 µatm, un coefficient R de 0,72, et un coefficient R² de 52 %.

Année	fCO ₂ (µatm)		SST (°C)		SSS		NP *	PD (%) **
	RMSE	Biais	RMSE	Biais	RMSE	Biais		
2004	16,89	-9,08	0,36	0,28	0,354	-0,029	689	8,24
2006	11,83	-5,35	0,42	-0,19	0,492	-0,279	1833	21,92
2007	12,47	1,37	0,33	0,21	0,226	0,028	2926	35
2008	11,46	5,56	0,56	0,12	0,12	0,044	1158	13,85
2009	5,97	1,32	0,35	0,13	0,155	-0,061	1755	20,99
2010	14,01	3,69	0,58	-0,2	0,667	0,232	3170	46,13
2012	10,69	-3,27	0,53	-0,18	0,462	-0,173	3702	53,87
Total	11,93	-0,25	0,47	-0,03	0,439	-0,027	15233	/

TABLEAU 6.4 – Test de la classe C1 de la saison du printemps : RMSE et biais (de fCO₂, SST, et SSS) correspondants à chaque année. * nombre de points, ** pourcentage de données par rapport à la BD de test

Cette erreur est influencée par plusieurs facteurs qui varient suivant les années. Si l'on s'intéresse aux deux années où le RMSE est le plus large, à savoir 2004 et 2010, on

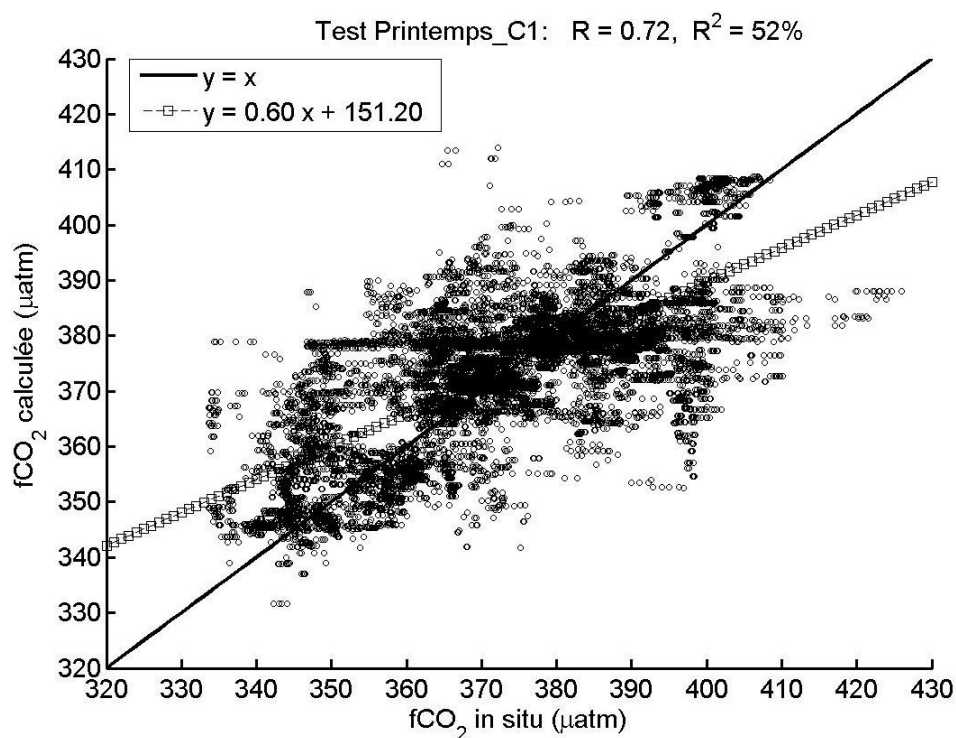


FIGURE 6.5 – Test de la classe C1 de la saison du printemps : fCO₂ calculée par le RN à partir des données satellitaire, en fonction de fCO₂ *in situ*.

remarque que pour l'année 2004, l'erreur est influencée par le nombre de données qui est très petit (8,24 %) par rapport aux autres années. Ce n'est pas le cas pour l'année 2010 où le pourcentage de données est très grand (46,13 %), mais le RMSE atteint une valeur de 14,01 µatm. Cette valeur du RMSE dépend de l'erreur sur la SSS (de SMOS) qui est assez large pour cette années ($\sim 0,7$). En essayant d'appliquer un seuil de 0,4 sur les erreurs de SSS SMOS (puisque l'incertitude de SSS SMOS est entre 0,3 et 0,4 [Boutin *et al.*, 2013; Kolodziejczyk *et al.*, 2016]), le RMSE pour cette année descend à 12,29 µatm. Ce même seuil a été appliqué sur la SSS de la BD globale (SODA et SMOS), un RMSE de 11,28 a été obtenu pour la fCO₂, avec un coefficient R de 0,8, et un coefficient R² de 60 %.

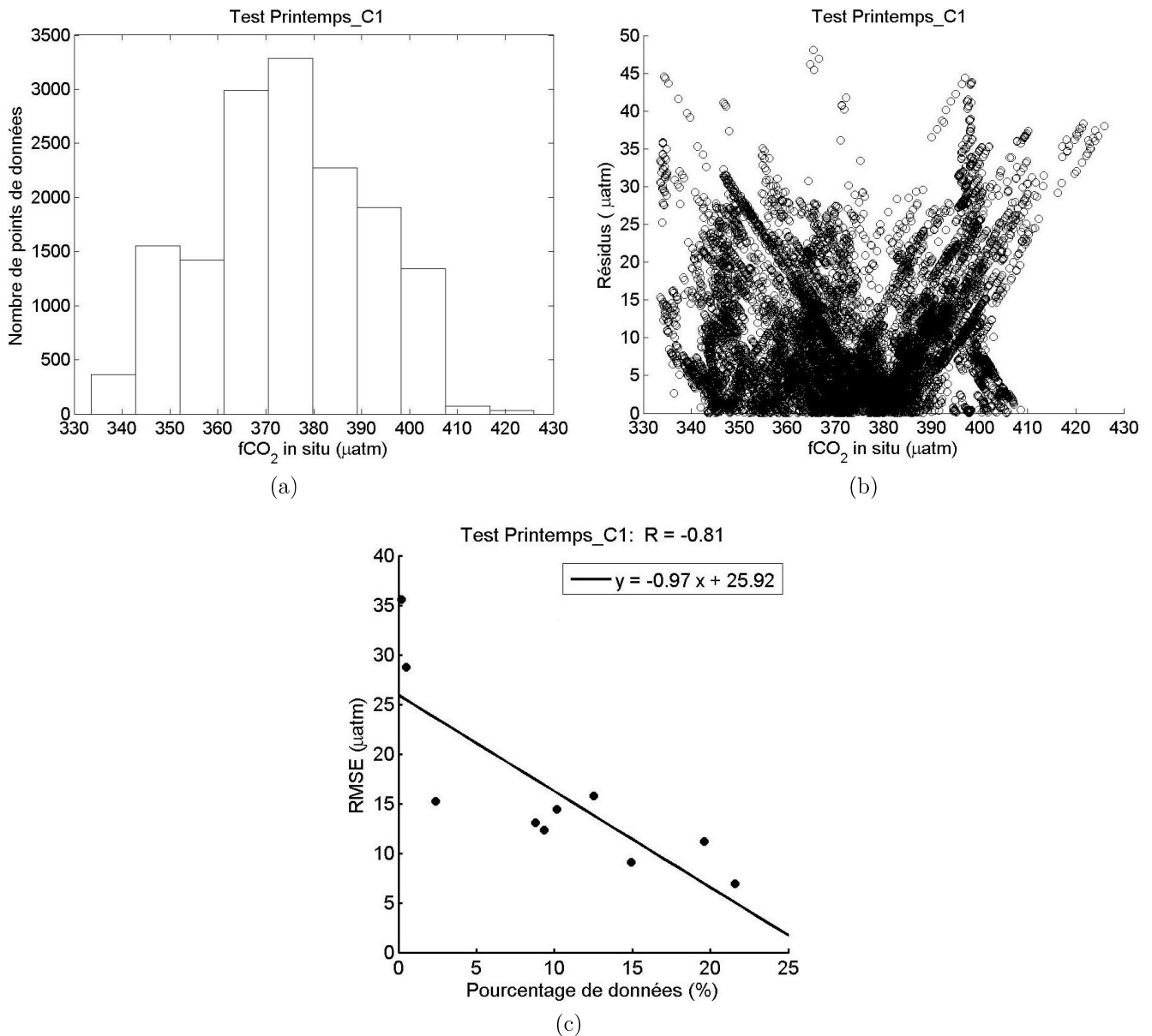


FIGURE 6.6 – Test de la classe C1 de la saison du printemps : (a) *histogramme de fCO₂*, (b) *relation entre les résidus et les valeurs de fCO₂*, (c) *relation entre le pourcentage de données (par rapport à la BD de test) de chaque intervalle de l’histogramme, et les RMSE correspondants.*

Comme pour la phase d’apprentissage, la **Figure 6.6a** contient l’histogramme de fCO₂, accompagné des résidus (**Figure 6.6b**). Il est remarquable que ces deux paramètres sont très liés, et que le RN estime mieux les valeurs de fCO₂ les plus présentes dans la BD du test.

Les 10 intervalles de l’histogramme (**Figure 6.6a**) sont présentés dans le **Tableau 6.5**, accompagnés des pourcentages de données ainsi que les RMSEs de fCO₂. De même que pour l’étape d’apprentissage, l’erreur est large pour les intervalles extrêmes présentant un faible pourcentage de données (cette relation est très visible dans la **Figure 6.6c**).

Or, il est à noter que ~ 90 % des points présentent un résidu inférieur à $20 \mu\text{atm}$.

N°	Borne inf. (μatm)	Borne sup. (μatm)	NP *	PD (%) **	RMSE (μatm)
1	334,22	342,22	362	2,38	15,25
2	343,46	351,46	1550	10,18	14,44
3	352,69	360,69	1421	9,33	12,33
4	361,93	369,93	2990	19,63	11,17
5	371,17	379,17	3286	21,57	6,91
6	380,4	388,4	2272	14,91	9,05
7	389,64	397,64	1907	12,52	15,75
8	398,88	406,88	1341	8,8	13,1
9	408,11	416,11	74	0,49	28,72
10	417,35	425,35	30	0,2	35,59

TABLEAU 6.5 – Test de la classe C1 de la saison du printemps : relation entre le nombre de points et le RMSE, pour chaque intervalle de l’histogramme de $f\text{CO}_2$. * nombre de points ** pourcentage de données par rapport au nombre total de données de la BD de test

6.2.2 Printemps : Classe 2 (C2)

Pour cette deuxième classe, le RN optimal est celui ayant 6 neurones cachés.

6.2.2.1 Apprentissage

Les résultats de l’apprentissage de la classe 2 du printemps sont présentés dans le **Tableau 6.6** et la **Figure 6.7**. Un RMSE global de $10,74 \mu\text{atm}$ a été obtenu, avec un très faible biais de $-0,02 \mu\text{atm}$, un coefficient R de 0,81 et un coefficient R^2 de 66 %. Sur les 13 années, la valeur du RMSE varie entre 6,78 et $25,67 \mu\text{atm}$.

L’année 2001 présente le plus grand RMSE ($25,67 \mu\text{atm}$), dû à la même raison que pour l’apprentissage de la classe C1 (**Section 6.2.1.1**), à savoir le faible pourcentage de données (ici 0,24 %) comparé aux autres années, accompagné du fait que la $f\text{CO}_2$ présente une moyenne un peu faible par rapport à la tendance générale de la BD d’apprentissage ($348,29 \pm 5,77 \mu\text{atm}$ et $361,59 \pm 18,31 \mu\text{atm}$, respectivement). Pour l’année 2002, et contrairement à la classe C1, le RMSE n’est pas très large (ici $13,23 \mu\text{atm}$). Bien que cette année présente aussi une faible moyenne de $f\text{CO}_2$ ($344,42 \pm 11,5 \mu\text{atm}$) comparée à la moyenne générale, son écart-type est plus grand (par rapport à 2001 de C2, et 2001, 2002 de C1), et le nombre de données l’est aussi (ici 7,23 %).

Année	RMSE (μatm)	Biais (μatm)	NP *	PD (%) **
2001	25,67	-22,89	122	0,24
2002	13,03	-10,2	3676	7,23
2003	7,65	-4,45	3740	7,35
2004	7,36	-4	3023	5,94
2005	7,51	-3,08	2345	4,61
2006	10,53	-5,67	4485	8,82
2007	7,93	-1,51	5573	10,96
2008	6,78	-0,2	3407	6,7
2009	15,88	1,99	5641	11,09
2010	12,08	7,06	7641	15,02
2011	10,25	4,08	2503	4,92
2012	7,7	3,06	7202	14,16
2013	16,5	7,98	1502	2,95
Total	10,74	-0,02	50860	/

TABLEAU 6.6 – Apprentissage de la classe C2 de la saison du printemps : RMSE et biais correspondants à chaque année. * nombre de points, ** pourcentage de données par rapport à la BD d'apprentissage

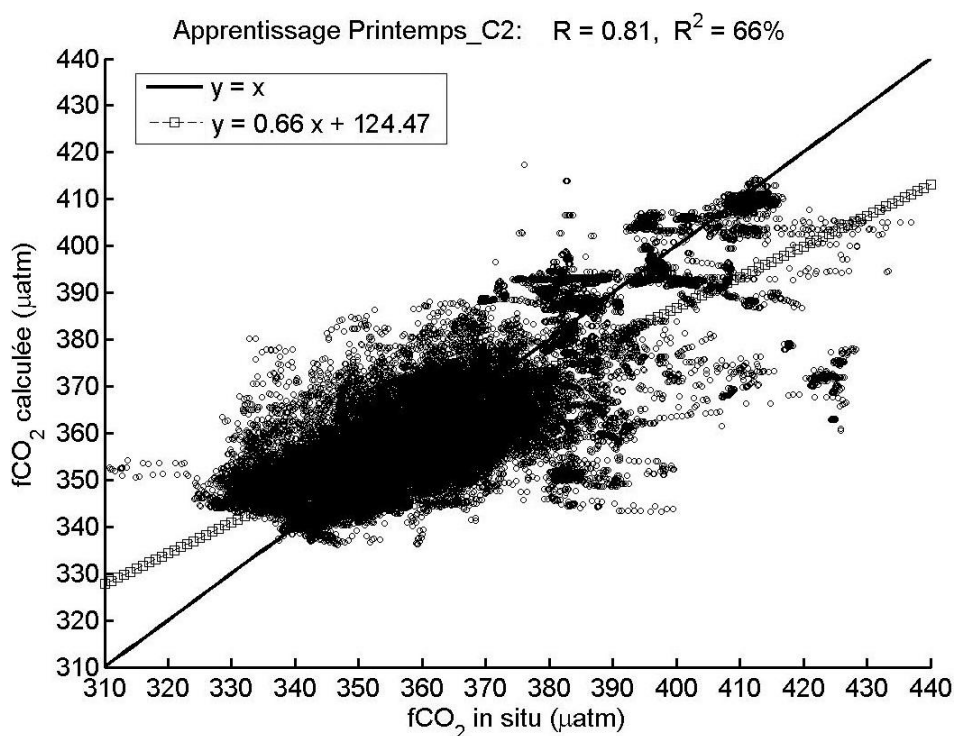


FIGURE 6.7 – Apprentissage de la classe C2 de la saison du printemps : fCO_2 estimée par le RN en fonction de fCO_2 in situ.

Pour l'année 2009, les très fortes valeurs des résidus surviennent le 06/05/2009 (voir

Figure 6.8a). Pendant ce jour, la $f\text{CO}_2$ connaît de fortes variabilités; entre 6h et 7h du matin elle passe de 391,53 à 426,7 μatm (voir **Figure 6.8b**).

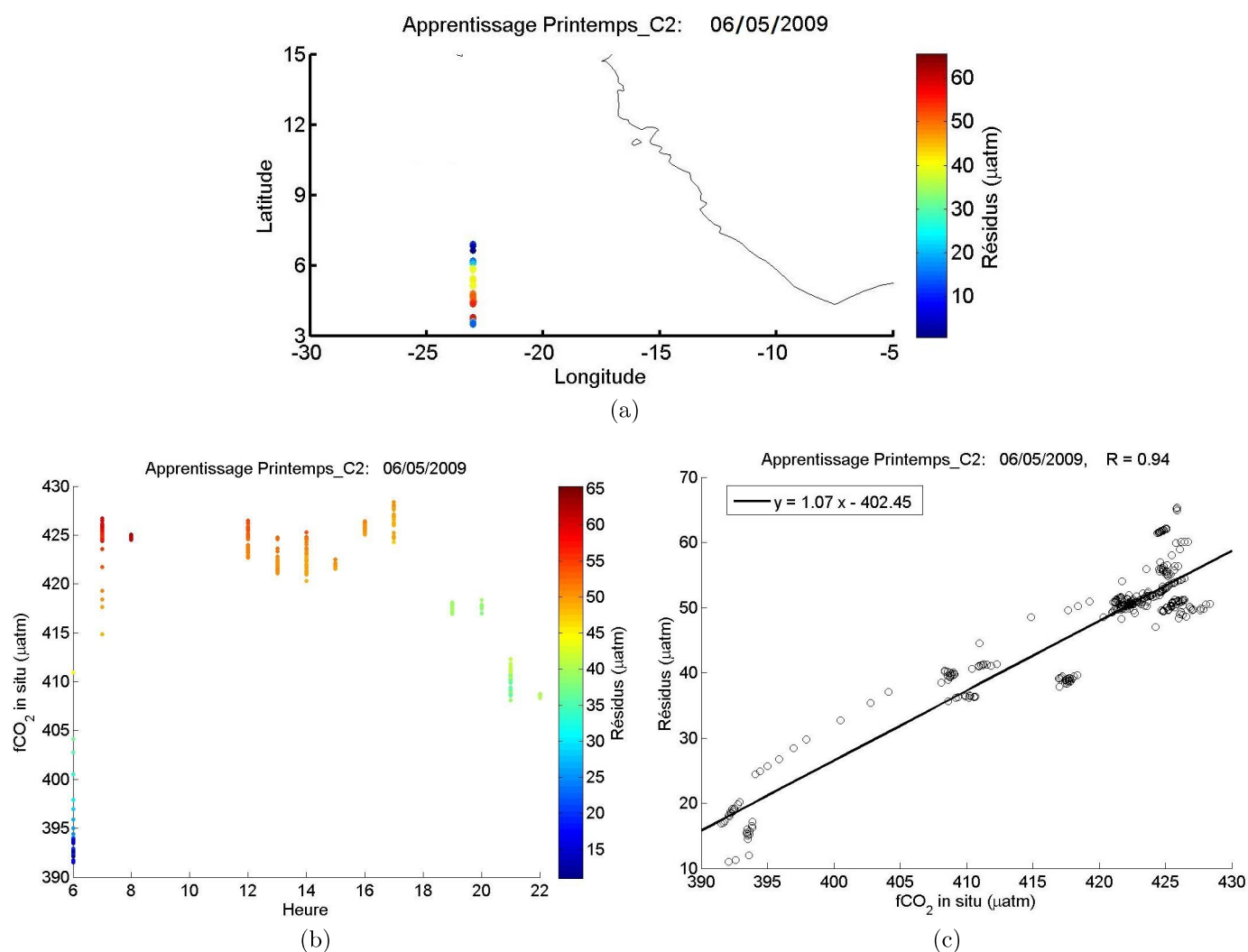


FIGURE 6.8 – Apprentissage de la classe C2 de la saison du printemps : Résidus du jour 06/05/2009. (a) positions géographiques des points, (b) variation de $f\text{CO}_2$ en fonction des heures, (c) relation entre les résidus et les valeurs de $f\text{CO}_2$

Cette forte augmentation de $f\text{CO}_2$ (35 μatm) est accompagnée d'une forte augmentation des résidus du RN. Comme vu dans la **Figure 6.8b**, entre 6h et 7h, l'erreur passe de 11 à 65 μatm . La relation entre les résidus du RN et l'augmentation de $f\text{CO}_2$ est bien illustrée dans la **Figure 6.8c**, où le coefficient R est de 0,94.

Par ailleurs, une grande partie des valeurs de $f\text{CO}_2$ de ce jour dépasse 400 μatm ($\sim 90\%$). Or, les valeurs de $f\text{CO}_2$ dépassant 400 μatm ne représentent que $\sim 5\%$ de toute la BD d'apprentissage (**Tableau 6.7**). Tous ces facteurs contribuent à ce que le RMSE de 2009 atteigne une valeur de 15,88 μatm .

Pour l'année 2013, les larges résidus apparaissent au mois de mars (**Figure 6.9a**). En effet, avec une moyenne de $384,47 \pm 4,57$, la $f\text{CO}_2$ est très forte pendant ce mois comparée aux reste de la BD d'apprentissage ($361,59 \pm 18,31$). Plus encore, si l'on s'intéresse que sur

la zone où se situent les points de ce mois ; de 21,5°N-70°O à 23,5°N-65°O, on remarque dans la **Figure 6.9b** que par rapport aux mois de mars des autres années, la moyenne de fCO₂ est très élevée en 2013 ; l'écart entre la moyenne de cette année et la moyenne générale dans cette zone (356,03 µatm) est de 28,44 µatm.

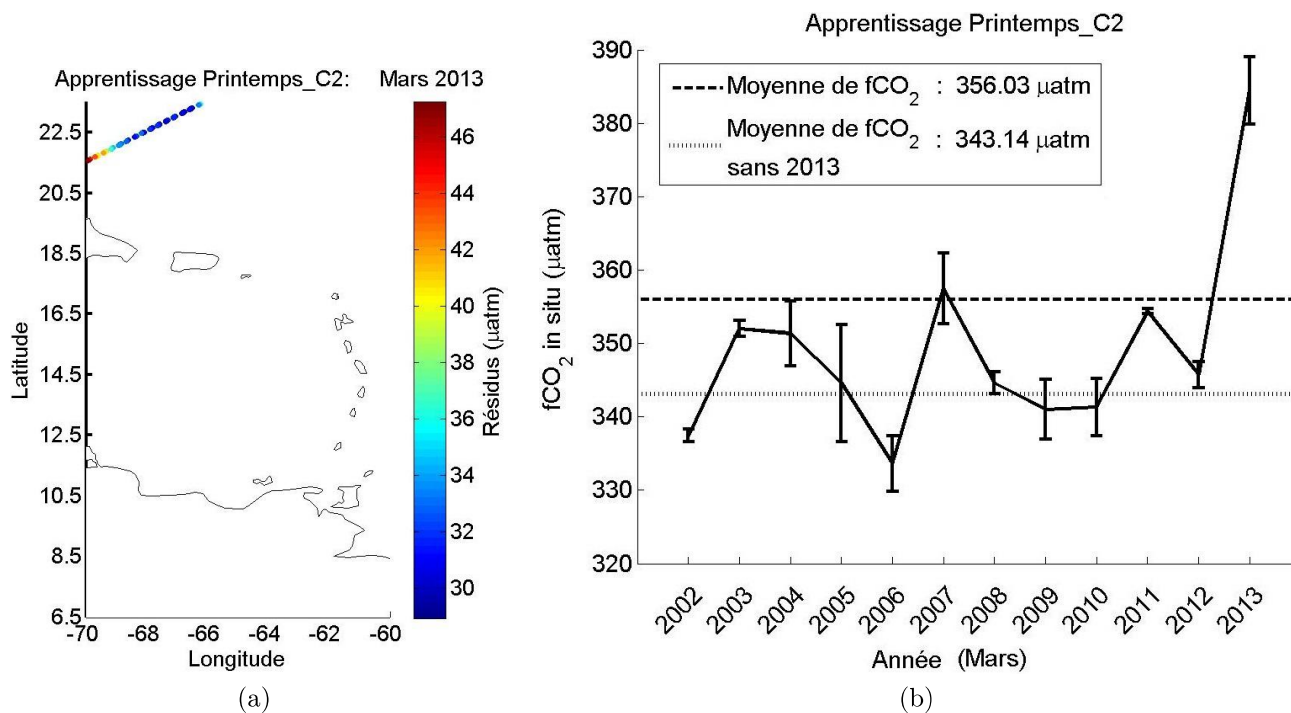


FIGURE 6.9 – Apprentissage de la classe C2 : (a) positions géographiques des points (mars 2013), (b) moyennes et écart-types de fCO₂ (pendant les mois de mars de toutes les années) dans la zone 21,5°N-70°O à 23,5°N-65°O.

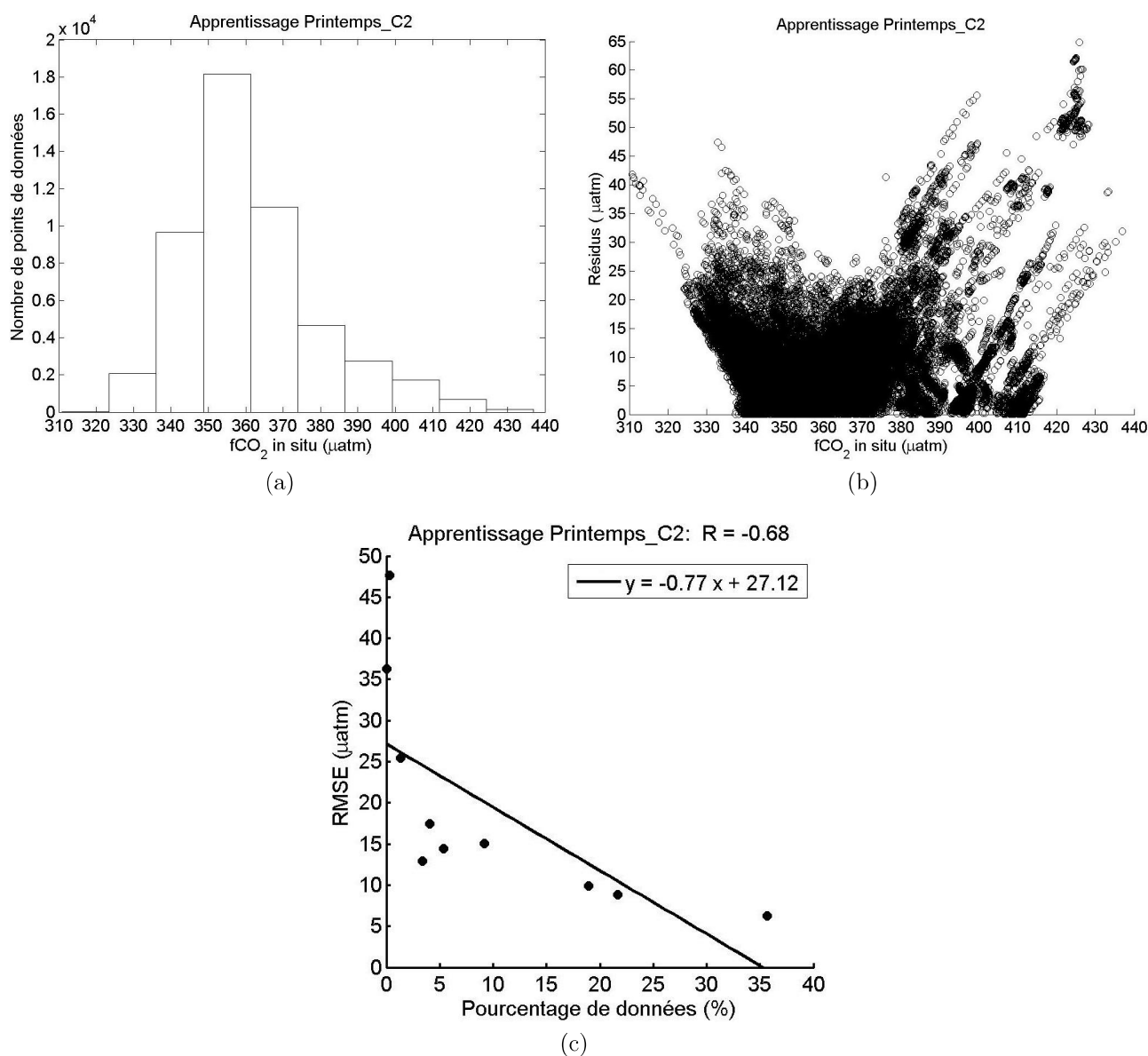


FIGURE 6.10 – Apprentissage de la classe C2 de la saison du printemps : (a) histogramme de fCO_2 , (b) relation entre les résidus et les valeurs de fCO_2 , (c) relation entre le pourcentage de données (par rapport à la BD d'apprentissage) de chaque intervalle de l'histogramme, et les RMSE correspondants.

Comme pour la classe C1, on constate, à partir des deux **Figure 6.10a** et **6.10b**, que le RN apprend mieux les valeurs de fCO_2 ayant une forte présence dans la BD d'apprentissage. Les RMSE des 10 intervalles de l'histogramme sont présentés dans le **Tableau 6.7**. La relation entre la densité des valeurs de fCO_2 et les résidus du RN est aussi vérifiée pour cette classe (voir **Figure 6.10c**).

N°	Borne inf. (μatm)	Borne sup. (μatm)	NP *	PD (%) **	RMSE (μatm)
1	311,13	323,13	24	0,05	36,27
2	323,75	335,75	2082	4,09	17,43
3	336,36	348,36	9651	18,98	9,93
4	348,98	360,98	18143	35,67	6,28
5	361,6	373,6	11006	21,64	8,83
6	374,21	386,21	4657	9,16	15,03
7	386,83	398,83	2737	5,38	14,41
8	399,44	411,44	1719	3,38	12,93
9	412,06	424,06	684	1,34	25,46
10	424,67	436,67	157	0,31	47,63

TABLEAU 6.7 – Apprentissage de la classe C2 de la saison du printemps : relation entre le nombre de points et le RMSE, pour chaque intervalle de l’histogramme de fCO₂. * nombre de points ** pourcentage de données par rapport au nombre total de données de la BD d’apprentissage

6.2.2.2 Validation

Les résultats de l’étape de validation sont présentés dans le **Tableau 6.8** et la **Figure 6.11**. Les résultats sont similaires à ceux de l’étape d’apprentissage ; une RMSE de 10,72 μatm avec un faible biais de -0,01 μatm , un coefficient R de 0,81, et un coefficient R² de 66 %. Suite à ces résultats, on peut dire que le RN est capable de produire la fCO₂ à partir d’un ensemble différent de données *in situ*.

Année	RMSE (μatm)	Biais (μatm)	NP *	PD (%) **
2001	25,51	-21,37	23	0,23
2002	13,04	-10,2	735	7,22
2003	7,69	-4,52	749	7,36
2004	7,23	-3,87	604	5,93
2005	7,53	-3,06	469	4,61
2006	10,52	-5,63	896	8,8
2007	7,95	-1,52	1120	11,01
2008	6,75	-0,21	685	6,73
2009	15,82	1,98	1131	11,11
2010	12,04	7,03	1528	15,01
2011	10,28	4,14	501	4,92
2012	7,73	3,1	1437	14,12
2013	16,52	8,1	299	2,94
Total	10,72	-0,01	10177	/

TABLEAU 6.8 – Validation de la classe C2 de la saison du printemps : RMSE et biais correspondants à chaque année. * nombre de points de données ** pourcentage de données par rapport au nombre total de données de la BD de validation

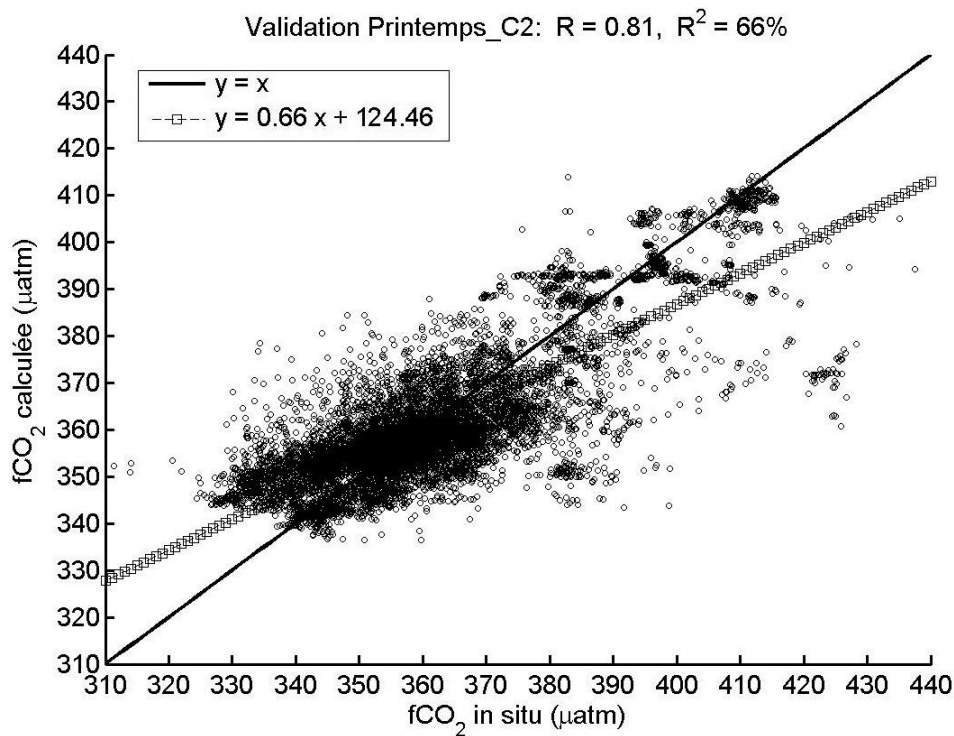


FIGURE 6.11 – Validation de la classe C2 de la saison du printemps : fCO₂ calculée par le RN en fonction de fCO₂ *in situ*.

6.2.2.3 Test

Les résultats de l'étape de test sont présentés dans le **Tableau 6.9** et la **Figure 6.12**. Un RMSE global de 12,21 µatm a été obtenu, avec un biais de -0,27 µatm, un coefficient R de 0,75, et un coefficient R² de 56%.

Comme pour l'étape d'apprentissage, les deux années où le RMSE est le plus large sont 2009 et 2013, et ceci dans les mêmes endroits (**Figures 6.8a** et **6.9a**) et pour les mêmes raisons que l'étape d'apprentissage (**Section 6.2.2.1**).

Pour certaines des années restantes, on remarque une augmentation du RMSE de l'ordre de 1 à 5 µatm par rapport à la phase d'apprentissage, ceci est dû à l'augmentation du nombre de données dans cette étape (53310 vs. 50860 en apprentissage), accompagnée des erreurs présentes sur les données de SST et de SSS. En effet, en appliquant un seuil de 0,4 sur la SSS (SODA et SMOS), un RMSE global de 11,6 a été retrouvé, avec un coefficient R de 0,78, et un coefficient R² de 62%.

Année	fCO ₂ (µatm)		SST (°C)		SSS		NP *	PD (%) **
	RMSE	Biais	RMSE	Biais	RMSE	Biais		
2003	12,7	-9,56	0,21	-0,06	0,462	-0,38	4489	13,26
2004	8,45	-4,61	0,33	0,05	0,221	-0,065	3625	10,7
2005	11,21	-6,92	0,39	-0,17	0,606	-0,315	2814	8,31
2006	13,52	-8,84	0,39	0,07	0,489	-0,305	5381	15,89
2007	8,16	0,39	0,47	0,27	0,256	-0,013	6693	19,76
2008	8,3	-0,52	0,57	0,2	0,261	-0,003	4092	12,08
2009	15,34	1,03	0,53	-0,03	0,252	-0,099	6772	20
2010	13,71	8,58	0,41	0,12	0,31	0,069	8939	45,97
2011	13,67	2,77	0,48	-0,14	0,667	0,132	1902	9,78
2012	10,14	0,71	0,37	0,13	0,636	-0,157	6954	35,76
2013	18,33	10,91	0,29	0,17	0,321	0,131	1649	8,48
Total	12,21	-0,27	0,42	0,08	0,417	-0,098	53310	/

TABLEAU 6.9 – Test de la classe C2 de la saison du printemps : RMSE et biais (de fCO₂, SST, et SSS) correspondant à chaque année.

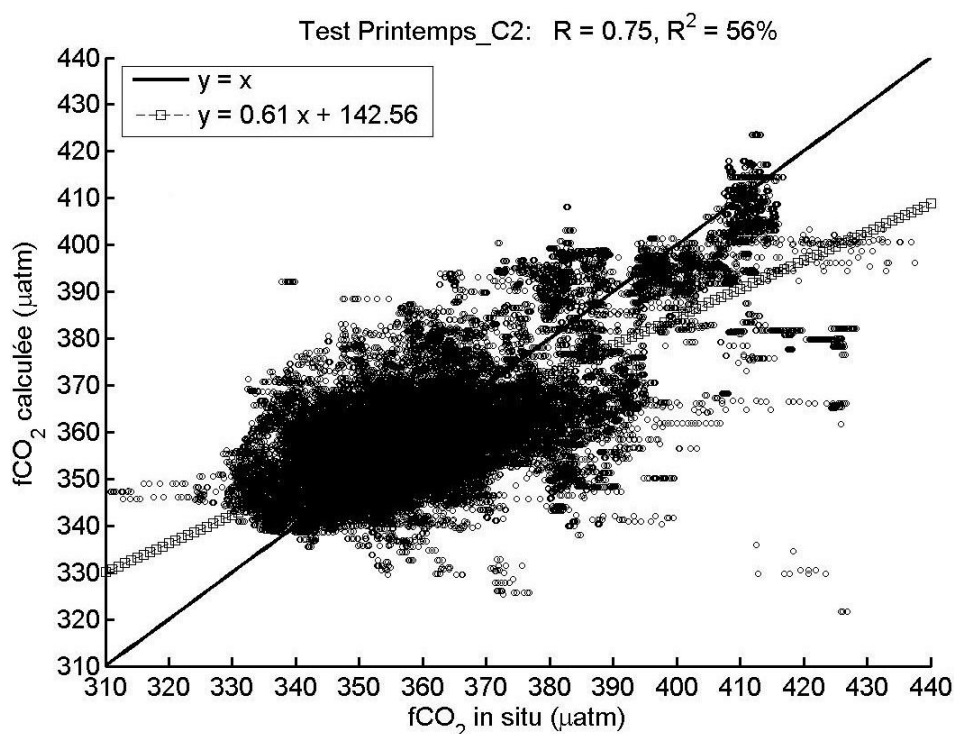


FIGURE 6.12 – Test de la classe C2 de la saison du printemps : fCO₂ calculée par le RN à partir des données satellitaires, en fonction de fCO₂ *in situ*.

Pour cette étape aussi on affiche l'histogramme de fCO₂ (**Figure 6.13a**), accompagné des résidus du test en fonction des valeurs de fCO₂ (**Figure 6.13b**).

Plus de 90 % des valeurs de fCO₂ (en sortie du RN) ont une part résiduelle inférieure à 20 µatm. Parmi les 10 % qui restent (résidu > 20 µatm), 83 % concernent les valeurs

de $f\text{CO}_2$ supérieurs à $380 \mu\text{atm}$ (**Figure 6.13b**), ce qui correspond aux $f\text{CO}_2$ ayant une faible densité dans la BD du test (**Figure 6.13a**). Là aussi, la relation entre les résidus du RN et le pourcentage des données est vérifiée (voir **Figure 6.13c** et **Tableau 6.10**).

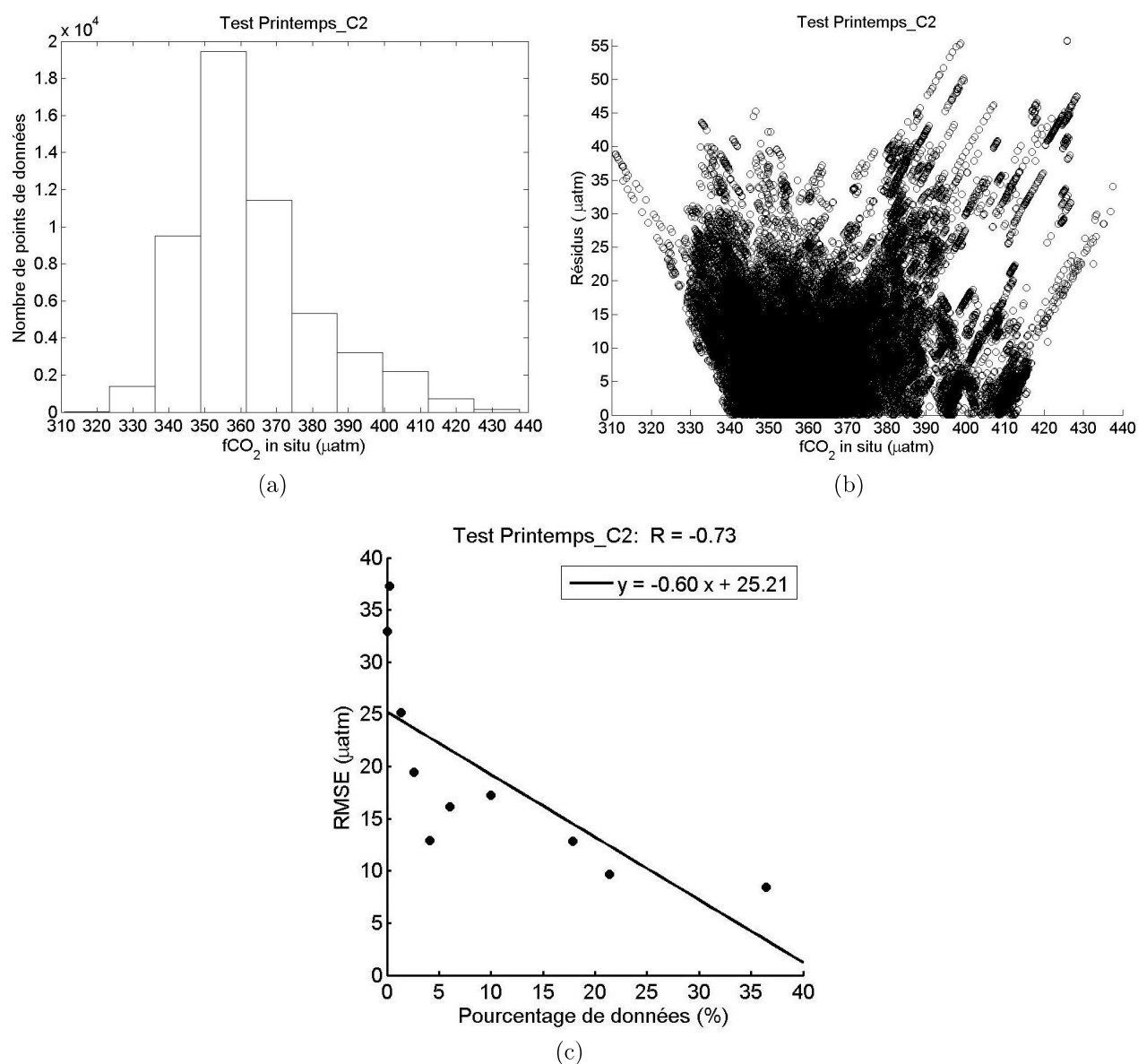


FIGURE 6.13 – Test de la classe C2 de la saison du printemps : (a) *histogramme de $f\text{CO}_2$* , (b) *relation entre les résidus et les valeurs de $f\text{CO}_2$* , (c) *relation entre le pourcentage de données (par rapport à la BD de test) de chaque intervalle de l’histogramme, et les RMSE correspondants.*

N°	Borne inf. (μatm)	Borne sup. (μatm)	NP *	PD (%) **	RMSE (μatm)
1	311,16	323,16	29	0,05	32,91
2	323,83	335,83	1386	2,6	19,48
3	336,5	348,5	9509	17,84	12,85
4	349,18	361,18	19427	36,44	8,45
5	361,85	373,85	11416	21,41	9,68
6	374,52	386,52	5304	9,95	17,21
7	387,19	399,19	3205	6,01	16,17
8	399,86	411,86	2182	4,09	12,89
9	412,54	424,54	706	1,32	25,13
10	425,21	437,21	146	0,27	37,29

TABLEAU 6.10 – Test de la classe C2 de la saison du printemps : relation entre le nombre de points et le RMSE, pour chaque intervalle de l’histogramme de fCO₂. * nombre de points ** pourcentage de données par rapport au nombre total de données de la BD de test

6.2.3 Printemps : Classe 3 (C3)

Pour cette troisième classe, le RN optimal est celui ayant 9 neurones cachés.

6.2.3.1 Apprentissage

Les résultats de l’étape d’apprentissage de la 3^{ème} classe de la saison du printemps sont présentés dans le **Tableau 6.11** et la **Figure 6.14**. Le RMSE global obtenu est de l’ordre de 11,27 μatm , avec un biais de 0,17 μatm , un coefficient R de 0,84, et un coefficient R² de 71 %. Comme pour la classe C2, les données sont disponibles sur une large échelle temporelle (13 années), et suivant ces années, les RMSEs varient de 5,75 à 21,17 μatm . Les deux années 2001 et 2002 sont celles où le RMSE est large (21,17 et 17,13 μatm respectivement), dû aux mêmes raisons que dans les autres classes (C1 : **Sections 6.2.1.1** et C2 : **Sections 6.2.2.1**).

Année	RMSE (μatm)	Biais (μatm)	Nombre de points	PD (%) *
2001	21,17	-19,02	500	0,87
2002	17,13	-13,2	755	1,31
2003	5,73	-2,91	2457	4,27
2004	9,09	-1,06	1150	2
2005	11,36	-5,33	1906	3,31
2006	8,72	-2,28	7437	12,93
2007	10,8	-0,52	5165	8,98
2008	12,73	-2,89	5634	9,79
2009	11,45	-1,29	4530	7,87
2010	12,88	4,27	19003	33,03
2011	11,96	-2,83	955	1,66
2012	7,59	1,11	7534	13,09
2013	8,83	5,45	508	0,88
Total	11,27	0,17	57534	/

TABLEAU 6.11 – Apprentissage de la classe C3 de la saison du printemps : RMSE et biais correspondants à chaque année. * pourcentage de données par rapport à la taille de la BD d'apprentissage

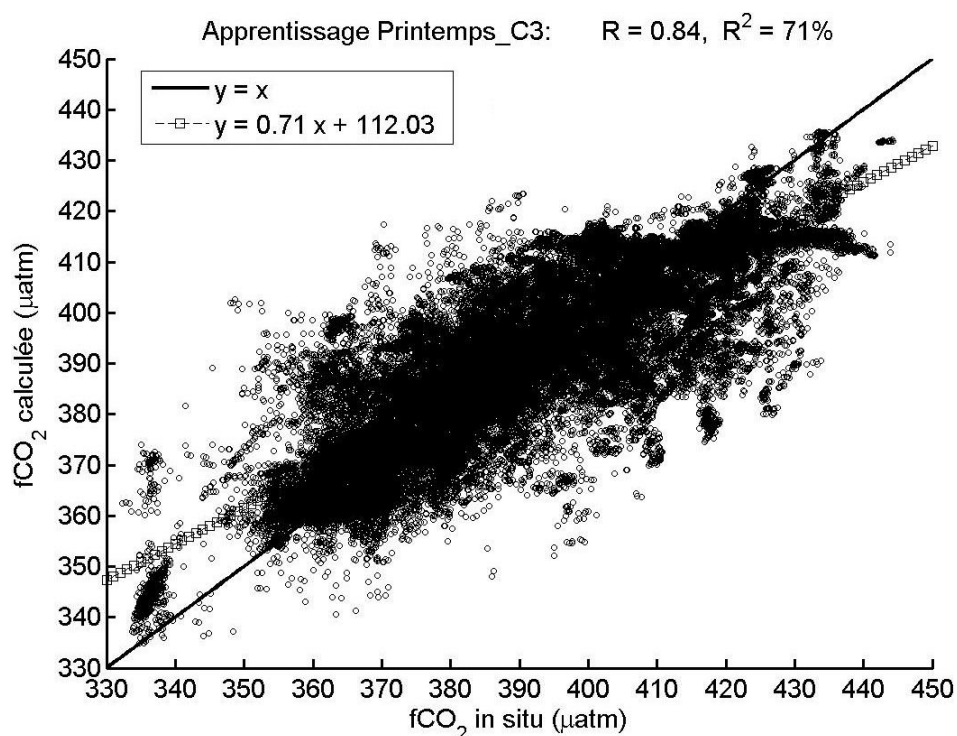


FIGURE 6.14 – Apprentissage de la classe C3 de la saison du printemps : fCO₂ estimée par le RN en fonction de fCO₂ *in situ*.

Comme on peut le voir dans les deux **Figures 6.15a** et **6.15b**, pour cette classe aussi la densité de fCO₂ et les résidus du RN sont très liés. Les intervalles de l'histogramme

illustrés dans la **Figure 6.15a** sont présentés dans le **Tableau 6.12**, accompagnés des valeurs de RMSE correspondants. La relation entre le pourcentage de données et le RMSE est bien visible dans la **Figure 6.15c**.

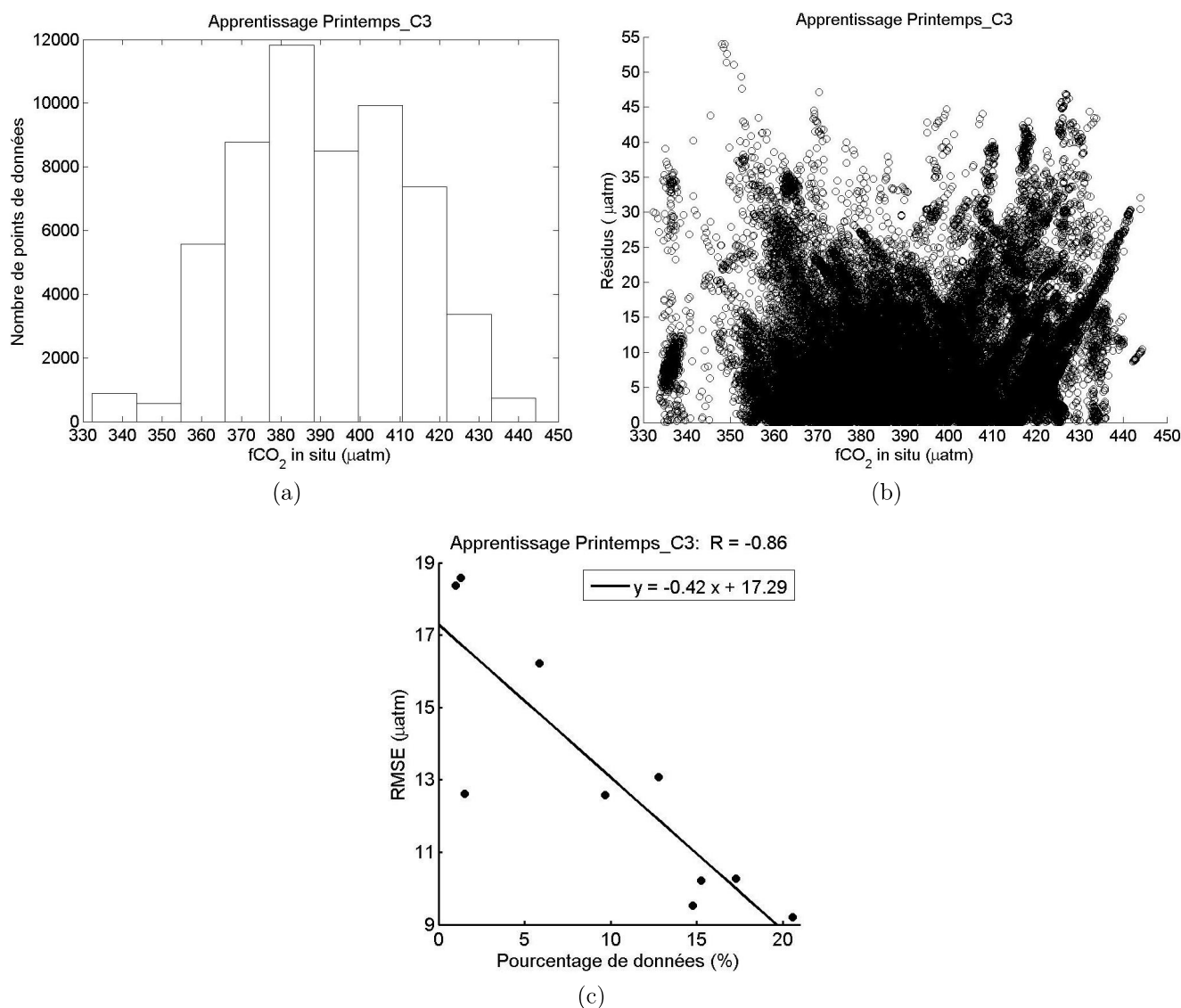


FIGURE 6.15 – Apprentissage de la classe C3 de la saison du printemps : (a) histogramme de fCO_2 , (b) relation entre les résidus et les valeurs de fCO_2 , (c) relation entre le pourcentage de données (par rapport à la BD d'apprentissage) de chaque intervalle de l'histogramme, et les RMSE correspondants.

N°	Borne inf. (μatm)	Borne sup. (μatm)	NP *	PD (%) **	RMSE (μatm)
1	332,95	342,95	879	1,53	12,62
2	344,15	354,15	567	0,99	18,37
3	355,34	365,34	5577	9,69	12,57
4	366,53	376,53	8786	15,27	10,22
5	377,72	387,72	11831	20,56	9,21
6	388,91	398,91	8501	14,78	9,53
7	400,11	410,11	9934	17,27	10,26
8	411,3	421,3	7365	12,8	13,08
9	422,49	432,49	3360	5,84	16,22
10	433,68	443,68	734	1,28	18,59

TABLEAU 6.12 – Apprentissage de la classe C3 de la saison du printemps : relation entre le nombre de points et le RMSE, pour chaque intervalle de l’histogramme de $f\text{CO}_2$. * nombre de points ** pourcentage de données par rapport au nombre total de données de la BD d’apprentissage

6.2.3.2 Validation

Les résultats de l’étape de validation sont présentés dans le **Tableau 6.13** et la **Figure 6.16**. Il est clair que le résultat est proche de celui de l’apprentissage (RMSE : 11,28 μatm , biais : 0,17 μatm , R : 0,84, R^2 : 71%), avec aussi des RMSEs élevés dans 2001 et 2002. On peut dire que ce RN est capable de produire la $f\text{CO}_2$ à partir d’un ensemble différent de données *in situ*.

Année	RMSE (μatm)	Biais (μatm)	NP *	PD (%) **
2001	21,1	-19,16	100	0,87
2002	17,43	-13,26	152	1,32
2003	5,83	-2,98	489	4,26
2004	9,12	-0,98	229	1,99
2005	11,34	-5,36	381	3,32
2006	8,68	-2,24	1486	12,94
2007	10,81	-0,55	1037	9,03
2008	12,82	-2,86	1121	9,76
2009	11,48	-1,25	905	7,88
2010	12,87	4,29	3790	32,99
2011	11,84	-2,61	191	1,66
2012	7,59	1,04	1505	13,1
2013	8,64	5,4	102	0,89
Total	11,28	0,17	11488	/

TABLEAU 6.13 – Validation de la classe C3 de la saison du printemps : RMSE et biais correspondants à chaque année. * nombre de points de données ** pourcentage de données par rapport au nombre total de données de la BD de validation

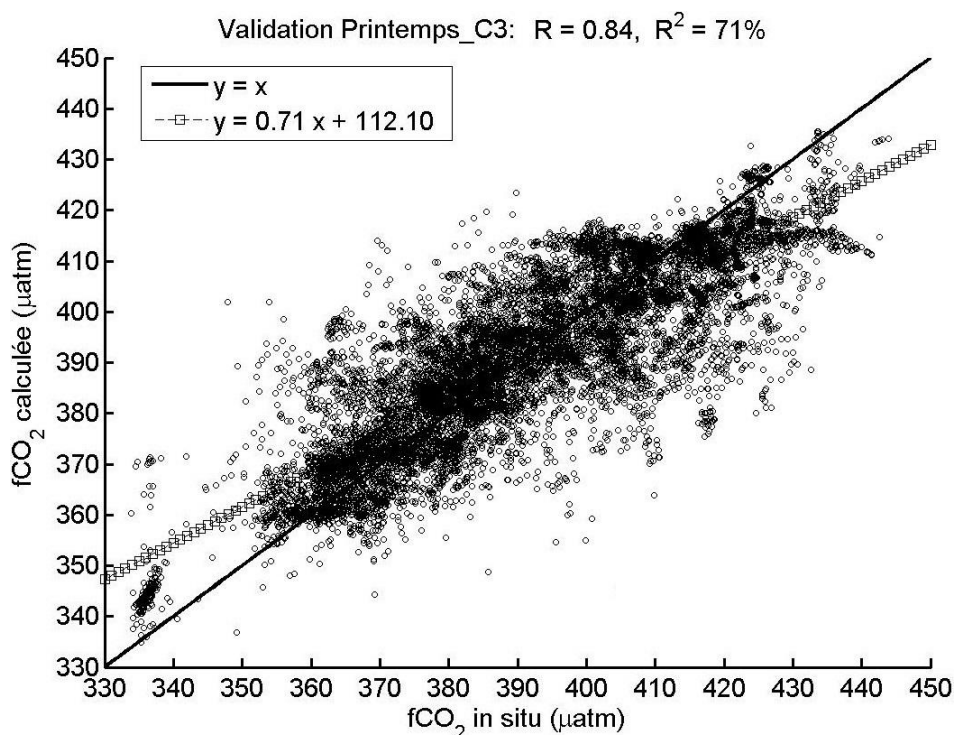


FIGURE 6.16 – Validation de la classe C3 de la saison du printemps : fCO₂ calculée par le RN en fonction de fCO₂ *in situ*.

6.2.3.3 Test

Les résultats de l'étape de test sont présentés dans le **Tableau 6.14** et la **Figure 6.17**. On peut constater que, contrairement aux deux précédentes classes, le RMSE global obtenu pour cette classe est assez large (15,91 µatm). Le biais obtenu est de 0,39 µatm, le coefficient R est de 0,67, et le coefficient R² est de 49 %.

On a remarqué que dans cette classe, plus de 42 % des points présentent un résidu de SSS qui dépasse 0,4, et le RMSE est directement lié à ces résidus. Bien que le RMSE global de SSS (0,508) dépasse légèrement les RMSEs des deux autres classes (C1 : 0,439 ; **Tableau 6.4**, et C2 : 0,417, **Tableau 6.9**), ceci a une influence assez importante sur le RMSE de fCO₂ de cette classe. En effet, en essayant de recalculer ce RMSE en appliquant un seuil de 0,4 sur les résidus de SSS, on a obtenu un RMSE global de fCO₂ de l'ordre de 12,9 µatm (3 µatm de différence), avec un coefficient R de 0,79, et un coefficient R² de 63 %.

Ceci s'applique aussi sur la majorité des années où le RMSE est large. Par exemple pour l'année 2005, 65 % des points ont un résidu de SSS qui dépasse 0,4. En appliquant le même seuil (0,4) sur les résidus de SSS, la valeur du RMSE passe de 19,99 à 12,11 µatm. Ceci est de même que pour d'autres années : par ex. en 2010 le RMSE passe de 18,66 à 14,88 µatm, et en 2012 il passe de 16,56 à 11,43 µatm.

Année	fCO ₂ (µatm)		SST (°C)		SSS		NP *	PD (%) **
	RMSE	Biais	RMSE	Biais	RMSE	Biais		
2003	6.14	2,28	0,51	0,44	0,434	-0,38	2946	8,83
2004	12.39	-0,54	0,33	0,21	0,407	-0,231	1377	4,13
2005	19.99	-8,84	0,4	0,15	0,595	-0,403	2137	6,4
2006	12.31	-4,19	0,44	0,1	0,462	-0,32	8870	26,57
2007	13.09	0,74	0,51	0,37	0,384	-0,204	5971	17,89
2008	15.9	-3,87	0,63	0,1	0,295	-0,162	6743	20,2
2009	14.59	-1,96	0,43	0,08	0,439	-0,269	5335	15,98
2010	18.66	5,43	0,47	0,03	0,627	0,46	20233	67,12
2011	15.99	-1,78	0,53	0,27	0,299	-0,01	941	3,12
2012	16.56	-0,46	0,34	0,09	0,499	0,337	8516	28,25
2013	12.21	5,72	0,4	0,33	0,783	-0,3	455	1,51
Total	15.91	0,39	0,47	0,12	0,508	0,05	63524	/

TABLEAU 6.14 – Test de la classe C3 de la saison du printemps : RMSE et biais (de fCO₂, SST, et SSS) correspondant à chaque année.

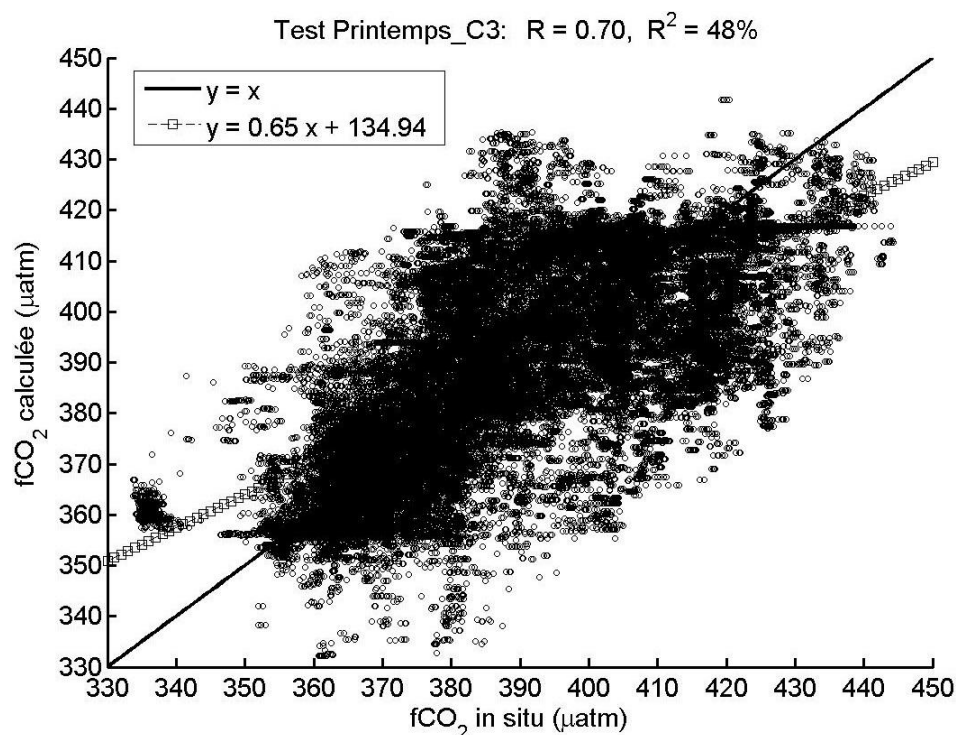


FIGURE 6.17 – Test de la classe C3 de la saison du printemps : fCO₂ calculée par le RN à partir des données satellitaires, en fonction de fCO₂ *in situ*.

Puisque les résidus de fCO₂ dépendent directement des résidus de SSS, on note (Figures 6.18a et 6.18b) que les larges résidus de fCO₂ ne concernent pas seulement

les valeurs de $f\text{CO}_2$ situées à l'extrême de l'intervalle de la BD de test. C'est pour cette raison que la relation entre le nombre de points de $f\text{CO}_2$ et les valeurs de RMSE n'est pas très visible dans le **Tableau 6.15** et la **Figure 6.18c**.

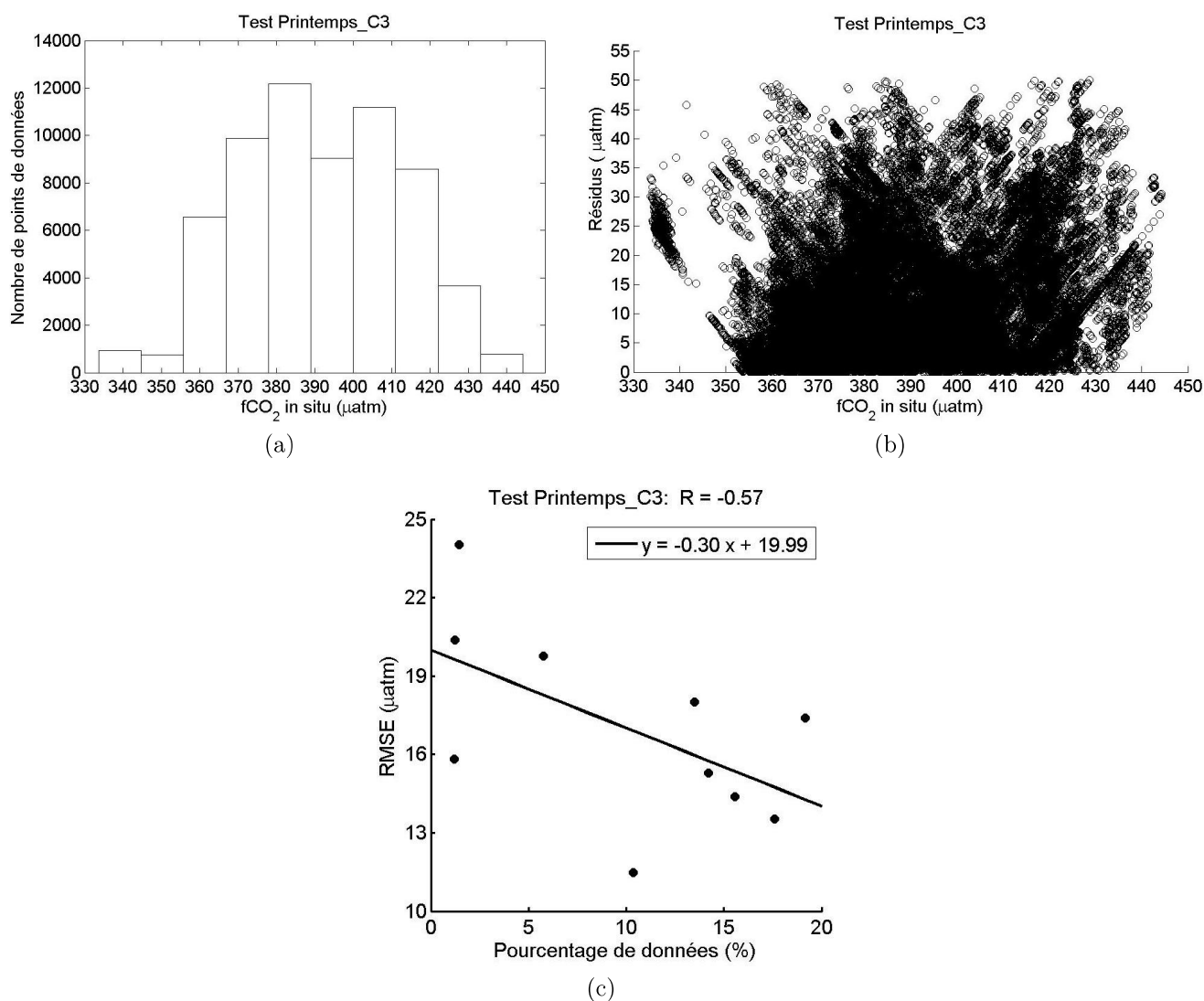


FIGURE 6.18 – Test de la classe C3 de la saison du printemps : (a) histogramme de $f\text{CO}_2$, (b) relation entre les résidus et les valeurs de $f\text{CO}_2$, (c) relation entre le pourcentage de données (par rapport à la BD de test) de chaque intervalle de l'histogramme, et les RMSE correspondants.

N°	Borne inf. (μatm)	Borne sup. (μatm)	NP *	PD (%) **	RMSE (μatm)
1	334,22	344,22	920	1,45	24,02
2	345,27	355,27	751	1,18	15,83
3	356,33	366,33	6574	10,35	11,48
4	367,39	377,39	9878	15,55	14,39
5	378,45	388,45	12169	19,16	17,4
6	389,51	399,51	9033	14,22	15,28
7	400,57	410,57	11172	17,59	13,52
8	411,63	421,63	8570	13,49	18
9	422,69	432,69	3667	5,77	19,77
10	433,75	443,75	790	1,24	20,37

TABLEAU 6.15 – Test de la classe C3 de la saison du printemps : relation entre le nombre de points et le RMSE, pour chaque intervalle de l’histogramme de $f\text{CO}_2$. * nombre de points ** pourcentage de données par rapport au nombre total de données de la BD de test

6.2.4 Hiver : Classe 1 (C1)

Pour cette première classe de la saison de l’hiver, en raison de la taille restreinte de la BD de construction du modèle (3478 points), on a choisi d’utiliser tous ces points pour l’apprentissage des RNs, et d’éliminer la phase dite de validation.

Le RN optimal obtenu est celui qui contient 4 neurones cachés.

6.2.4.1 Apprentissage

Les résultats de l’étape d’apprentissage sont présentés dans le **Tableau 6.16** et la **Figure 6.19**. Un RMSE global de 9 μatm a été obtenu, avec un biais de -0,23 μatm , un coefficient R de 0,67, et un coefficient R^2 de 45%.

Année	RMSE (μatm)	Biais (μatm)	NP *	PD (%) **
2005	10,42	-0,03	2099	60,35
2006	20,55	-20,44	11	0,32
2009	7,84	2,67	198	5,69
2010	4,43	-1,14	716	20,59
2012	3,26	-2,41	421	12,1
2013	23,90	23,84	33	0,95
Total	9	-0,23	3478	/

TABLEAU 6.16 – Apprentissage de la classe C1 de la saison de l’hiver : RMSE et biais correspondants à chaque année. * nombre de points, ** pourcentage de données par rapport à la BD d’apprentissage

Les faibles valeurs de R et R^2 sont due principalement au grand manque de données dans cette classe. En effet, en comparant avec la classe C1 de la saison du printemps ; ayant 12983 points pour la phase d'apprentissage, on constate que la taille de cette BD est très petite, ce qui rend très difficile la tâche d'apprentissage.

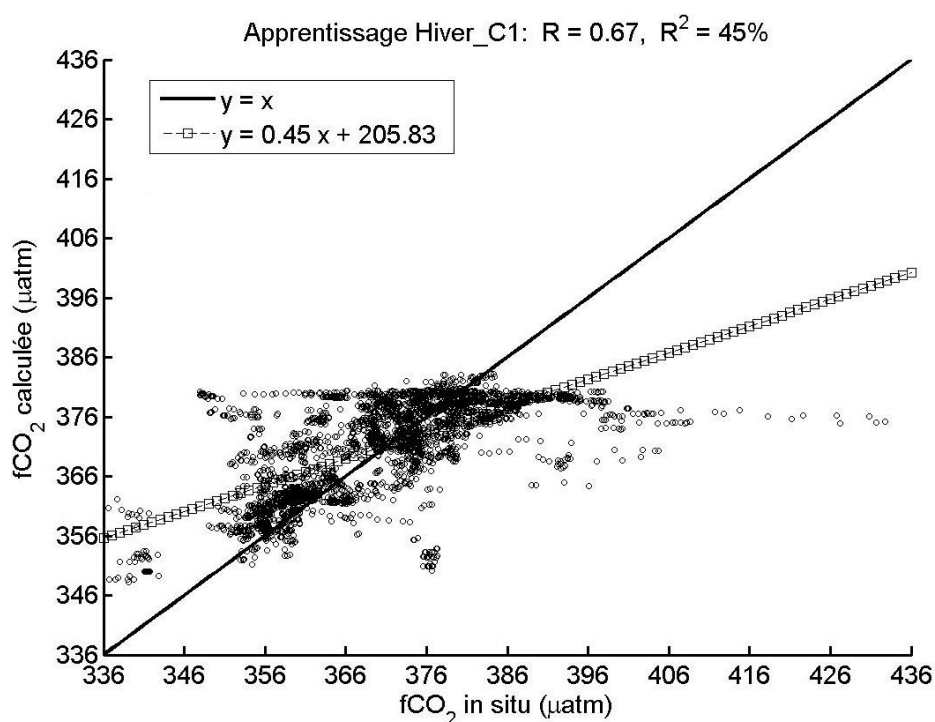


FIGURE 6.19 – Apprentissage de la classe C1 de la saison de l'hiver : fCO₂ estimée par le RN en fonction de fCO₂ *in situ*.

Les deux années où le RMSE est le plus large sont 2006 et 2013. Comme on voit dans le **Tableau 6.16**, ces deux années représentent un pourcentage inférieur à 1% de la BD d'apprentissage.

De même que pour les classes précédentes, les valeurs des résidus sont très liés à la densité des données. Comme illustré dans la **Figure 6.20**, la relation entre l'histogramme dans **(a)** et les résidus dans **(b)** est très visible. Cette relation est présentée dans la **Figure 6.20c**, et le détail des valeurs est présenté dans le **Tableau 6.17**.

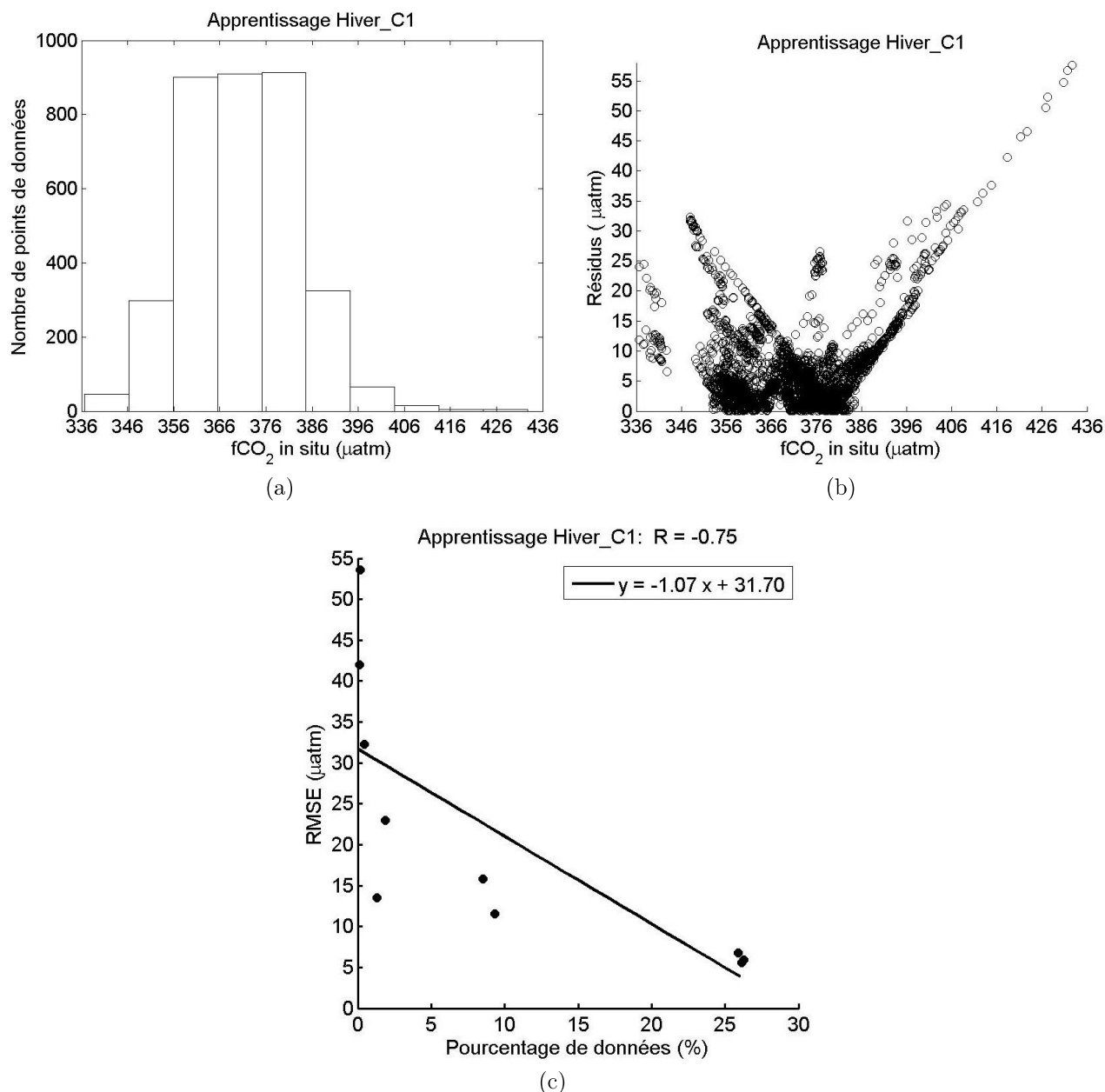


FIGURE 6.20 – Apprentissage de la classe C1 de la saison de l’hiver : (a) *histogramme de fCO_2* , (b) *relation entre les résidus et les valeurs de fCO_2* , (c) *relation entre le pourcentage de données (par rapport à la BD d’apprentissage) de chaque intervalle de l’histogramme, et les RMSE correspondants.*

N°	Borne inf. (μatm)	Borne sup. (μatm)	NP *	PD (%) **	RMSE (μatm)
1	337,45	345,45	45	1,29	13,48
2	347,06	355,06	297	8,54	15,83
3	356,67	364,67	900	25,88	6,82
4	366,28	374,28	909	26,14	5,56
5	375,89	383,89	913	26,25	5,96
6	385,5	393,5	324	9,32	11,53
7	395,11	403,11	66	1,9	22,97
8	404,72	412,72	15	0,43	32,25
9	414,33	422,33	4	0,12	41,98
10	423,94	431,94	5	0,14	53,62

TABLEAU 6.17 – Apprentissage de la classe C1 de la saison de l’hiver : relation entre le nombre de points et le RMSE, pour chaque intervalle de l’histogramme de fCO₂. * nombre de points ** pourcentage de données par rapport au nombre total de données de la BD d’apprentissage

6.2.4.2 Test

Les résultats de l’étape de test sont présentés dans le **Tableau 6.18** et la **Figure 6.21**. Un RMSE global de 10,62 μatm a été obtenu, avec un coefficient R de 0,54, et un R² de 29%. La légère augmentation du RMSE, et la diminution du R et R² par rapport à l’apprentissage, est due à la contribution des résidus de SST et de SSS (0,56 °C et 0,398 respectivement, **Tableau 6.18**).

Année	fCO ₂ (μatm)		SST (°C)		SSS		NP *	PD (%) **
	RMSE	Biais	RMSE	Biais	RMSE	Biais		
2005	11,63	-0,25	0,65	-0,2	0,422	-0,343	2099	90,94
2006	20,13	-19,93	0,23	0,06	0,358	-0,334	11	0,48
2009	9,97	5,51	0,31	-0,16	0,178	-0,031	198	8,58
2010	7,57	-1,99	0,29	0	0,516	-0,205	310	40,58
2012	3,12	-2,19	0,29	0,25	0,228	0,045	421	55,1
2013	22,31	22,27	0,05	-0,03	0,192	0,166	33	4,32
Total	10,62	-0,15	0,56	-0,11	0,398	-0,25	3072	/

TABLEAU 6.18 – Test de la classe C1 de la saison de l’hiver : RMSE et biais (de fCO₂, SST, et SSS) correspondants à chaque année. * nombre de points, ** pourcentage de données par rapport à la BD de test

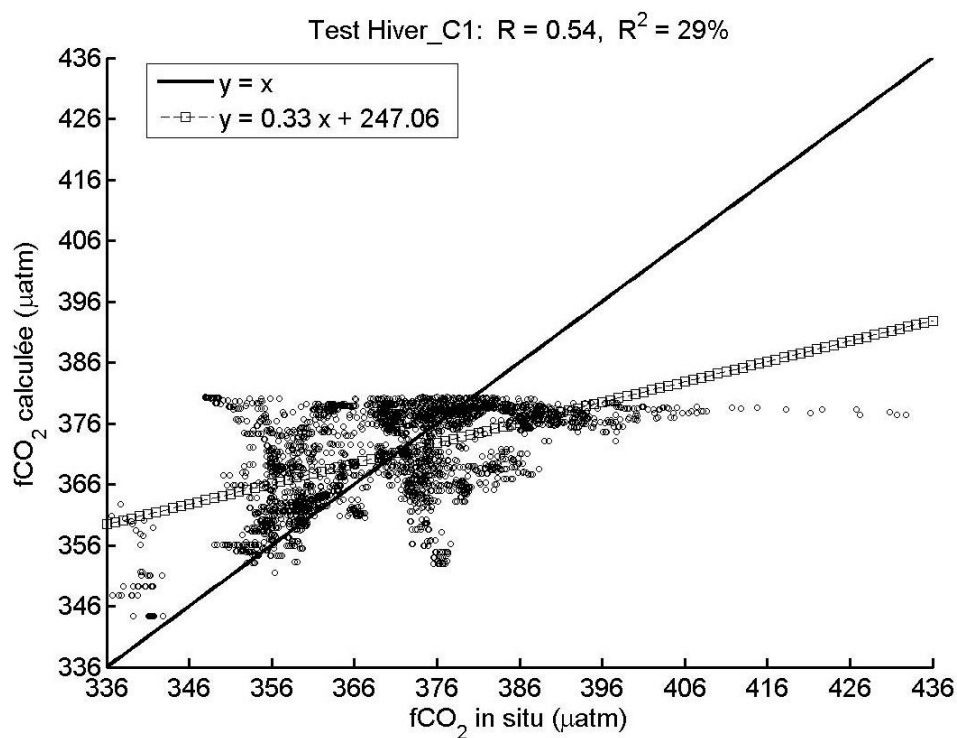


FIGURE 6.21 – Test de la classe C1 de la saison de l’hiver : fCO₂ calculée par le RN à partir des données satellitaires, en fonction de fCO₂ *in situ*.

Pour cette phase aussi, la relation entre le nombre de points et les résidus est confirmée. L’illustration de cette relation est présentée dans la **Figure 6.22** (le **Tableau 6.19** contient le détail des valeurs).

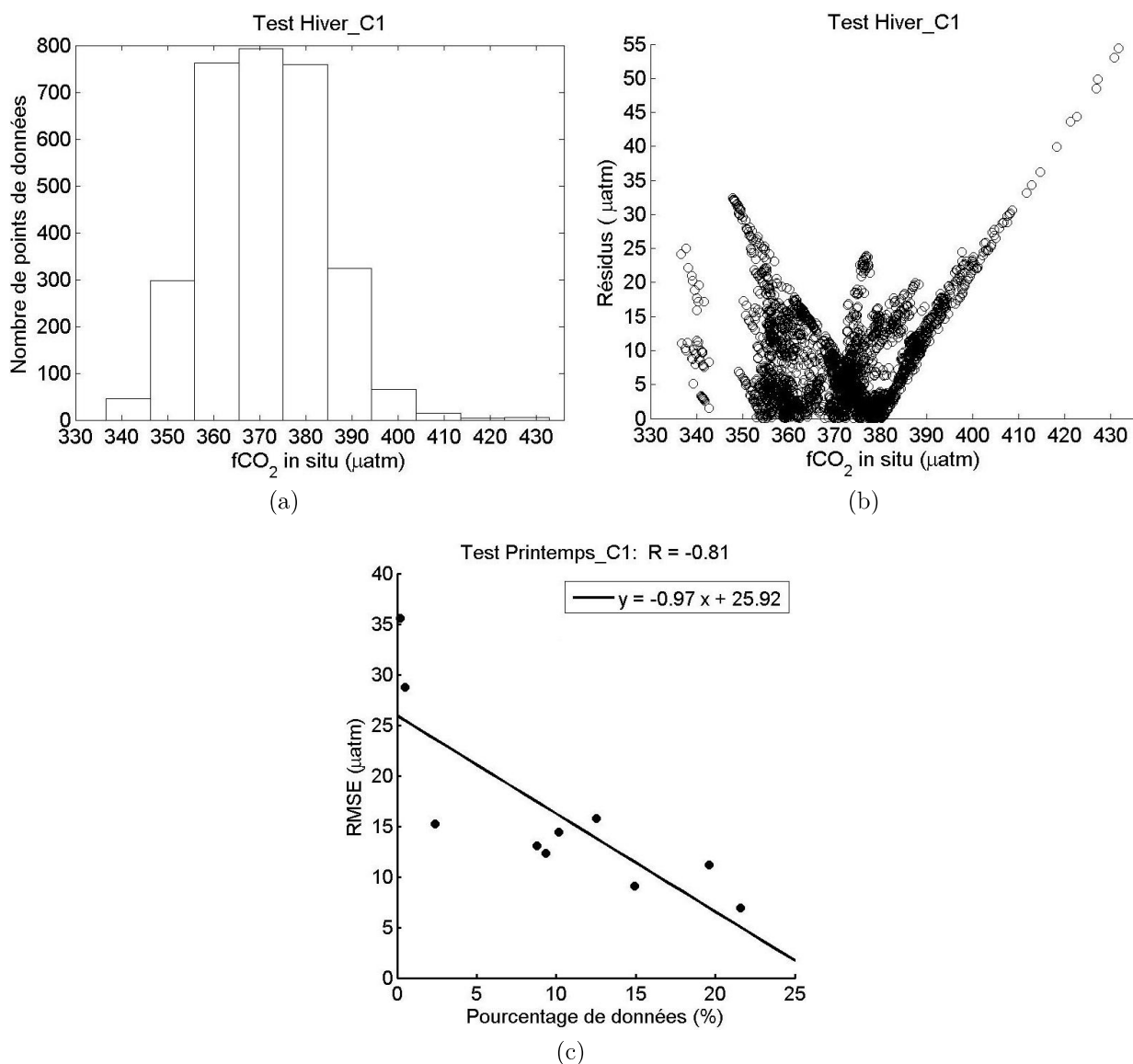


FIGURE 6.22 – Test de la classe C1 de la saison de l’hiver : (a) *histogramme de $f\text{CO}_2$* , (b) *relation entre les résidus et les valeurs de $f\text{CO}_2$* , (c) *relation entre le pourcentage de données (par rapport à la BD de test) de chaque intervalle de l’histogramme, et les RMSE correspondants.*

N°	Borne inf.	Borne sup.	NP *	PD (%) **	RMSE (μatm)
1	337,45	345,45	45	1,46	11,87
2	347,06	355,06	297	9,67	17,58
3	356,67	364,67	763	24,84	9,1
4	366,28	374,28	793	25,81	6,51
5	375,89	383,89	760	24,74	7,89
6	385,5	393,5	324	10,55	12,47
7	395,11	403,11	66	2,15	21,59
8	404,72	412,72	15	0,49	29,34
9	414,33	422,33	4	0,13	40,05
10	423,94	431,94	5	0,16	51,48

TABLEAU 6.19 – Test de la classe C1 de la saison de l’hiver : relation entre le nombre de points et le RMSE, pour chaque intervalle de l’histogramme de $f\text{CO}_2$. * nombre de points
** pourcentage de données par rapport au nombre total de données de la BD de test

6.2.5 Hiver : Classe 2 (C2)

Le RN optimal retrouvé pour la classe C2, est celui ayant 2 neurones cachés.

6.2.5.1 Apprentissage

Les résultats de l’apprentissage de la classe C2 sont présentés dans le **Tableau 6.20** et la **Figure 6.23**. Le RMSE global obtenu est de l’ordre de 10,18 μatm accompagné d’un biais presque nul (0,001 μatm). Les deux coefficients R et R^2 sont très forts; 0,94 et 88% respectivement.

Année	RMSE (μatm)	Biais (μatm)	NP *	PD (%) **
2002	17,22	-6,05	37	0,11
2003	6,5	-2,06	576	1,72
2004	4,95	-1,51	3027	9,04
2005	8,1	-2,45	2291	6,84
2006	12,16	-3,83	4857	14,51
2007	7,61	5,73	646	1,93
2008	6,76	-5,34	1447	4,32
2009	10,56	-0,83	2624	7,84
2010	10,44	4,12	3034	9,06
2011	11,02	3,26	3545	10,59
2012	9,23	-1,09	5740	17,15
2013	12,2	2,67	3767	11,25
2014	12,25	4,55	1884	5,63
Total	10,18	0,001	33475	/

TABLEAU 6.20 – Apprentissage de la classe C2 de la saison de l’hiver : RMSE et biais correspondants à chaque année. * nombre de points, ** pourcentage de données par rapport à la BD d’apprentissage

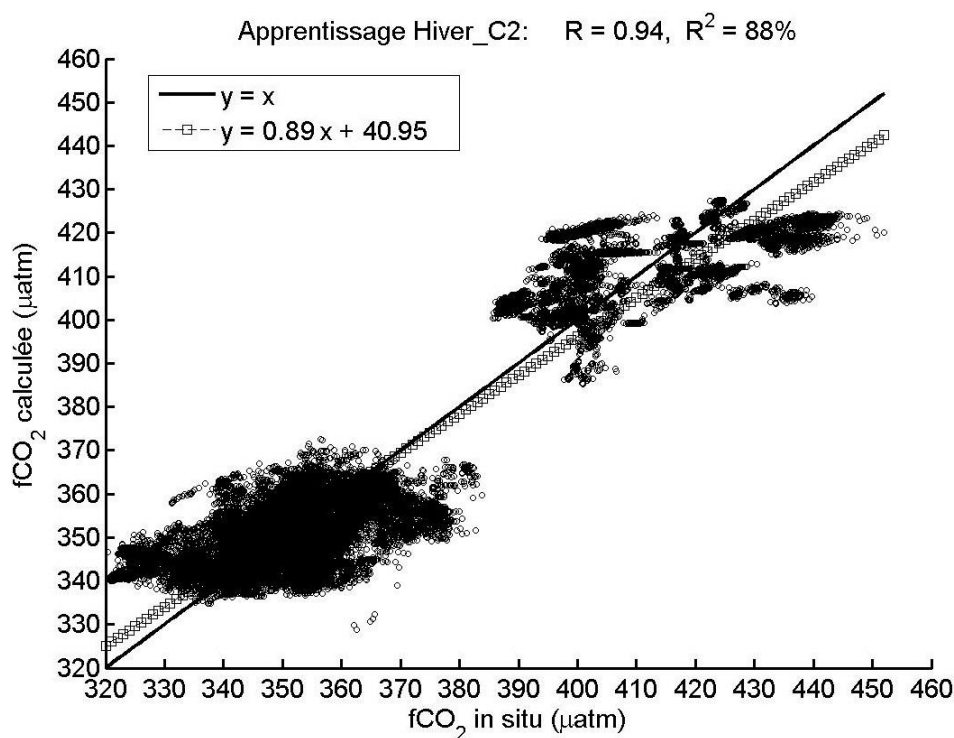


FIGURE 6.23 – Apprentissage de la classe C2 de la saison de l’hiver : fCO₂ estimée par le RN en fonction de fCO₂ *in situ*.

On constate dans l’histogramme des données d’apprentissage de cette classe (**Figure 6.24a**), que la distribution est bimodale (existence de deux modes statistiques). Ainsi, afin de mieux voir la relation qui existe entre le pourcentage de données et le RMSE, l’histogramme a été divisé en deux parties, et les relations ont été présentées en deux graphes (**Figures 6.24c** et **6.24d**). Le **Tableau 6.21** contient le détail de ces valeurs.

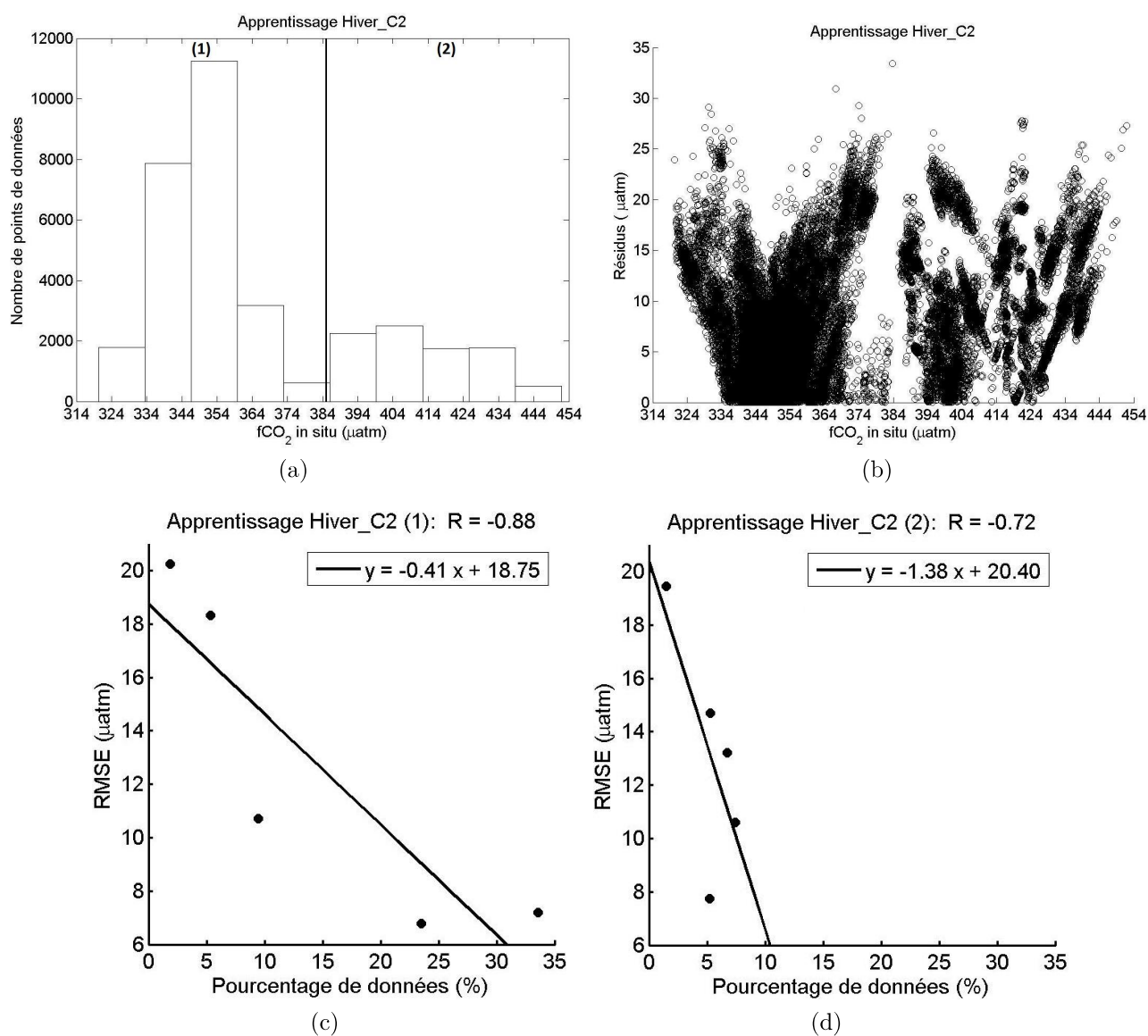


FIGURE 6.24 – Apprentissage de la classe C2 de la saison de l’hiver : (a) *histogramme de fCO_2* , (b) *relation entre les résidus et les valeurs de fCO_2* , (c) *relation entre le pourcentage de données (par rapport à la BD d’apprentissage) de chaque intervalle de l’histogramme, et les RMSE correspondants.*

N°	Borne inf. (μatm)	Borne sup. (μatm)	NP *	PD (%) **	RMSE (μatm)
1	320,92	332,92	1789	5,34	18,31
2	334,08	346,08	7870	23,51	6,78
3	347,23	359,23	11239	33,57	7,2
4	360,39	372,39	3178	9,49	10,71
5	373,54	385,54	620	1,85	20,24
6	386,7	398,7	2251	6,72	13,21
7	399,86	411,86	2493	7,45	10,6
8	413,01	425,01	1758	5,25	7,75
9	426,17	438,17	1771	5,29	14,71
10	439,32	451,32	506	1,51	19,45

TABLEAU 6.21 – Apprentissage de la classe C2 de la saison de l’hiver : relation entre le nombre de points et le RMSE, pour chaque intervalle de l’histogramme de fCO₂. * nombre de points ** pourcentage de données par rapport au nombre total de données de la BD d’apprentissage

6.2.5.2 Validation

Les résultats de l’étape de validation sont présentés dans le **Tableau 6.22** et la **Figure 6.25**. Le RMSE global obtenu est de 10,15 μatm , avec un biais de 0,02 μatm , un coefficient R de 0,95 et un R² de 90%.

Année	RMSE (μatm)	Biais (μatm)	NP *	PD (%) **
2002	17,48	-3,63	6	0,09
2003	6,56	-2,26	115	1,72
2004	4,95	-1,56	606	9,06
2005	8,05	-2,47	456	6,81
2006	12,13	-3,79	972	14,52
2007	7,75	5,91	129	1,93
2008	6,66	-5,2	289	4,32
2009	10,55	-0,84	525	7,85
2010	10,34	4,12	606	9,06
2011	11,04	3,33	709	10,59
2012	9,17	-1,09	1148	17,15
2013	12,19	2,67	754	11,27
2014	12,24	4,58	377	5,63
Total	10,15	0,02	6692	/

TABLEAU 6.22 – Validation de la classe C2 de la saison de l’hiver : RMSE et biais correspondants à chaque année. * nombre de points de données ** pourcentage de données par rapport au nombre total de données de la BD de validation

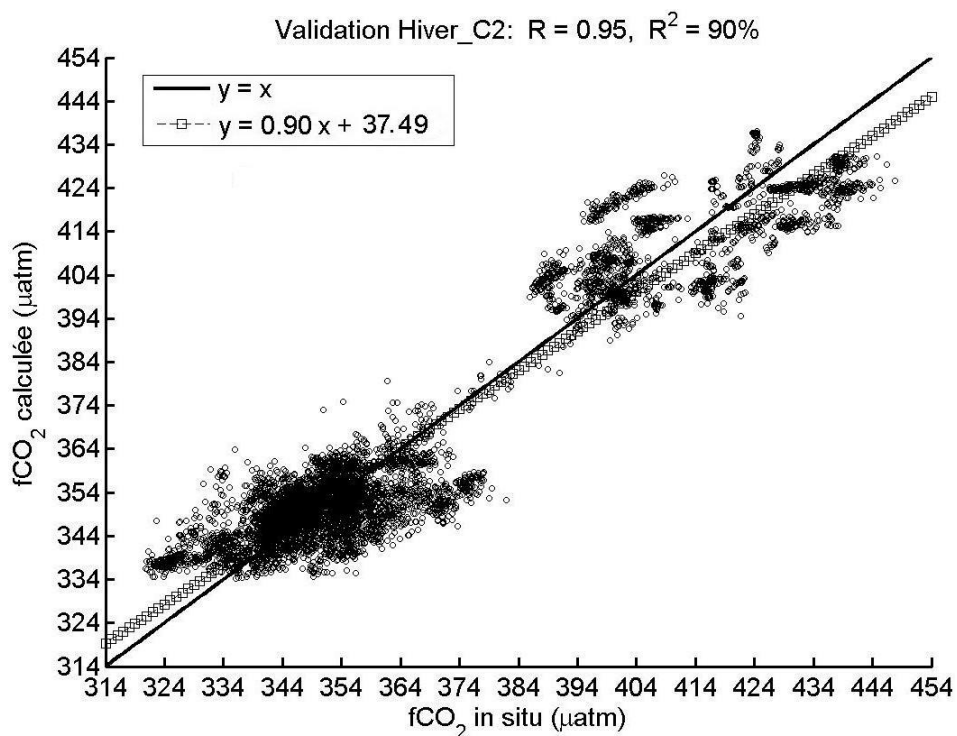


FIGURE 6.25 – Validation de la classe C2 de la saison de l’hiver : fCO₂ estimée par le RN en fonction de fCO₂ *in situ*.

6.2.5.3 Test

Les résultats de la phase de test de cette classe sont présentés dans le **Tableau 6.23** et la **Figure 6.26**. La valeur du RMSE obtenu est de 10,7 µatm avec un biais de -1,12 µatm, un coefficient R de 0,93 et un coefficient R² de 86%.

Année	fCO ₂ (µatm)		SST (°C)		SSS		NP *	PD (%) **
	RMSE	Biais	RMSE	Biais	RMSE	Biais		
2002	9,08	-8,88	0,19	-0,1	0,2	-0,105	7	0,04
2003	7,23	-1,61	0,15	0,07	0,266	0,02	691	3,72
2004	8,07	-6,16	0,2	-0,03	0,48	-0,317	3633	19,57
2005	8,24	-4,13	0,5	0,21	0,373	-0,233	2747	14,8
2006	14,82	-7,71	0,26	0,02	0,638	-0,419	5829	31,39
2007	7,89	5,18	0,16	0,07	0,168	-0,112	775	4,17
2008	7,78	-4,19	0,42	0,05	0,319	0,072	1736	9,35
2009	10,78	-3,4	0,25	-0,01	0,31	-0,15	3149	16,96
2010	10,94	3,54	0,4	-0,04	0,543	0,065	3321	20,2
2011	12,07	6,22	0,43	0,32	0,512	0,331	2944	17,9
2012	7,24	1,08	0,31	0,17	0,568	0,115	5812	35,34
2013	11,85	4,2	0,32	0,21	0,413	0,193	4367	26,56
Total	10,7	-1,12	0,33	0,1	0,493	-0,056	35011	/

TABLEAU 6.23 – Test de la classe C2 de la saison de l’hiver : RMSE et biais (de fCO₂, SST, et SSS) correspondants à chaque année. * nombre de points, ** pourcentage de données par rapport à la BD de test

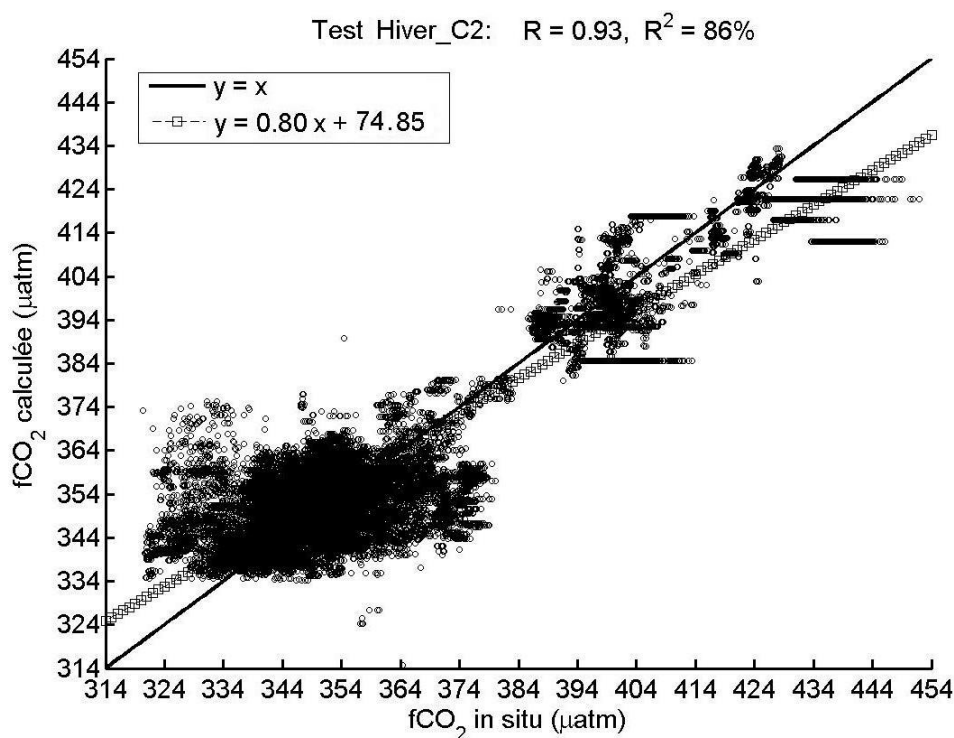


FIGURE 6.26 – Test de la classe C2 de la saison de l’hiver : fCO₂ calculée par le RN à partir des données satellitaires, en fonction de fCO₂ *in situ*.

Comme pour toutes les phases de tests décrites auparavant, les résidus de SST et de

SSS contribuent à l'augmentation du RMSE. Par ailleurs, la relation entre la densité des données et le RMSE reste valide aussi. Cette relation est illustrée dans la **Figure 6.27**.

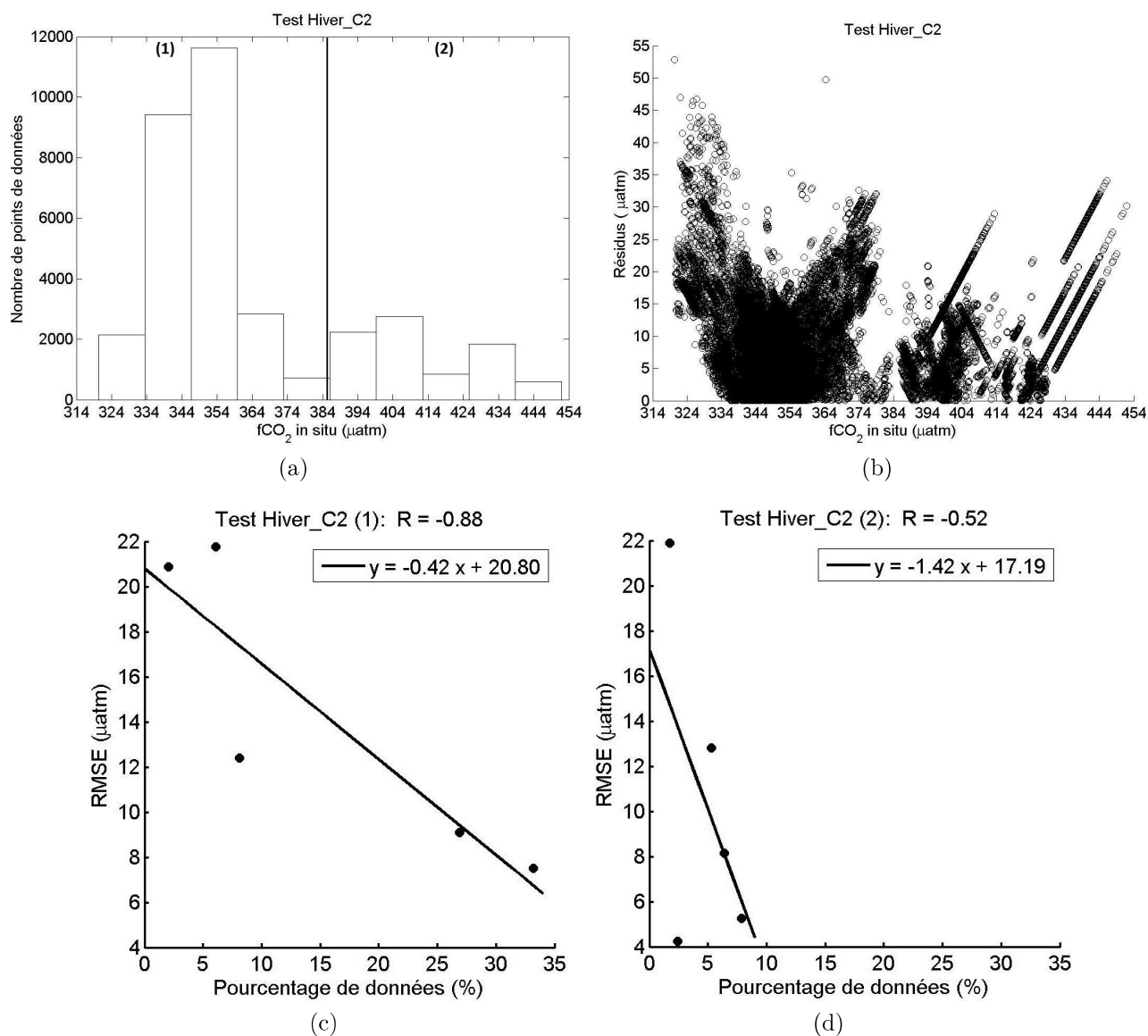


FIGURE 6.27 – Test de la classe C2 de la saison de l'hiver : (a) relation entre les résidus et l'histogramme de $f\text{CO}_2$, (b) relation entre le pourcentage de données (par rapport à la BD de test) de chaque intervalle de l'histogramme, et les RMSE correspondants.

De même que pour la phase d'apprentissage, en raison de la distribution bimodale des données, l'histogramme a été divisé en deux, les relations des deux parties sont présentées dans la **Figure 6.27c** et **6.27d**. Le **Tableau 6.24** contient le détails des valeurs de cette relation.

N°	Borne inf. (μatm)	Borne sup. (μatm)	NP *	PD (%) **	RMSE (μatm)
1	320,92	332,92	2141	6,12	21,75
2	334,08	346,08	9408	26,87	9,1
3	347,23	359,23	11619	33,19	7,5
4	360,39	372,39	2844	8,12	12,41
5	373,54	385,54	726	2,07	20,89
6	386,7	398,7	2230	6,37	8,15
7	399,86	411,86	2750	7,85	5,27
8	413,01	425,01	842	2,4	4,25
9	426,17	438,17	1848	5,28	12,82
10	439,32	451,32	603	1,72	21,89

TABLEAU 6.24 – Test de la classe C2 de la saison de l’hiver : relation entre le nombre de points et le RMSE, pour chaque intervalle de l’histogramme de fCO₂. * nombre de points
** pourcentage de données par rapport au nombre total de données de la BD de test

6.2.6 Hiver : Classe 3 (C3)

Pour cette troisième classe, le RN optimal est celui ayant 7 neurones cachés.

6.2.6.1 Apprentissage

Les résultats de l’étape d’apprentissage de cette classe sont présentés dans le **Tableau 6.25** et la **Figure 6.28**. Le RMSE global obtenu est de l’ordre de 9,9 μatm , avec un faible biais de 0,02 μatm , un coefficient R de 0,91 et un R² de 83%.

Année	RMSE (μatm)	Biais (μatm)	Nombre de points	PD (%) *
2002	11,82	-0,44	3970	10,34
2003	9,83	1,95	3230	8,41
2004	7,15	-4,99	1554	4,05
2005	5,68	-1,66	2763	7,2
2006	7,48	-3,17	895	2,33
2007	5,06	-0,39	2736	7,13
2008	6,13	-2,31	1623	4,23
2009	10,2	-1,47	2984	7,77
2010	9,94	0,44	12684	33,05
2011	13,55	0,26	3665	9,55
2012	19,26	10,68	338	0,88
2013	7,54	4,07	1803	4,7
2014	24,7	23,75	139	0,36
Total	9,9	0,02	38384	/

TABLEAU 6.25 – Apprentissage de la classe C3 de la saison de l’hiver : RMSE et biais correspondants à chaque année. * pourcentage de données par rapport à la taille de la BD d’apprentissage

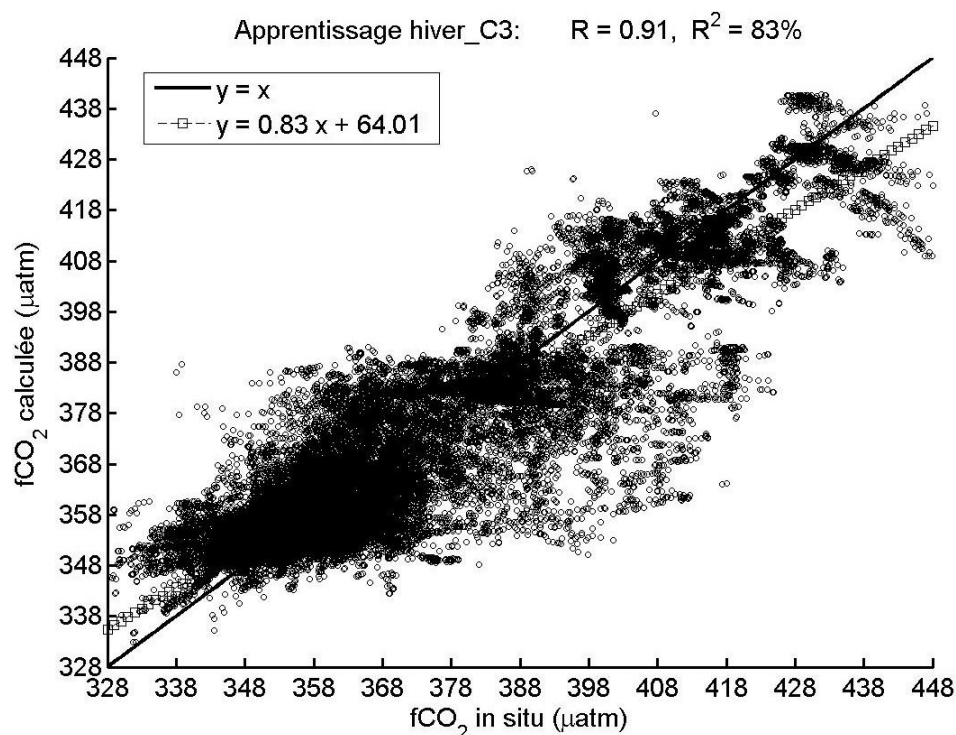


FIGURE 6.28 – Apprentissage de la classe C3 de la saison d’hiver : fCO₂ estimée par le RN en fonction de fCO₂ *in situ*.

Les deux années où le RMSE est le plus large sont 2012 et 2014. Pour l’année 2012, comme on peut le constater sur la **Figure 6.29a**, ces points sont situés à l’embouchure du fleuve Amazone. On remarque dans la **Figure 6.29b** que les valeurs des résidus augmentent avec l’augmentation de la fCO₂. Cette augmentation de fCO₂ est accompagnée d’une augmentation de SST, et d’une forte variabilité de SSS.

Pour l’année 2014, le RMSE est large pour deux raisons : la première est la fCO₂ qui a une moyenne de 440,72 μatm, ce qui est trop élevé comparé à la moyenne générale de fCO₂ pour cette classe, qui est elle de 370,86 μatm. La deuxième est le faible pourcentage de données que contient cette année (0,36, **Tableau 6.25**). En effet, les points où l’erreur dépasse 20 μatm sont associés aux valeurs de fCO₂ supérieures à 436 μatm. Ces points représentent 96% des données pendant cette année, cependant, ils ne représentent que 0,8% de la BD globale d’apprentissage.

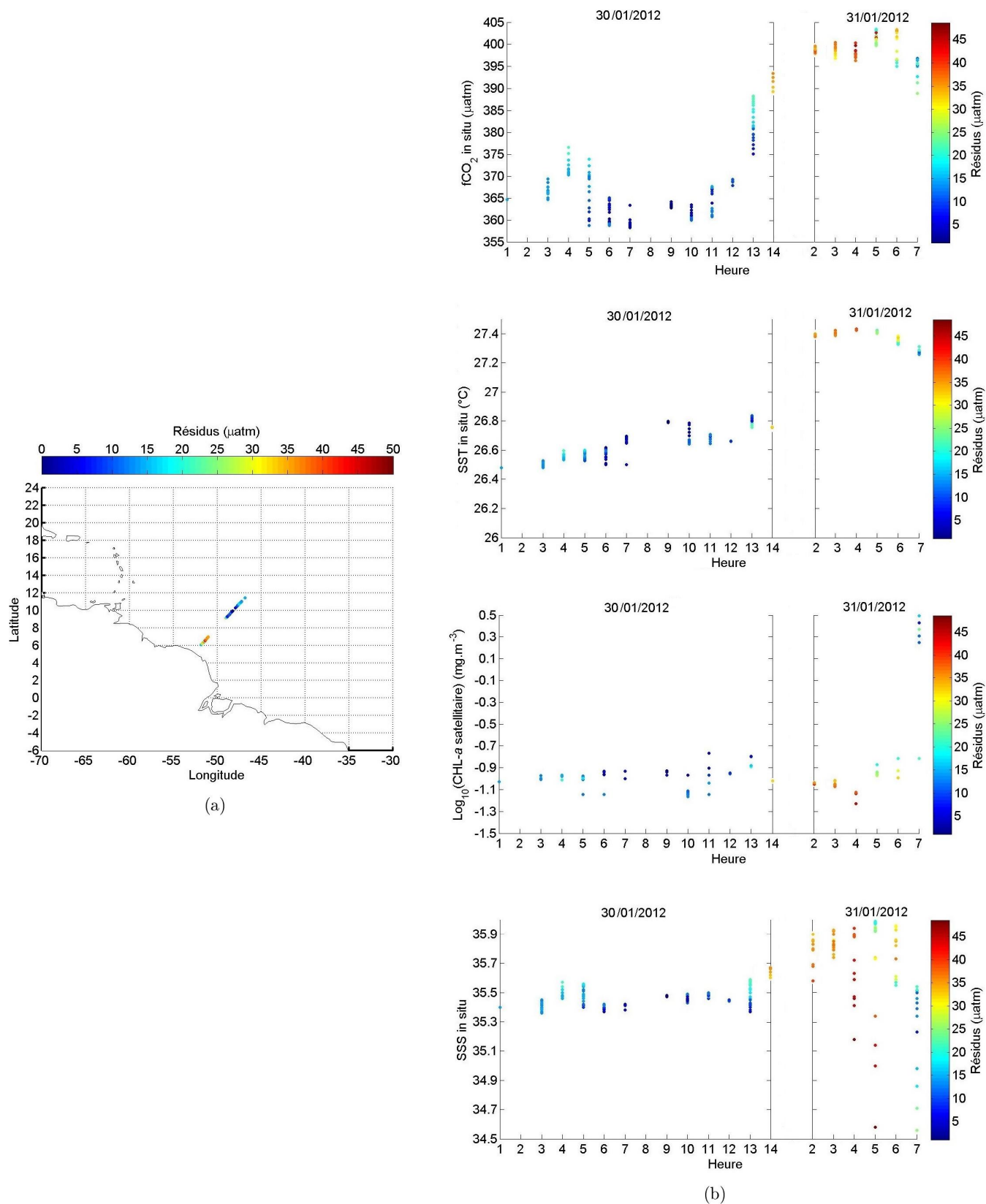


FIGURE 6.29 – Apprentissage de la classe C3 de la saison de l'hiver, entre le 30 et le 31 janvier 2012 : (a) positions géographiques des points, (b) variabilités de $f\text{CO}_2$, SST, Chl-a, et de SSS en fonction des heures de la journée.

La **Figure 6.30** illustre bien la relation entre le nombre de données et les valeurs des résidus. Le RMSE ainsi que le pourcentage de données associé à chaque intervalle de l'histogramme (**Figure 6.30a**) est présenté dans le **Tableau 6.26**.

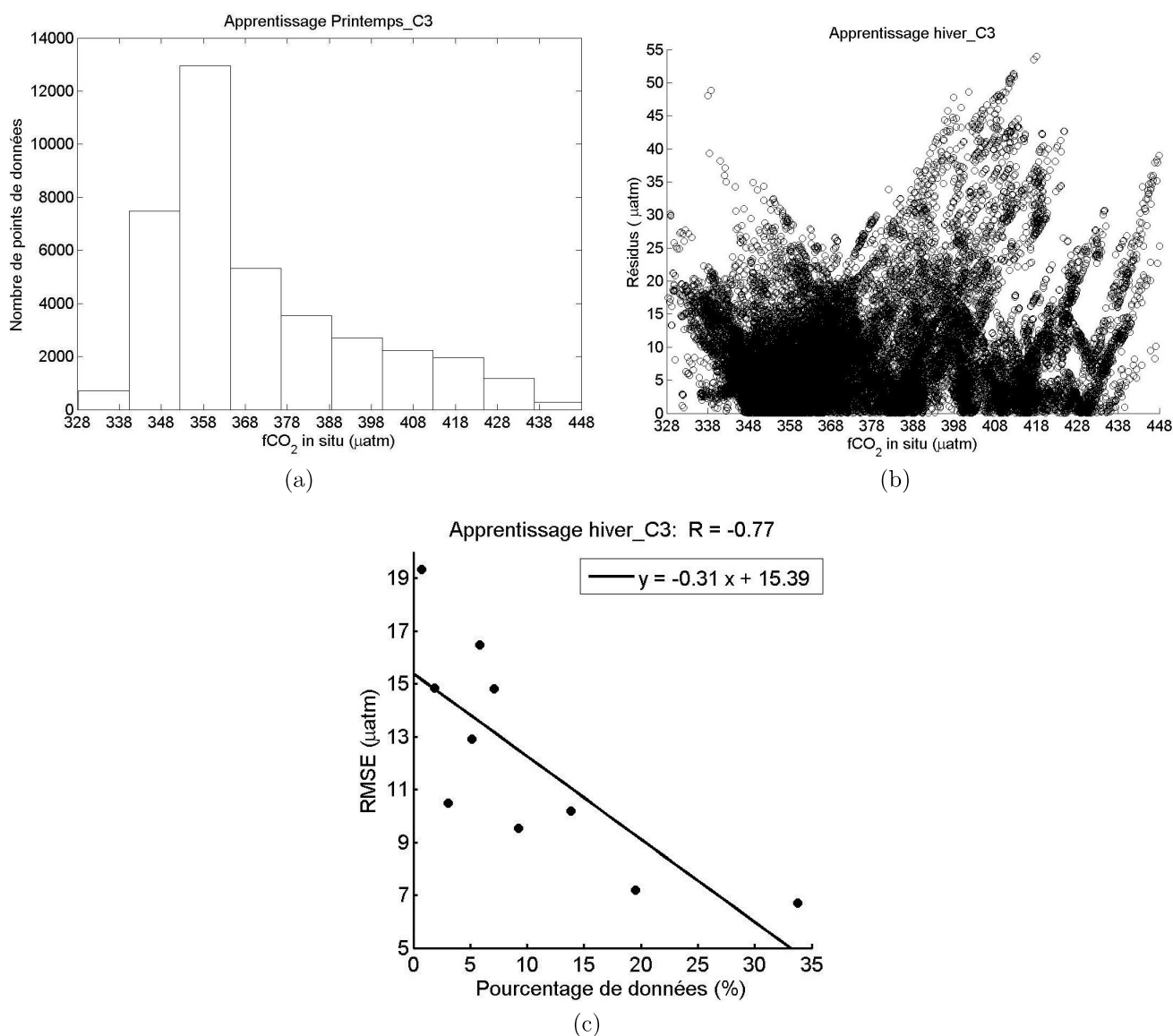


FIGURE 6.30 – Apprentissage de la classe C3 de la saison de l'hiver : (a) *histogramme de fCO_2* , (b) *relation entre les résidus et les valeurs de fCO_2* , (c) *relation entre le pourcentage de données (par rapport à la BD d'apprentissage) de chaque intervalle de l'histogramme, et les RMSE correspondants.*

N°	Borne inf. (μatm)	Borne sup. (μatm)	NP *	PD (%) **	RMSE (μatm)
1	328,33	340,33	707	1,84	14,84
2	340,38	352,38	7482	19,49	7,2
3	352,43	364,43	12953	33,75	6,69
4	364,48	376,48	5322	13,87	10,19
5	376,53	388,53	3536	9,21	9,55
6	388,58	400,58	2713	7,07	14,8
7	400,63	412,63	2239	5,83	16,48
8	412,68	424,68	1970	5,13	12,91
9	424,73	436,73	1181	3,08	10,48
10	436,78	448,78	281	0,73	19,33

TABLEAU 6.26 – Apprentissage de la classe C3 de la saison de l’hiver : relation entre le nombre de points et le RMSE, pour chaque intervalle de l’histogramme de fCO₂. * nombre de points ** pourcentage de données par rapport au nombre total de données de la BD d’apprentissage

6.2.6.2 Validation

Les résultats de l’étape de validation sont présentés dans le **Tableau 6.27** et la **Figure 6.31**. Les valeurs retrouvées sont très proches de celles de l’apprentissage, un RMSE de 9,88 μatm a été obtenu, associé à un faible biais de 0,01, un coefficient R de 0,91, et un R² de 83%.

Année	RMSE (μatm)	Biais (μatm)	NP *	PD (%) **
2002	11,83	-0,56	796	10,37
2003	9,6	1,9	646	8,41
2004	7,24	-4,99	311	4,05
2005	5,76	-1,68	555	7,23
2006	7,41	-3,69	178	2,32
2007	5,01	-0,33	548	7,14
2008	6,21	-2,34	324	4,22
2009	10,28	-1,55	597	7,78
2010	9,94	0,51	2533	32,99
2011	13,47	0,28	733	9,55
2012	19,63	10,55	69	0,9
2013	7,53	3,99	360	4,69
2014	23,46	22,7	27	0,35
Total	9,88	0,01	7677	/

TABLEAU 6.27 – Validation de la classe C3 de la saison de l’hiver : RMSE et biais correspondants à chaque année. * nombre de points de données ** pourcentage de données par rapport au nombre total de données de la BD de validation

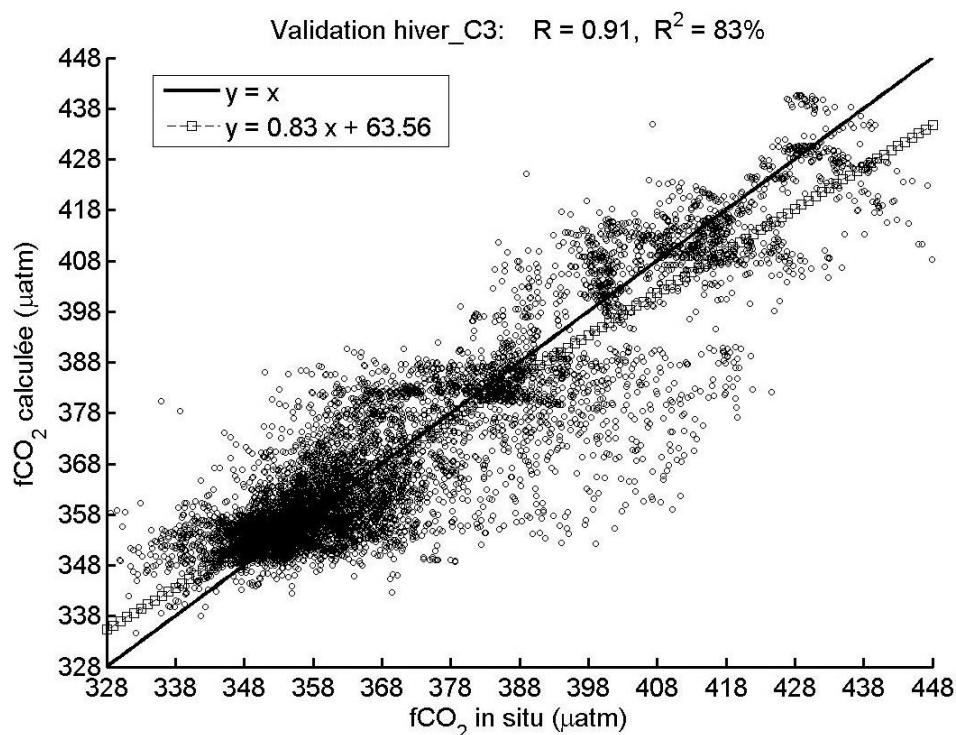


FIGURE 6.31 – Validation de la classe C3 de la saison d’hiver : fCO₂ calculée par le RN en fonction de fCO₂ *in situ*.

6.2.6.3 Test

Les résultats de la partie test de cette classe sont présentés dans le **Tableau 6.28** et la **Figure 6.32**. Le RMSE global obtenu est de l’ordre de 12,77 µatm, avec un biais de 1,55 µatm, un coefficient R de 0,87 et un R² de 76%. On tient à souligner que les RMSEs de SSS de SMOS (entre 2010 et 2013) sont très larges, ce qui affecte largement le RMSE de fCO₂ (~2 µatm). En effet, en appliquant un seuil de 0,4 sur les résidus de SSS, on obtient un RMSE global sur la fCO₂ d’une valeur de 10,19 µatm. En outre, en appliquant ce seuil de 0,4 séparément sur les années de 2010 à 2013, on obtient des RMSEs de : 11,85, 12,48, 12,85, et 8,28 µatm respectivement.

Année	fCO ₂ (µatm)		SST (°C)		SSS		NP *	PD (%) **
	RMSE	Biais	RMSE	Biais	RMSE	Biais		
2002	13.01	11,04	0,75	0,73	0,318	-0,144	660	3,37
2003	11.64	4,84	0,36	0,25	0,358	0,02	3876	19,77
2004	9.05	-6,22	0,32	0,22	0,596	-0,444	1865	9,51
2005	6.66	-1,43	0,37	0,23	0,245	-0,065	3318	16,93
2006	8.43	0,39	0,65	0,53	0,436	-0,26	1073	5,47
2007	7.49	3,54	0,39	0,29	0,281	-0,232	3284	16,75
2008	6.03	0,32	0,67	0,63	0,461	-0,296	1947	9,93
2009	12.41	0	0,47	0,27	0,464	0,016	3581	18,27
2010	13.56	1,09	0,37	0,15	0,802	0,432	11939	69,98
2011	20.97	3,53	0,68	0,62	0,726	0,011	3169	18,58
2012	26.96	12,9	0,44	0,2	0,96	-0,032	407	2,39
2013	13.39	3,23	0,35	0,28	0,799	0,445	1545	9,06
Total	12.77	1,55	0,45	0,29	0,613	0,089	36664	/

TABLEAU 6.28 – Test de la classe C3 de la saison de l’hiver : RMSE et biais (de fCO₂, SST, et SSS) correspondants à chaque année. * nombre de points, ** pourcentage de données par rapport à la BD de test

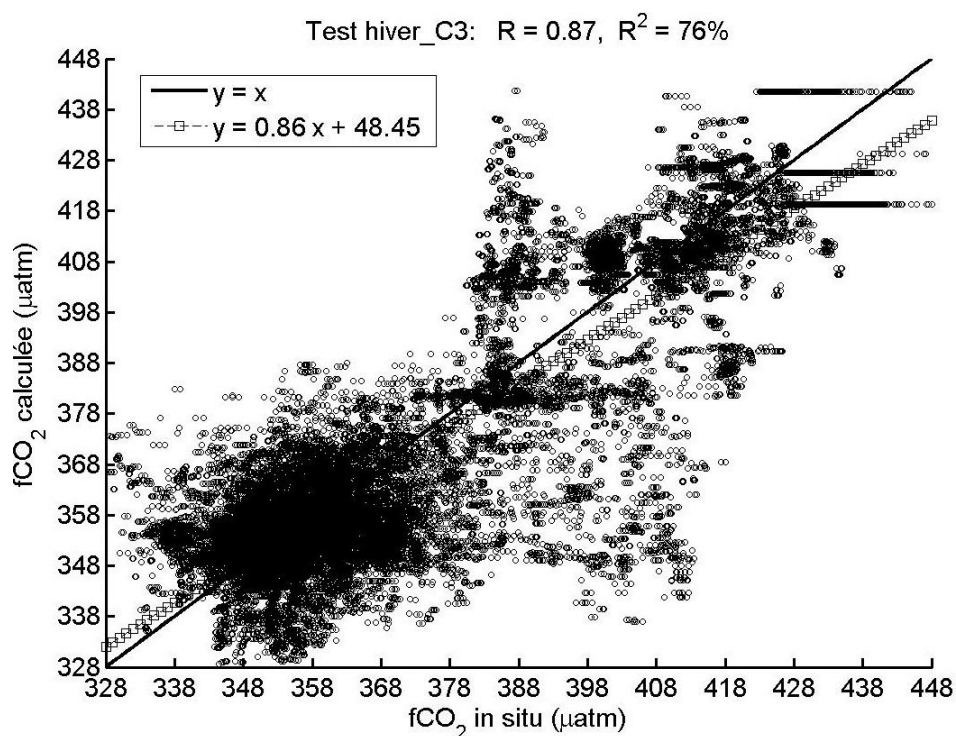


FIGURE 6.32 – Test de la classe C3 de la saison de l’hiver : fCO₂ calculée par le RN à partir des données satellitaires, en fonction de fCO₂ *in situ*.

De même que pour les classes précédentes, on présente dans la **Figure 6.33** la relation

qui existe entre la densité des valeurs de $f\text{CO}_2$ ainsi que les résidus du modèles. Le détail des valeurs est présenté dans le **Tableau 6.29**.

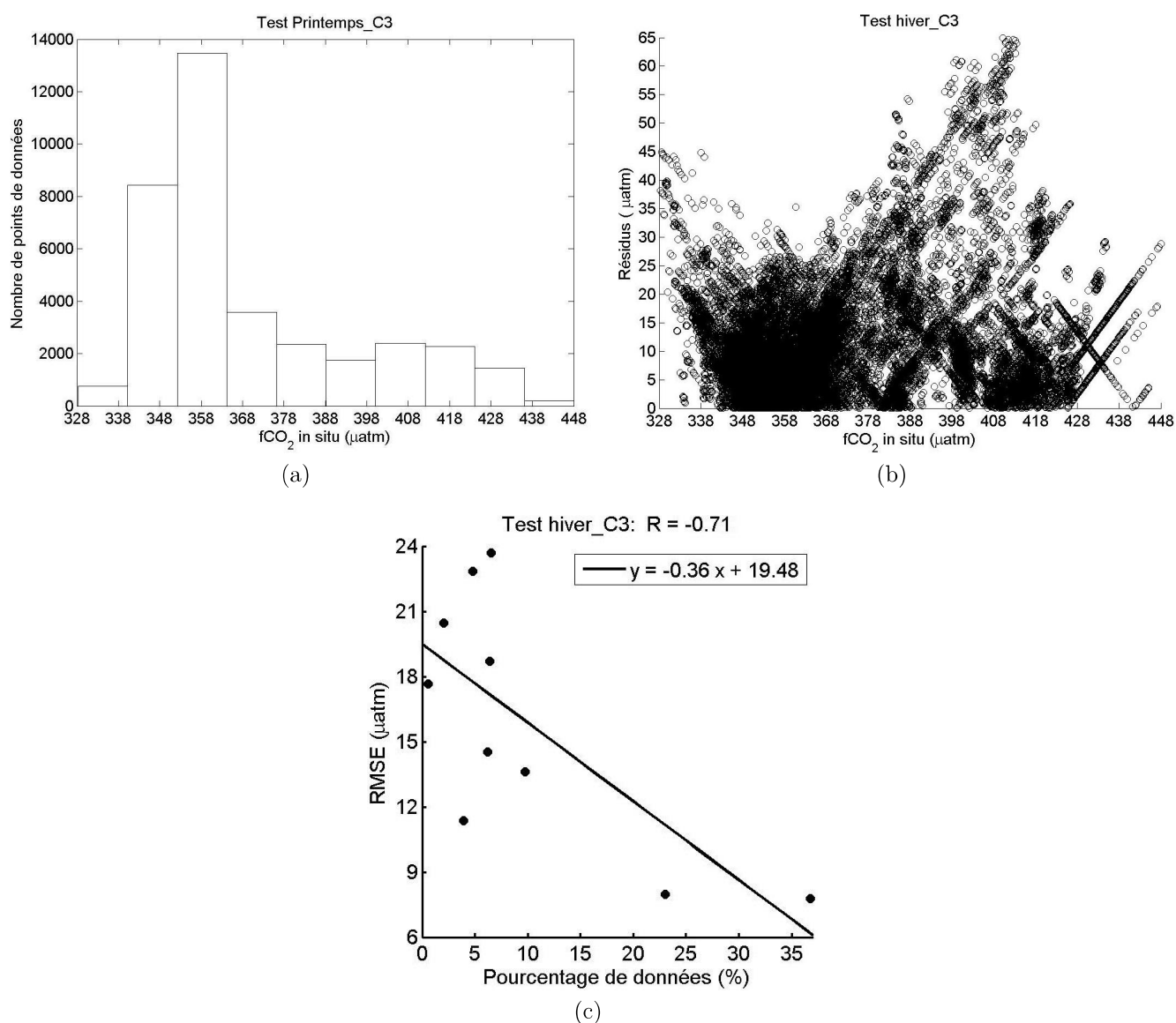


FIGURE 6.33 – Test de la classe C3 de la saison de l'hiver : (a) relation entre les résidus et l'histogramme de $f\text{CO}_2$, (b) relation entre le pourcentage de données (par rapport à la BD de test) de chaque intervalle de l'histogramme, et les RMSE correspondants.

N°	Borne inf. (μatm)	Borne sup. (μatm)	NP *	PD (%) **	RMSE (μatm)
1	329,29	339,29	752	2,05	20,46
2	341,27	351,27	8437	23,01	7,98
3	353,25	363,25	13471	36,74	7,79
4	365,23	375,23	3584	9,78	13,63
5	377,21	387,21	2350	6,41	18,69
6	389,19	399,19	1758	4,79	22,83
7	401,17	411,17	2405	6,56	23,7
8	413,15	423,15	2268	6,19	14,53
9	425,13	435,13	1433	3,91	11,36
10	437,11	447,11	206	0,56	17,67

TABLEAU 6.29 – Test de la classe C3 de la saison de l’hiver : relation entre le nombre de points et le RMSE, pour chaque intervalle de l’histogramme de fCO₂. * nombre de points
** pourcentage de données par rapport au nombre total de données de la BD de test

Chapitre 7

ÉLABORATION DES CARTES DE $f\text{CO}_2$

Après avoir créé tous les RNs de toutes les saisons, on présente dans ce chapitre les résultats finaux de cette thèse, à savoir l'élaboration des fichiers 'CSV' (*Comma-separated values*) pour la création des cartes de $f\text{CO}_2$ à partir des données satellitaires.

Les données mensuelles de $f\text{CO}_2$ produites couvrent tout l'océan Atlantique tropical (de 23,5°S-70°O à 23,5°N-15°E) sur une couverture temporelle allant de 2002 à 2013. Pour ceci, deux principales étapes ont été suivies, la première est la structuration des données satellitaires en un seul tableau pour chaque mois, et la deuxième est le calcul de $f\text{CO}_2$ en utilisant les RNs créés.

Afin de réaliser la première étape, comme vu dans le schéma **1** de la **Figure 7.1**, les données de SST et Chl-*a* de MODIS-A ont déjà été combinées au préalable (**Section 5.4**, code MATLAB : **Annexe A.6**). Ainsi, puisque les données satellitaires de SSS sont indisponibles pour les mois allant de juillet 2002 à décembre 2009, on a utilisé la SSS de SODA (**Section 5.3**) pour ces mois et la SSS de SMOS pour le reste (de janvier 2010 à décembre 2013). Ces données sont donc combinées avec la SST et la Chl-*a* suivant le code MATLAB de l'**Annexe A.16.1**. Ceci produit un nombre total de 138 matrices, chacune d'elle correspond à un seul mois et contient : **longitude - latitude - SST - Chl-*a* - SSS**.

La deuxième étape (**2** dans la **Figure 7.1**) représente le calcul de $f\text{CO}_2$ pour chaque matrice produite dans la première étape. Pour ceci, l'algorithme suivant a été appliqué pour chaque saison :

1. Suivant la saison, récupérer les RNs de toutes les classes,
2. Pour chaque classe, récupérer les domaines de validation des paramètres (préalablement créés et enregistrés ; **Annexe A.13.1**),
3. Pour chaque mois de chaque année, récupérer la matrice des données satellitaires créée dans l'étape **1**,
 - Utiliser les domaines de validation pour diviser la matrice en 2 ou 3 parties (suivant le nombre de classe de la saison),
 - Pour chaque partie, calculer la $f\text{CO}_2$ en utilisant le RN correspondant,
 - Combiner les parties en une seule matrice.

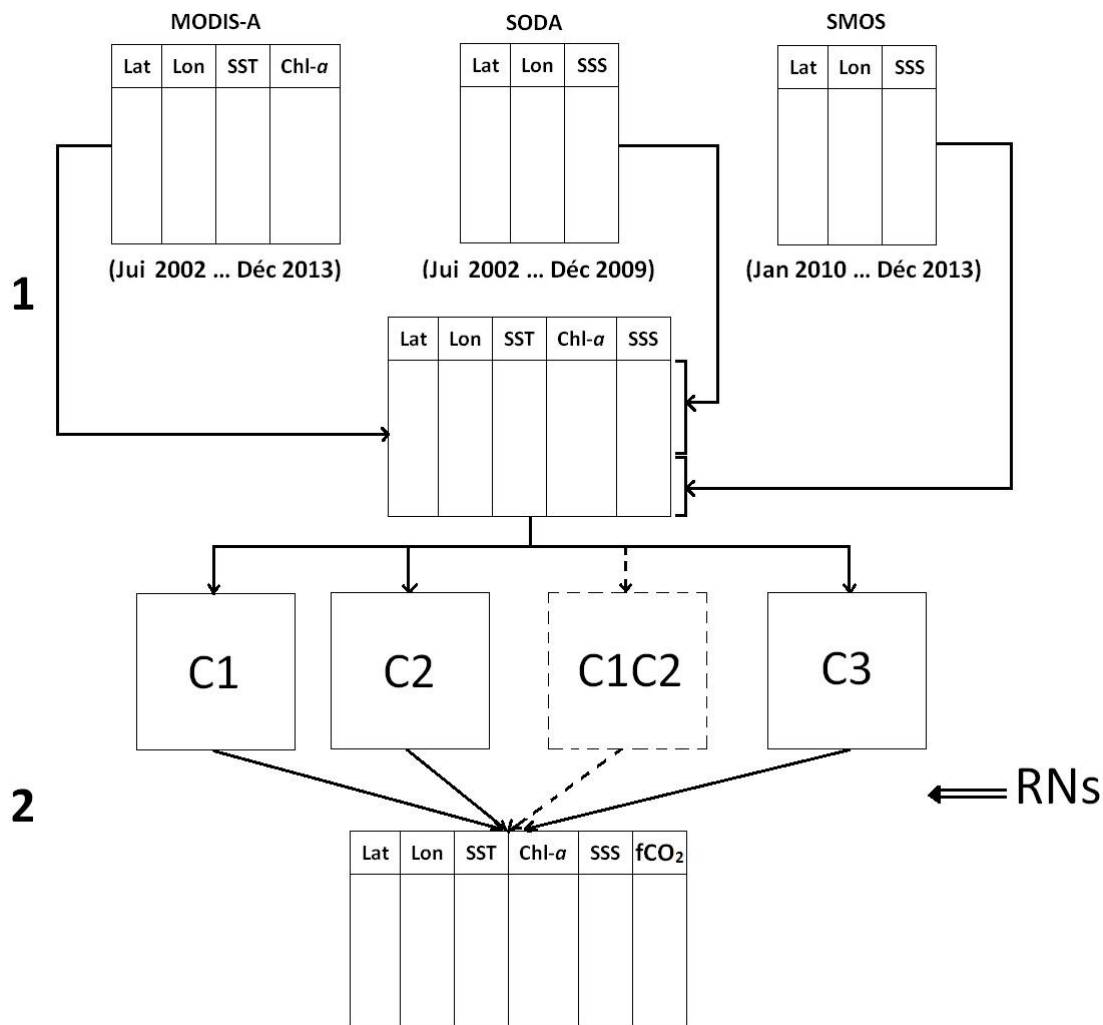


FIGURE 7.1 – Les deux principales étapes suivies pour le calcul de fCO_2 à partir des données satellitaires (et assimilées).

Le code MATLAB permettant le déroulement de cet algorithme est présenté dans l'Annexe A.16.2.

À la fin de ces étapes, pour les 138 matrices, une colonne a été ajoutée correspondant à la fCO_2 . Pour chacune de ces matrices, le nombre de points avoisine ou dépasse le 1 million, et le nombre total des points de toutes les matrices est égal à 194330767, distribués sur les quatre saisons (voir **Tableau 7.1**).

Saison	Nombre de points
Hiver	34691142
Printemps	34935408
Été	36411219
Automne	37808355

TABLEAU 7.1 – Nombre de points de données propre à chaque saison, pour la période allant de juillet 2002 à décembre 2013.

La **Figure 7.2** illustre les moyennes mensuelles des quatre paramètres (SST, SSS, Chl-*a*, et $f\text{CO}_2$) sur toute la zone Atlantique tropical, entre juillet 2002 et décembre 2013. On remarque que pour les quatre paramètres, les moyennes mensuelles sont quasiment stables d'une année à une autre. On note une légère diminution des moyennes de SSS à partir de l'année 2010 (données de SMOS), accompagné d'une légère diminution des moyennes de Chl-*a* et de $f\text{CO}_2$.

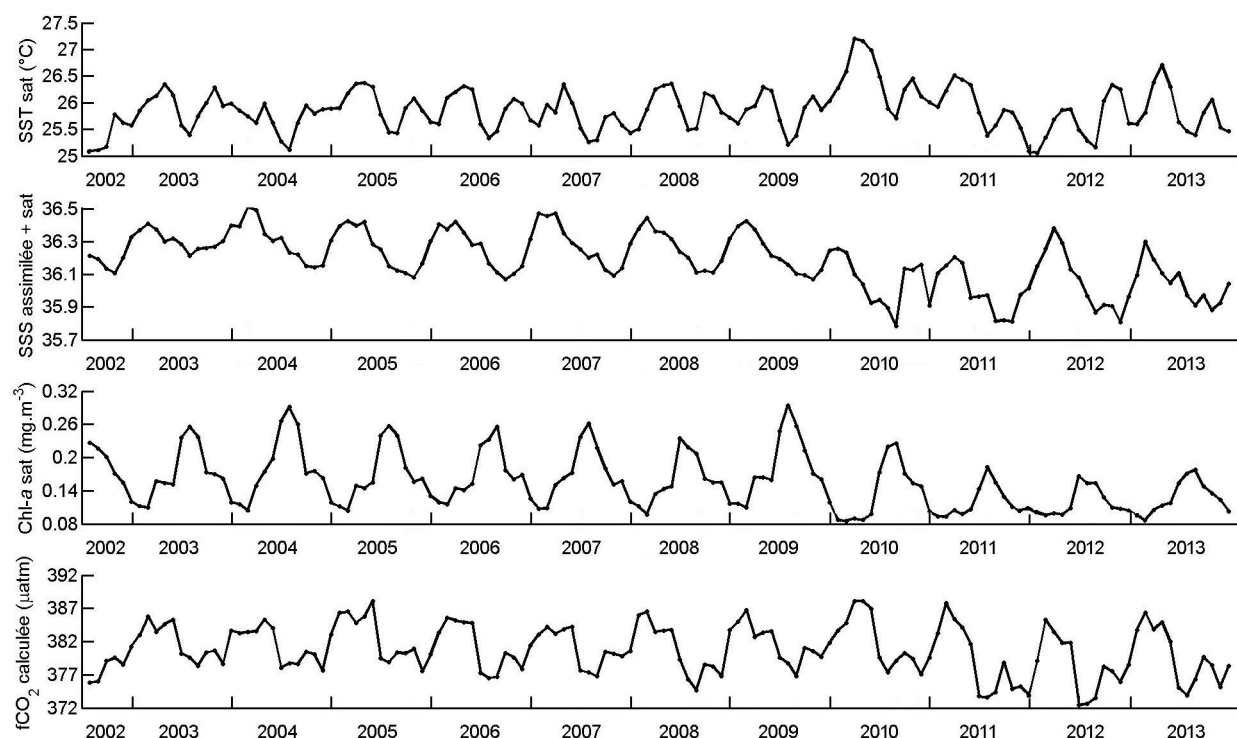


FIGURE 7.2 – Moyennes mensuelles de SST, SSS, Chl-*a*, et $f\text{CO}_2$ entre juillet 2002 et décembre 2013.

Afin de mieux voir le cycle annuel de chaque paramètre, on présente dans les deux **Figures 7.3** et **7.4** la variabilité de chaque paramètre durant chaque l'année.

À partir de ces figures on peut noter plusieurs remarques, la première est que pour toutes les années, la $f\text{CO}_2$ est plus élevée en hiver (où elle atteint son maximum) et au printemps, et est plus faible en été (où elle atteint son minimum) et en automne. La SST quand à elle, augmente pendant le printemps et l'automne, et diminue pendant l'hiver et

l'été. Pour la $\text{Chl-}a$, on remarque qu'elle commence à augmenter à partir du printemps, elle atteint son maximum en été, elle commence ensuite à diminuer jusqu'à ce qu'elle atteigne son minimum en hiver. Pour la SSS, elle atteint son maximum entre l'hiver et le printemps, ensuite elle commence à diminuer jusqu'à ce qu'elle atteigne son minimum en automne.

Les autres remarques concernent la variabilité de la $f\text{CO}_2$ par rapport aux autres paramètres. En effet, il est très visible que la $f\text{CO}_2$ est très influencée par la SST et la $\text{Chl-}a$. On peut citer quatre cas de variabilité qui se répètent quasiment dans toutes les années :

1. Quand la SST augmente et la $\text{Chl-}a$ augmente, la $f\text{CO}_2$ diminue (ex. **Figure 7.3e** 2-5/2006),
2. Quand la SST diminue et la $\text{Chl-}a$ diminue, la $f\text{CO}_2$ augmente (ex. **Figure 7.4a** 11-12/2008),
3. Quand la SST augmente et la $\text{Chl-}a$ diminue, la $f\text{CO}_2$ augmente (ex. **Figure 7.3c** 7-9/2004),
4. Quand la SST diminue et la $\text{Chl-}a$ augmente, la $f\text{CO}_2$ diminue (ex. **Figure 7.3f** 5-7/2007).

Pour certains mois, tels que 9-10/2010 et 9-11/2012 (**Figure 7.4c** et **7.4e** respectivement), on remarque que même si la SST augmente et la $\text{Chl-}a$ diminue, la $f\text{CO}_2$ diminue, ceci est visiblement dû à la diminution de SSS pendant ces mois.

Les tableaux contenant les valeurs du minimum, maximum, moyenne, et écart-type de chaque paramètre sont présentées en **Annexes D.1, D.2, D.3, et D.4**.

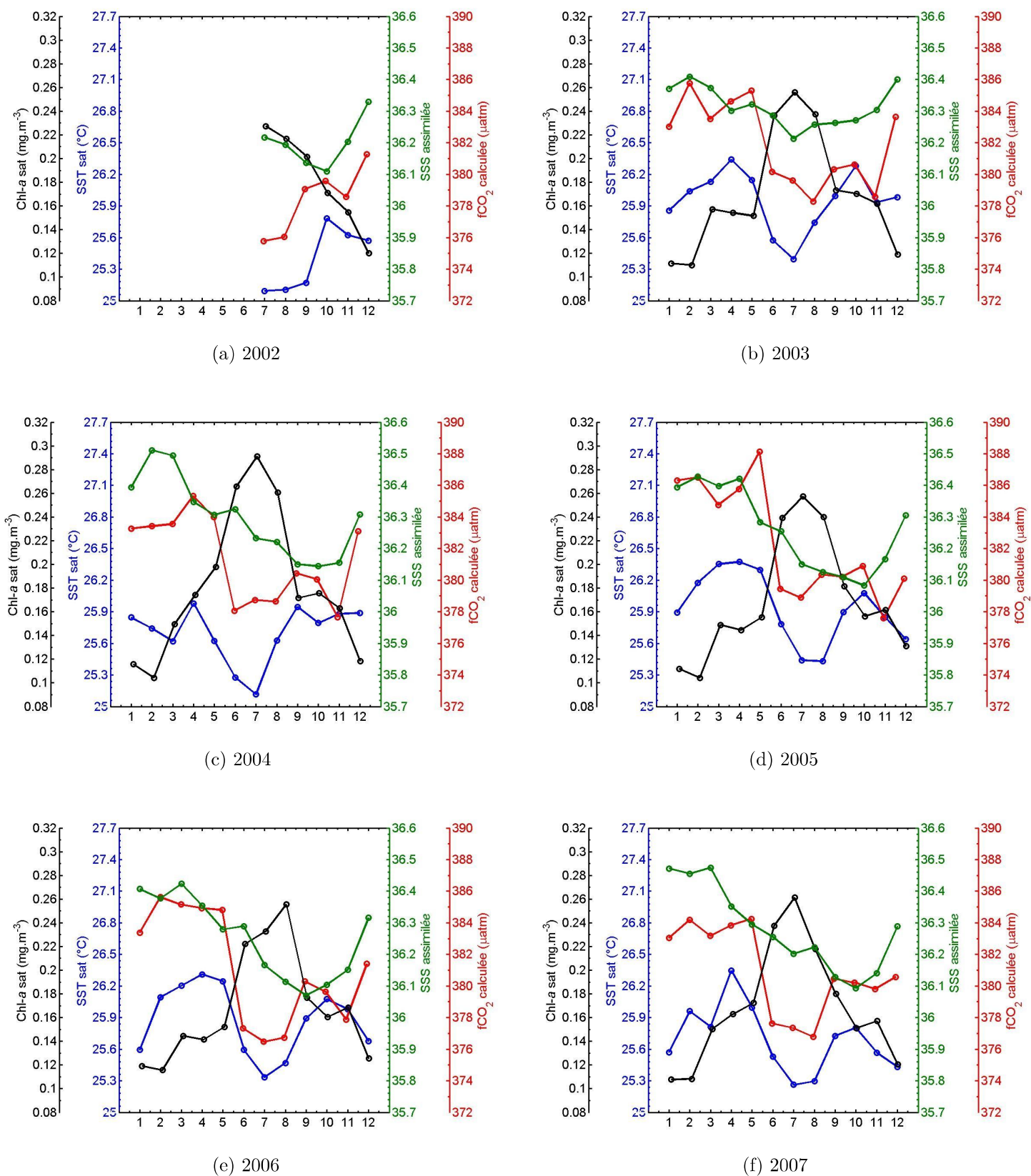


FIGURE 7.3 – Présentation de la variabilité annuelle de SST, SSS, Chl-*a*, et fCO₂, pour les années allant de 2002 à 2007.

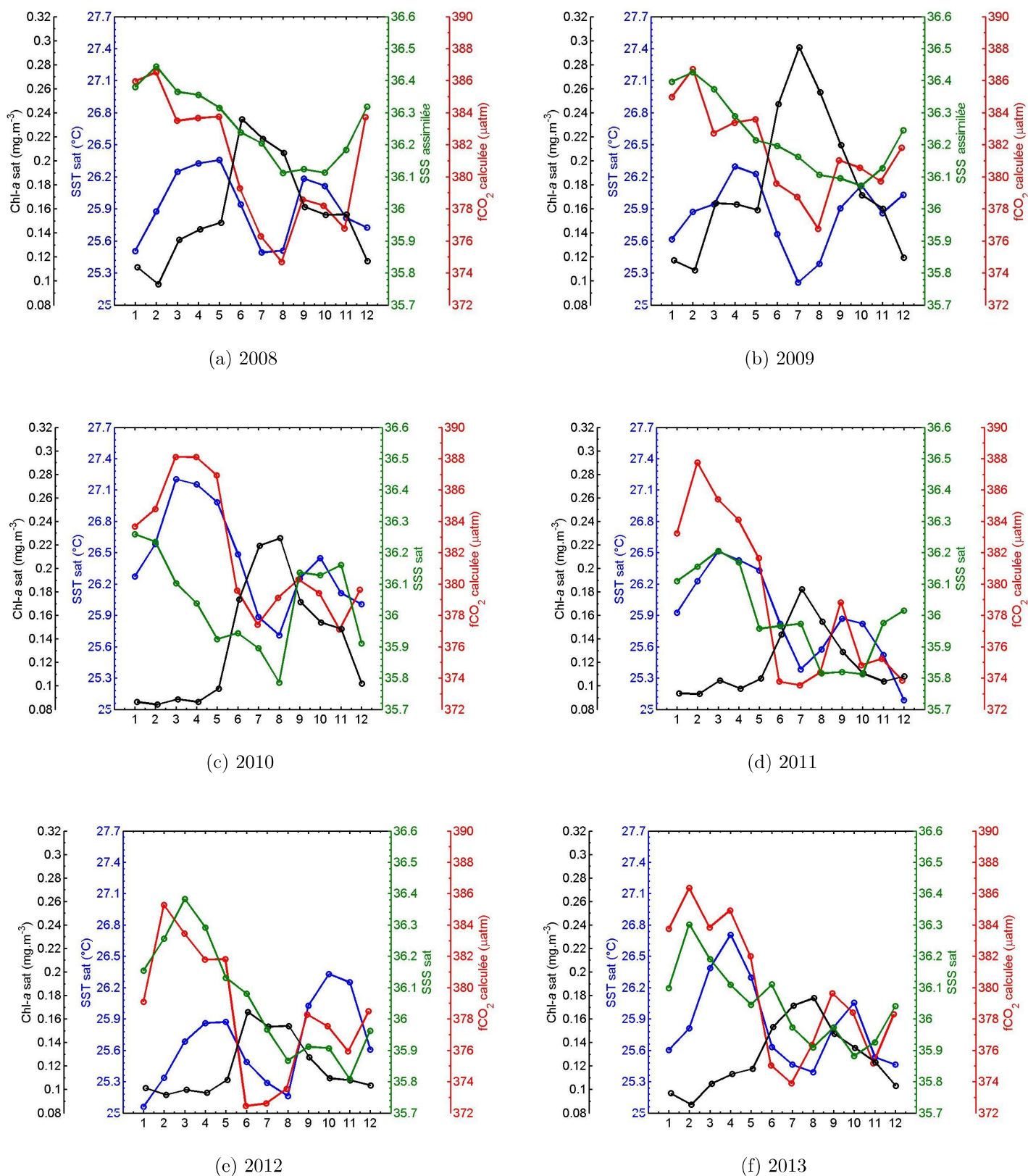


FIGURE 7.4 – Présentation de la variabilité annuelle de SST, SSS, Chl-*a*, et fCO_2 , pour les années allant de 2008 à 2013.

La dernière étape de tout ce traitement de données proposé dans cette étude, se résume dans la production de fichiers de type '.csv' contenant la compilation de toutes ces données. Pour cela, le code MATLAB présenté en **Annexe A.16.3** a été utilisé. Le nombre de fichiers produits est le même que le nombre de matrices, à savoir 138. L'entête des fichiers est la suivante : **ID,latitude,longitude,SST,Chl-a,Log10(Chl-a),latitude,longitude,SSS,fCO₂,depth**. Les premières latitudes et longitudes concernent la SST et la Chl-a, quant-aux deuxièmes, elles concernent la SSS. Le paramètre 'depth' a été ajouté pour permettre l'affichage des cartes via le logiciel 'ODV', ce paramètre a été fixé à 1.

Un extrait des paramètres de l'année 2012/2013 est présenté dans les figures : **7.5** pour la SST, **7.6** pour la Chl-a, **7.7** pour la SSS, et **7.8** pour la fCO₂.

Ces figures illustrent bien comment, suivant les mois de l'année, les paramètres varient ensemble. On note dans ce qui suit les constatations qu'on a pu faire à partir de ces cartes mensuelles :

- En hiver, le gradient de la fCO₂ est très visible entre le Nord et le Sud du bassin ; la fCO₂ est très faible au Nord, et est très forte au Sud. De décembre à février :
 - * au Nord-Est, la SST diminue, la Chl-a est quasiment stable, la SSS augmente légèrement, et la fCO₂ augmente,
 - * au Nord-Ouest, la SST diminue, la Chl-a est stable, la SSS augmente, et la fCO₂ est presque stable,
 - * au Sud-Est, la SST augmente, la Chl-a diminue, la SSS est presque stable, et la fCO₂ augmente,
 - * au Sud-Ouest, la SST augmente, la Chl-a et la SSS sont assez stables, et la fCO₂ diminue légèrement.
- Entre l'hiver et le printemps, on remarque une diminution de la SST au Nord et une augmentation de celle-ci au Sud, la Chl-a augmente au niveau des zones côtières, la SSS diminue à l'Est, et la fCO₂ augmente au Nord et diminue au Sud.
- Au printemps, de mars à mai, on remarque que :
 - * au Nord, tous les paramètres restent pratiquement stables,
 - * au Sud-Est, la SST diminue, la Chl-a augmente, la SSS reste quasiment stable, et la fCO₂ diminue,
 - * au Sud-Ouest, tous les paramètres sont presque stables, sauf la fCO₂ qui subit une légère diminution.
- Entre le printemps et l'été, on remarque que la SST augmente au Nord du bassin, et diminue fortement au Sud, la Chl-a augmente fortement dans la partie centrale, la SSS diminue à l'Ouest et augmente à l'Est, et la fCO₂ augmente au Nord et diminue fortement au Sud.
- En été, de juin à août, on constate que :

* au Nord, la SST augmente, la Chl-*a* ainsi que la SSS sont quasiment stables, la fCO_2 quand-à-elle diminue un peu à l'Est, et augmente légèrement entre 0-10°N,

* au Sud, la Chl-*a* et la SSS restent stables, cependant, il y a une forte diminution de la SST et de la fCO_2 .

— Entre l'été et l'automne, il y a une augmentation de la SST au Nord, et une diminution de celle-ci au Sud, la Chl-*a* diminue au niveau de tout le bassin, la SSS reste presque stable, et la fCO_2 diminue au Nord et augmente au Sud.

— En automne, de septembre à novembre, on remarque que :

* au Nord, la SST diminue, la Chl-*a* diminue aussi mais légèrement, la SSS reste presque stable, et la diminution de la fCO_2 est très remarquable,

* au Sud, la SST augmente, la Chl-*a* diminue, la SSS reste stable là aussi, et la fCO_2 diminue légèrement à l'Est.

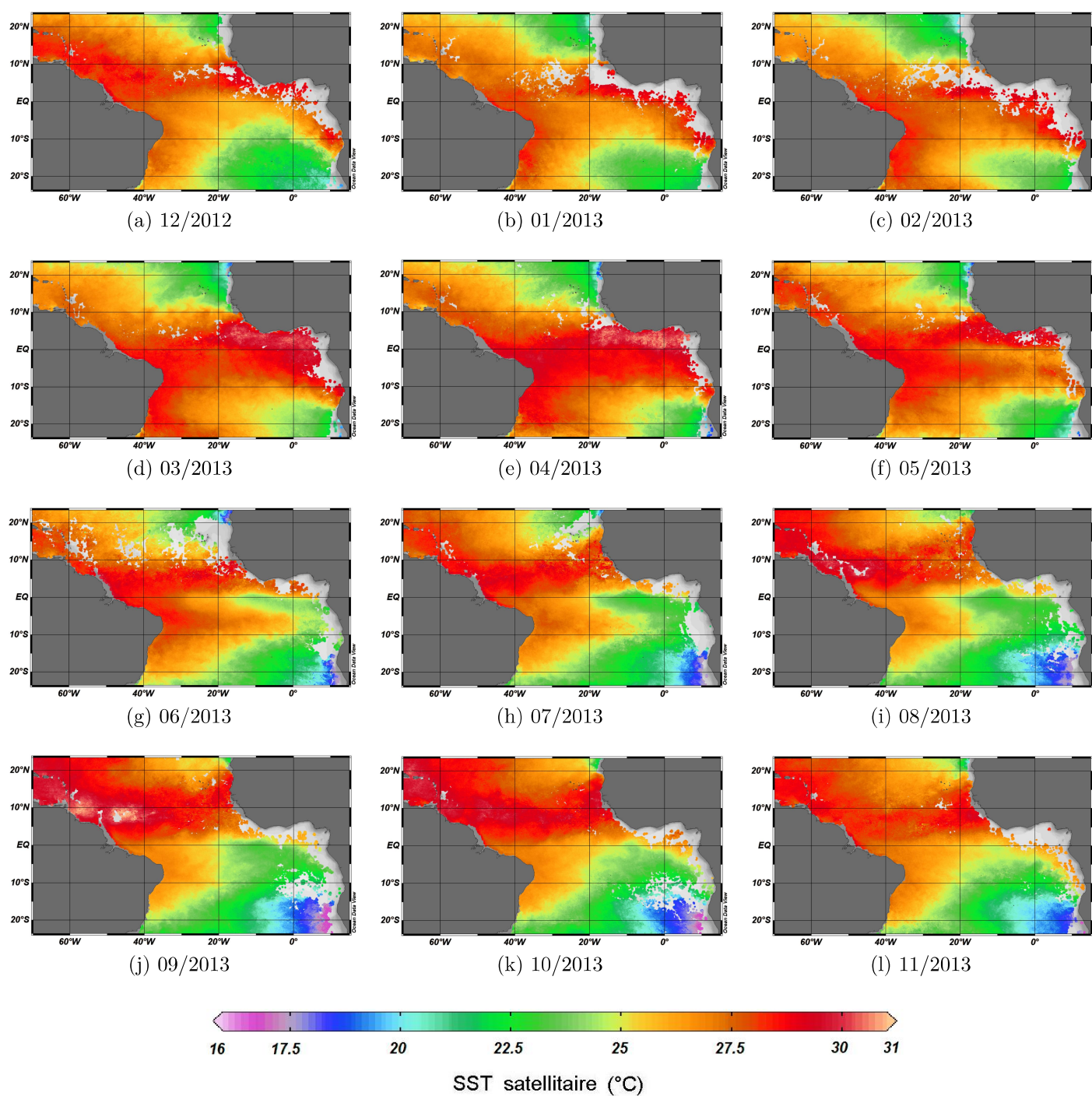


FIGURE 7.5 – Cartes mensuelles de SST satellitaires, fournies par le capteur **MODIS-A**, dans la zone 23,5°S-70°O à 23,5°N-15°E, entre décembre 2012 et novembre 2013, avec une résolution spatiale de 4 km × 4 km.

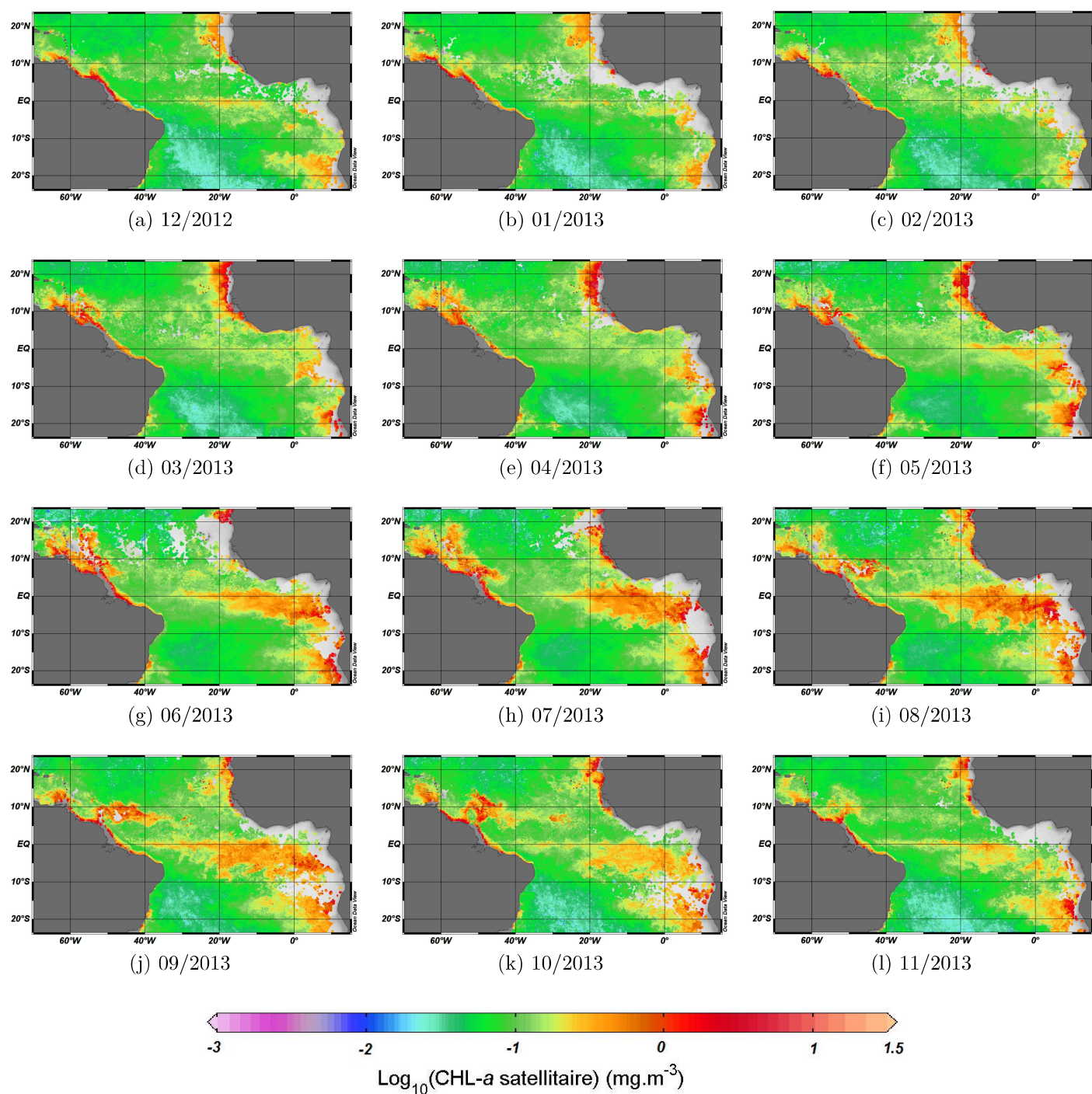


FIGURE 7.6 – Cartes mensuelles de $\text{Log}_{10}(\text{Chl-}a)$ satellitaires, fournies par le capteur **MODIS-A**, dans la zone 23,5°S-70°O à 23,5°N-15°E, entre décembre 2012 et novembre 2013, avec une résolution spatiale de 4 km × 4 km.

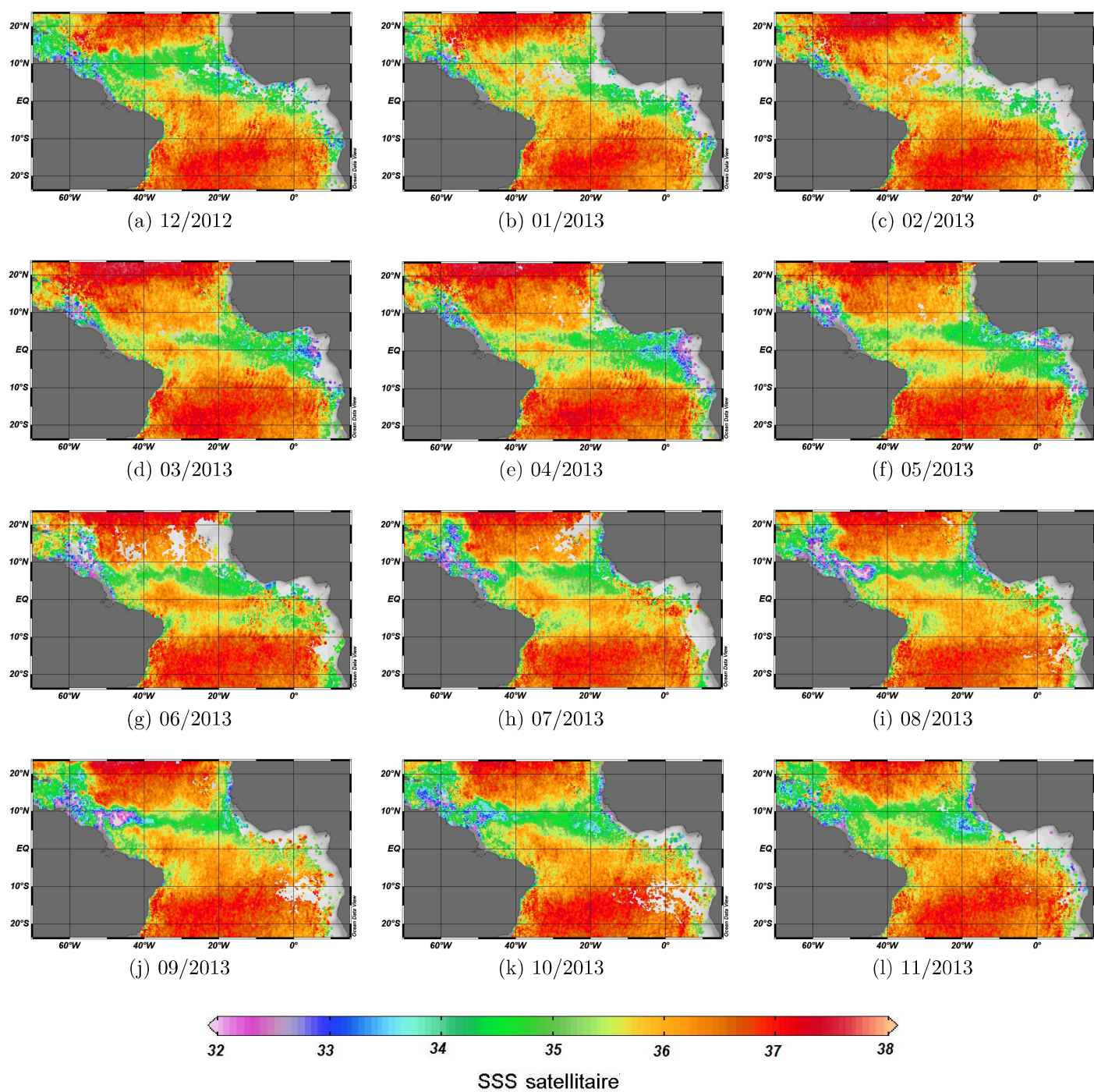


FIGURE 7.7 – Cartes mensuelles de SSS satellitaires, fournies par le capteur **SMOS**, dans la zone $23,5^{\circ}\text{S}$ - 70°O à $23,5^{\circ}\text{N}$ - 15°E , entre décembre 2012 et novembre 2013, avec une résolution spatiale de $25\text{ km} \times 25\text{ km}$.

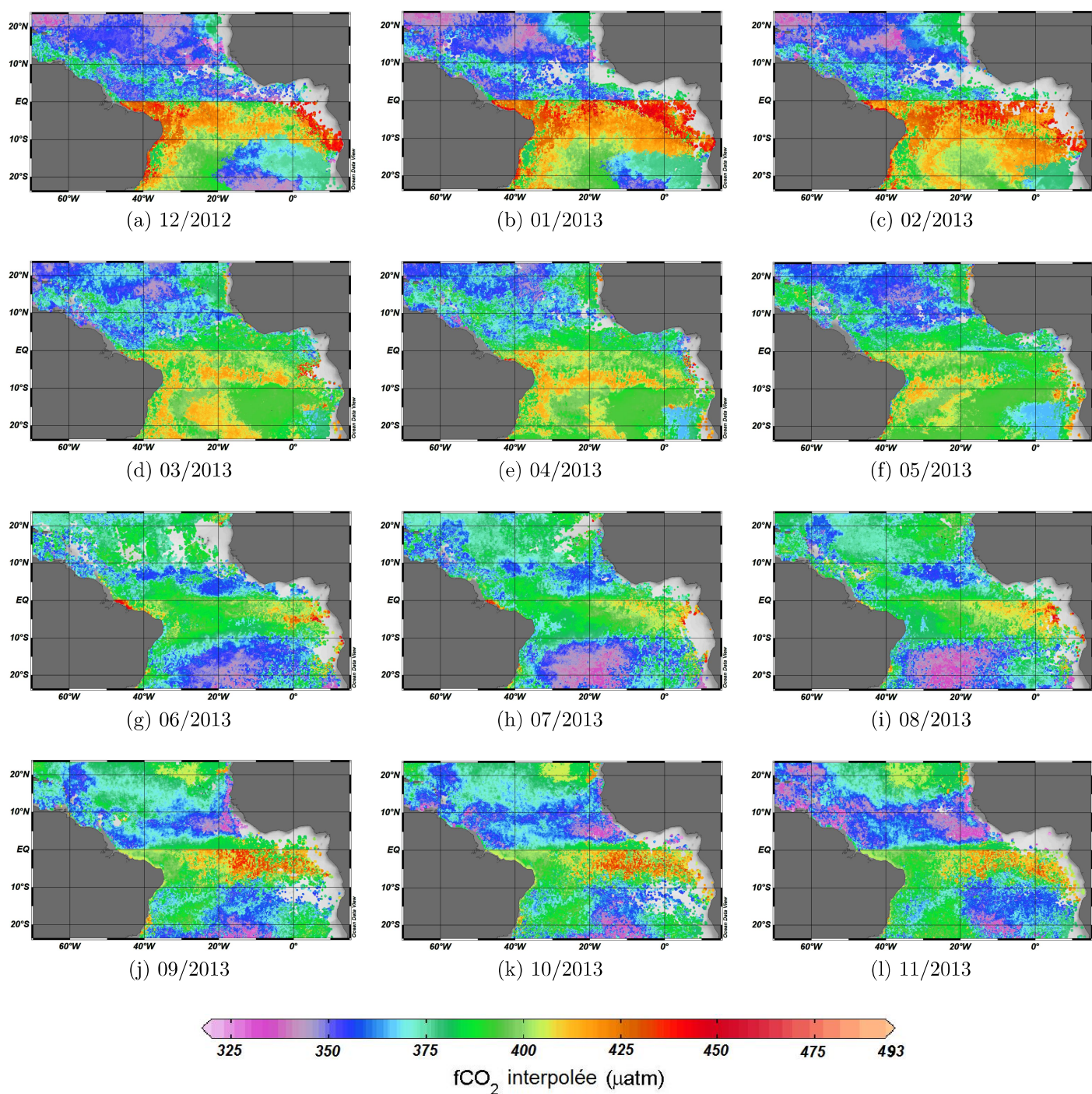


FIGURE 7.8 – Résultats de l'interpolation de fCO_2 mensuelles à partir des données satellitaires, produites par les RNs construits, dans la zone 23,5°S-70°O à 23,5°N-15°E, entre décembre 2012 et novembre 2013, avec une résolution spatiale de 4 km \times 4 km.

CONCLUSIONS ET PERSPECTIVES

L'objectif principale de cette thèse était de confirmer qu'on pouvait profiter de la haute résolution spatiotemporelle des données satellitaires, pour produire des cartes de $f\text{CO}_2$ de surface de bonne qualité, et ainsi permettre de mieux étudier la variabilité de ce paramètre durant tous les mois de l'année. Pour ceci, on a proposé dans cette thèse un ensemble d'outils, de techniques, et de programmes informatiques, qui permettent, en suivant un certain nombre d'étapes, de traiter automatiquement des données océanographiques de différents types (*in situ*, satellitaires, et assimilées).

Dans cette étude, nous nous sommes intéressés à la partie tropicale de l'océan Atlantique (de $23,5^\circ\text{S}$ - 70°O à $23,5^\circ\text{N}$ - 15°E). De par sa taille et son emplacement, ce bassin est caractérisé par des processus fortement dynamiques (**Chapitre 2**), ce qui rend son étude quelque peu particulière.

Le principe général de l'étude, consiste à utiliser les données *in situ* (considérées comme une référence) pour trouver des modèles d'interpolation de $f\text{CO}_2$, à partir d'un ou plusieurs paramètres ayant une certaine influence sur la variabilité de la $f\text{CO}_2$ de surface (**Chapitre 1**). Les paramètres qu'on a choisi sont : SST, SSS, et Chl-*a*. La période sur laquelle on a travaillé, s'étend de l'année 2000 à l'année 2014. Les données *in situ* utilisées ont été extraites de la BD SOCAT V.3 (**Section 2.8**). La méthode qui a été utilisée pour l'interpolation de la $f\text{CO}_2$ est les réseaux de neurones artificiels avec une architecture de type MLP (*multilayer perceptron*) ayant des connexions de type *feedforward*, et un apprentissage supervisé dit *backpropagation* (**Chapitre 4**).

La démarche qu'on a proposée est fondée sur deux principales étapes; la première concerne la préparation de toutes les données utilisées (**Chapitre 5**), et la deuxième la création des RNs pour l'interpolation de la $f\text{CO}_2$ (**Chapitre 6**).

En premier lieu, les données *in situ* ont été structurées par saison, afin de faciliter les traitements qui en ont suivi. On a procédé ensuite à l'application de quelques filtres sur les données *in situ*, afin de ne garder que les points situés dans l'*open ocean*, et d'éliminer toutes les données aberrantes qui nuisent au bon fonctionnement des RNs. Ces filtres ont été appliqués sous forme de conditions sur les paramètres : bathymétrie, SSS, $f\text{CO}_2$, et distance à la masse terrestre (**Section 5.1.1**). Après cela, pour chaque saison, on a étudié la variabilité de tous les paramètres *in situ* (SST, SSS, et $f\text{CO}_2$). On a constaté que, le

fait que ces données ne couvrent pas tout le bassin Atlantique tropical, elles ne peuvent pas être utilisées en entier pour calculer la corrélation entre chaque paramètre avec la latitude et la longitude. De ce fait, on a essayé de calculer ces corrélations en se basant sur quelques trajets qui couvrent le bassin du Nord au Sud (de 23,5°S- 23,5 °N) et qui s'étendent sur au moins une vingtaine de degrés de longitudes. On a alors montré que ces paramètres sont très bien corrélés avec la latitude, cependant le seul paramètre qui est bien corrélé avec la longitude, est la $f\text{CO}_2$. L'autre remarque qu'on a faite c'est que, pendant les deux saisons printemps et hiver, la $f\text{CO}_2$ est plus élevée au Sud par rapport au Nord, et augmente en se rapprochant de la latitude 6°S (**Sections 5.1.1.1** et **5.1.1.2**). Néanmoins, il y a beaucoup plus de variabilités pendant l'été et l'automne (**Annexe B**).

En deuxième lieu, on est passé au traitement des données satellitaires. Trois paramètres ont été utilisés, Chl-*a* (SeaWifs et MODIS-A), SST (MODIS-A), et SSS (SMOS) (**Section 5.2**). Étant donnée l'importante taille des bases de données satellitaires (des millions de points), cette partie a été beaucoup plus consacrée à la démarche suivie ainsi qu'au programmes informatiques utilisés, permettant de gérer ces données satellitaires de façon optimisée, de la lecture des fichiers bruts (HDF pour Chl-*a* et SST, et NetCDF pour SSS) jusqu'à la sauvegarde sous forme de matrice '.mat'. On a, entre autre, montré comment extraire et structurer les données assimilées de SSS SODA (**Section 5.3**). Afin de faciliter la suite du traitement, on a proposé de combiner tous ces types de données en une seule matrice globale.

En troisième lieu, l'étape réalisée était la classification des données. Comme on l'a évoqué dans le **Chapitre 2**, puisque la zone d'étude est très grande, les variabilités des processus océaniques sont très importantes, ce qui peut, on prenant toutes les données *in situ* (à la fois) pour faire l'apprentissage des RNs, avoir des conséquences très négatives sur la performance de ceux-ci. De ce fait, il a fallu trouver un moyen pour diviser ces données en plusieurs sous-ensembles, afin de limiter les variabilités au sein d'un même ensemble, et donc de permettre le bon déroulement des apprentissages des RNs. Pour se faire, on a choisi de se baser sur la SST, qui est un paramètre très important en océanographie, de par son intervention dans plusieurs processus, notamment l'évolution climatique ainsi que la circulation océanique [Carton *et al.*, 1996]. Ainsi, on a utilisé la méthode de classification appelée *K-means* (**Section 4.3.1**) pour créer 3 classes (pour chaque saison) des données satellitaires de SST allant de l'hiver 2002/2003 jusqu'à l'automne 2014. Pour chaque saison, le résultat consiste en un ensemble de 3 centroïdes (moyennes) accompagnés de 3 intervalles de SST représentant les 3 classes (**Section 5.5.1**). Ceci a permis de réduire la variabilité de la SST dans toutes les saisons. Les intervalles retrouvés lors de cette étape, ont été utilisés pour diviser chaque saison des données *in situ* en 3 (C1, C2, et C3), en se basant sur la SST *in situ*. Ainsi, les BDs de création des RNs ont été prêtes à l'utilisation.

Une fois toutes les données structurées, on a abordé la phase fondamentale de cette étude, à savoir l'interpolation de la $f\text{CO}_2$ (**Chapitre 6**). Pour chaque classe produite lors de l'étape précédente, on a essayé de trouver le meilleur RN qui, à partir de l'étape d'apprentissage, puisse représenter les relations qui existent entre la $f\text{CO}_2$ et : la SST, la SSS, et la Chl-*a*. Compte tenu de la complexité d'un système représentant toutes ces relations, on ne pouvait se baser que sur les résultats de l'apprentissage, pour déterminer le meilleur RN. De ce fait, on a proposé dans notre algorithme (**Figure 6.1**) de confronter chaque

RN à deux autres étapes après l'apprentissage ; la première ("validation") consistait à le valider avec des données *in situ*, extraites de la BDs de création du RN, et n'appartenant pas à la BD d'apprentissage. Ceci a permis de vérifier que les RNs étaient capables d'estimer la $f\text{CO}_2$ à partir de données *in situ* différentes. La deuxième étape, qu'on a nommé "test", est plus importante que la première car, étant donné que le but général de cette thèse était de calculer la $f\text{CO}_2$ en fonction des données satellitaires, celle-ci permet de vérifier si le RN créé était capable de le faire avec une bonne performance. L'algorithme qu'on a proposé a deux particularités : 1) afin d'augmenter les chances de trouver un RN très performant, on a joué sur le nombre de neurones cachés, de façon à créer un ensemble de 20 RNs tout en augmentant ce nombre de 1 à chaque itération, et 2) afin de garantir le bon fonctionnement du RN lors de la généralisation, le choix du RN optimal (parmi les 20) a été basé sur le RMSE minimal de l'étape de test. Un autre point qu'on considère important dans notre algorithme, concerne l'étape de création des 3 BDs d'apprentissage, de validation, et de test. En effet, la BD *in situ* a été divisée en deux sous-ensembles afin d'assurer les deux premières étapes ($\sim 80\%$ et 20% respectivement). Pour le test, on a pris toutes les données satellitaires (et assimilées) qui correspondent à toutes les données *in situ*, ainsi, pendant cette étape, les RNs ont été confrontés à un nombre de données bien supérieur à celui de l'apprentissage.

Pour les deux saisons été et automne, l'étape de classification a produit une première classe (C1) contenant un nombre d'années très réduit (3 années) par rapport à l'échelle temporelle globale (14 années), ne pouvant être pris en considération pour représenter la $f\text{CO}_2$ pendant ces deux saisons. On a donc pris la décision de combiner cette classe avec la classe 2.

Tous les résultats obtenus ont été analysés afin d'évaluer les performances des RNs. On a donc tiré quelques conclusions relatives au fonctionnement de ces RNs, et plus particulièrement aux raisons de l'augmentation des résidus de ceux-ci, qui sont très liés aux facteurs suivants :

- La densité des valeurs de $f\text{CO}_2$: ceci a été confirmé pour toutes les saisons. En effet, on a remarqué que le rendement des RNs était très affecté par les valeurs de $f\text{CO}_2$ situées dans les extrémités des intervalles de celle-ci ; caractérisées par une augmentation des valeurs des résidus. Ainsi, pour les années ayant un faible pourcentage de données, et dont la majorité des valeurs de $f\text{CO}_2$ avaient une faible densité, les RMSEs étaient larges. Par exemple pour la classe 1 de la saison du printemps, les RMSEs des deux années 2001 et 2002 étaient très larges ($30,94$ et $40,13 \mu\text{atm}$ respectivement), ce qui a affecté le RMSE global de cette classe de $1 \mu\text{atm}$ ($10,5 \mu\text{atm}$: avec 2001 et 2002, et $9,5 \mu\text{atm}$: sans 2001 et 2002).
- Le changement de variabilités des paramètres : ceci survient principalement dans les zones d'upwelling, où on constate de très fortes augmentations de la $f\text{CO}_2$, mais aussi dans la zone située à l'embouchure du fleuve Amazone, où on constate de fortes chutes de SSS. Quand ceci survient, il y a un changement dans les relations entre les paramètres, et donc de fortes augmentations des résidus. Par exemple pour l'année 2012 de la classe 3 de l'hiver, entre les deux jours 30 et 31 janvier, il y a une forte augmentation de la $f\text{CO}_2$ et de la SST, et une forte variation de la SSS (\pm une unité par heure) ce qui a fait augmenter la valeur du RMSE pour cette année

à 19 μatm .

- Les incertitudes des données satellitaires (et assimilées) : cela concerne l'étape de test ; caractérisée par une augmentation des RMSEs par rapport à l'étape d'apprentissage. Ceci est directement lié aux incertitudes des données satellitaires et assimilées, pour lesquelles les RMSEs peuvent aller de 0,3 à 0,7 °C pour la SST, et de 0,4 à 0,7 pour la SSS (suivant les saisons). En essayant de recalculer les RMSEs de test de toutes les classes, en appliquant un seuil de 0,5 °C pour la SST, et de 0,4 pour la SSS, on a montré (**Tableau 7.2**) que ces incertitudes contribuent à l'augmentation des RMSEs d'une valeur qui peut aller jusqu'à $\sim 4 \mu\text{atm}$.

Saison	Classe	RMSE (μatm)		
		Apprentissage	Test	Test (seuils)
Hiver	C1	9	10,62	8,83
	C2	10,18	10,7	10,27
	C3	9,9	12,77	10,12
Printemps	C1	10,54	11,93	10,55
	C2	10,74	12,21	11,06
	C3	11,27	15,91	12,1
Été	C1C2	12,81	15,34	12,95
	C3	11,73	14,37	10,92
Automne	C1C2	12,68	14,57	13,3
	C3	14,89	17,08	15,71

TABLEAU 7.2 – Récapitulatif des résultats obtenus dans cette étude. La colonne 'Test (seuils)' contient les RMSEs recalculés en appliquant un seuil de 0,5 °C sur la SST satellitaire, et un seuil de 0,4 sur la SSS satellitaire.

Finalement, l'importance de passer par toutes ces étapes : traiter les saisons séparément, passer par une classification des données, et enfin, choisir le meilleur RN en fonction des résultats de la généralisation, se résume dans les résultats obtenus au cours de cette thèse, qui nous ont permis de confirmer qu'il était possible d'interpoler la $f\text{CO}_2$ exclusivement à partir des données satellitaires, mais aussi de présenter des cartes mensuelles de $f\text{CO}_2$ avec une bonne résolution spatiale (4 km \times 4 km) sur une large échelle temporelle (de 2002 à 2013) et avec de très bonnes précisions, qui varient de 8,8 à 15,7 μatm suivant les saisons (**Chapitre 7**).

Perspectives

L'ensemble de tous ces résultats permet d'introduire des perspectives, dans le but de poursuivre cette étude, mais aussi pour ouvrir d'autres pistes de recherche :

- **À court terme :**

On a vu à partir des cartes mensuelles de $f\text{CO}_2$ (2012/2013) présentées dans la **Figure 7.8**, qu'avoir des données de $f\text{CO}_2$ avec une telle résolution spatiale (4

km \times 4 km), apporte une aide précieuse quant-à l'étude de la variabilité annuelle de celle-ci, mais aussi à sa variabilité en fonction des autres paramètres. Ainsi, la première perspective qui fera l'objet de la suite directe de ce travail, serai l'utilisation des données qu'on a compilées (les 138 fichiers CSV), pour étudier de façon approfondie, les relations qui existent entre la variabilité saisonnière et annuelle de la fCO₂, et celle des autres paramètres (SST, SSS, et Chl-*a*). Ceci permettra de mieux comprendre comment chaque paramètre influence sur l'autre.

Afin de valider les cartes de fCO₂ produites dans cette étude, on doit les comparer avec les cartes climatologiques qui existent [Takahashi *et al.*, 2009, 2014], mais aussi avec d'autres données *in situ*.

Ceci doit être suivi par l'utilisation de la fCO₂ calculée et la fCO₂ atmosphérique pour la localisation des zones sources et éventuellement des zones puits de CO₂ dans cette zone. Ensuite, il faut réaliser la cartographie du flux de CO₂ air-mer, afin de donner la possibilité de mieux étudier le rôle qu'a eu l'océan Atlantique tropical, dans la régulation de l'augmentation de la concentration du CO₂ dans l'atmosphère, et aussi dans la pénétration du CO₂ anthropique dans l'océan, ceci pour la période qui va de 2002 à 2013.

— **À moyen terme :**

Dès lors que l'intérêt de passer par une étape de classification a été confirmé dans cette thèse, on propose de refaire cette étape en se basant à la fois sur : la SST et la Chl-*a*, ensuite sur la SST et la SSS, et enfin sur la SST, la Chl-*a*, et la SSS. Et ainsi comparer les résultats, et voir quelle combinaison permet d'améliorer les performances des RNs.

Il semble aussi intéressant de localiser les endroits où les résidus des RNs ont beaucoup augmenté, à cause du manque de données ou des fortes variabilités de la fCO₂ (ou des autres paramètres), et qui ont affecté les performances des RNs. Ceci dans le but d'avoir plus d'informations sur ces endroits, mais aussi d'orienter l'échantillonnage vers ces zone là, afin d'avoir plus de points, ce qui permettra d'améliorer les performances des modèles.

— **À long terme :**

On suggère aussi d'étendre la démarche qu'on a proposée sur les autres océans tropicaux (Pacifique et Indien), tout en apportant les modifications nécessaires pour adapter cette démarche en fonction des caractéristiques connues de chaque océan.

Bibliographie

- ABEL, B. (2012). Régions approximatives en fréquence et en longueur d'onde du spectre électromagnétique. URL : https://commons.wikimedia.org/wiki/File:Domaines_du_spectre_
- ALI, M. M., KISHTAWAL, C. M. et JAIN, S. (2007). Predicting cyclone tracks in the north Indian Ocean : An artificial neural network approach. *Geophysical Research Letters*, 34(4):L04603. DOI : 10.1029/2006GL028353.
- ANDERSON, J. A. et ROSENFELD, E., éditeurs (1988). *Neurocomputing : Foundations of Research*. MIT Press, Cambridge, MA, USA.
- ANDRIÉ, C., OUDOT, C., GENTHON, C. et MERLIVAT, L. (1986). CO₂ fluxes in the tropical Atlantic during FOCAL cruises. *Journal of Geophysical Research : Oceans*, 91(C10):11741–11755. DOI : 10.1029/JC091iC10p11741.
- ARNAULT, S. (1987). Tropical Atlantic geostrophic currents and ship drifts. *Journal of Geophysical Research : Oceans*, 92(C5):5076–5088. DOI : 10.1029/JC092iC05p05076.
- BAKKER, D. C. E., PFEIL, B., LANDA, C. S., METZL, N., O'BRIEN, K. M., OLSEN, A., SMITH, K., COSCA, C., HARASAWA, S., JONES, S. D., NAKAOKA, S.-I., NOJIRI, Y., SCHUSTER, U., STEINHOFF, T., SWEENEY, C., TAKAHASHI, T., TILBROOK, B., WADA, C., WANNINKHOF, R., ALIN, S. R., BALESTRINI, C. F., BARBERO, L., BATES, N. R., BIANCHI, A. A., BONOU, F., BOUTIN, J., BOZEC, Y., BURGER, E. F., CAI, W.-J., CASTLE, R. D., CHEN, L., CHERICI, M., CURRIE, K., EVANS, W., FEATHERSTONE, C., FEELY, R. A., FRANSSON, A., GOYET, C., GREENWOOD, N., GREGOR, L., HANKIN, S., HARDMAN-MOUNTFORD, N. J., HARLAY, J., HAUCK, J., HOPPEMA, M., HUMPHREYS, M. P., HUNT, C. W., HUSS, B., IBÁÑEZ, J. S. P., JOHANNESSEN, T., KEELING, R., KITIDIS, V., KÖRTZINGER, A., KOZYR, A., KRASAKOPOULOU, E., KUWATA, A., LANDSCHÜTZER, P., LAUVSET, S. K., LEFÈVRE, N., LO MONACO, C., MANKE, A., MATHIS, J. T., MERLIVAT, L., MILLERO, F. J., MONTEIRO, P. M. S., MUNRO, D. R., MURATA, A., NEWBERGER, T., OMAR, A. M., ONO, T., PATERSON, K., PEARCE, D., PIERROT, D., ROBBINS, L. L., SAITO, S., SALISBURY, J., SCHLITZER, R., SCHNEIDER, B., SCHWEITZER, R., SIEGER, R., SKJELVAN, I., SULLIVAN, K. F., SUTHERLAND, S. C., SUTTON, A. J., TADOKORO, K., TELSZEWSKI, M., (2016). A multi-decade record of high-quality fCO₂ data in version 3 of the surface ocean CO₂ atlas (socat). *Earth System Science Data Discussions*, 2016:1–55. DOI : 10.5194/essd-2016-15.

- BAKKER, D. C. E., PFEIL, B., SMITH, K., HANKIN, S., OLSEN, A., ALIN, S. R., COSCA, C., HARASAWA, S., KOZYR, A., NOJIRI, Y., O'BRIEN, K. M., SCHUSTER, U., TELSZEWski, M., TILBROOK, B., WADA, C., AKL, J., BARBERO, L., BATES, N. R., BOUTIN, J., BOZEC, Y., CAI, W.-J., CASTLE, R. D., CHAVEZ, F. P., CHEN, L., CHERICI, M., CURRIE, K., de BAAR, H. J. W., EVANS, W., FEELY, R. A., FRANSSON, A., GAO, Z., HALES, B., HARDMAN-MOUNTFORD, N. J., HOPPEMA, M., HUANG, W.-J., HUNT, C. W., HUSS, B., ICHIKAWA, T., JOHANNESSEN, T., JONES, E. M., JONES, S. D., JUTTERSTRÖM, S., KITIDIS, V., KÖRTZINGER, A., LANDSCHÜTZER, P., LAUVSET, S. K., LEFÈVRE, N., MANKE, A. B., MATHIS, J. T., MERLIVAT, L., METZL, N., MURATA, A., NEWBERGER, T., OMAR, A. M., ONO, T., PARK, G.-H., PATERSON, K., PIERROT, D., RÍOS, A. F., SABINE, C. L., SAITO, S., SALISBURY, J., SARMA, V. V. S. S., SCHLITZER, R., SIEGER, R., SKJELVAN, I., STEINHOFF, T., SULLIVAN, K. F., SUN, H., SUTTON, A. J., SUZUKI, T., SWEENEY, C., TAKAHASHI, T., TJIPUTRA, J., TSURUSHIMA, N., van HEUVEN, S. M. A. C., VANDEMARK, D., VLAHOS, P., WALLACE, D. W. R., WANNINKHOF, R. et WATSON, A. J. (2014). An update to the Surface Ocean CO₂ Atlas (SOCAT version 2). *Earth System Science Data*, 6(1):69–90. DOI : 10.5194/essd-6-69-2014.
- BALAKRISHNAN, P. V., COOPER, M. C., JACOB, V. S. et LEWIS, P. A. (1994). A study of the classification capabilities of neural networks using unsupervised learning : A comparison with K-means clustering. *Psychometrika*, 59:509–525. DOI : 10.1007/BF02294390.
- BAÇÃO, F., LOBO, V. et PAINHO, M. (2005). Self-organizing Maps as Substitutes for K-Means Clustering. In SUNDERAM, V. S., van ALBADA, G. D., SLOOT, P. M. A. et DONGARRA, J., éditeurs : *Computational Science – ICCS 2005*, numéro 3516 de Lecture Notes in Computer Science, pages 476–483. Springer Berlin Heidelberg. DOI : 10.1007/11428862_65.
- BARALE, V., GOWER, J. F. R. et ALBEROTANZA, L. (2010). *Oceanography from Space*. Springer Netherlands.
- BARRETO, G. A. (2007). Time Series Prediction with the Self-Organizing Map : A Review. In HAMMER, D. B. et HITZLER, D. P., éditeurs : *Perspectives of Neural-Symbolic Integration*, volume 77, pages 135–158. Springer Berlin Heidelberg. DOI : 10.1007/978-3-540-73954-8_6.
- BARRETO, G. A. et ARAÚJO, A. F. R. (2002). Nonlinear Modeling of Dynamic Systems with the Self-Organizing Map. In DORRONSORO, J. R., éditeur : *Artificial Neural Networks – ICANN 2002*, volume 2415, pages 975–980. Springer Berlin Heidelberg. DOI : 10.1007/3-540-46084-5_158.
- BATTLE, M., BENDER, M. L., TANS, P. P., WHITE, J. W. C., ELLIS, J. T., CONWAY, T. et FRANCEY, R. J. (2000). Global Carbon Sinks and Their Variability Inferred from Atmospheric O₂ and $\delta^{13}\text{C}$. *Science*, 287(5462):2467–2470. DOI : 10.1126/science.287.5462.2467.
- BIASUTTI, M., BATTISTI, D. S. et SARACHIK, E. S. (2003). The Annual Cycle over the Tropical Atlantic, South America, and Africa*. *Journal of Climate*, 16(15):2491–2508. DOI : 10.1175/1520-0442(2003)016<2491:TACOTT>2.0.CO;2.

- BOCK, H.-H. (2007). Clustering Methods : A History of k-Means Algorithms. In BRITO, P., CUCUMEL, G., BERTRAND, P. et DE CARVALHO, F., éditeurs : *Selected Contributions in Data Analysis and Classification*, Studies in Classification, Data Analysis, and Knowledge Organization, pages 161–172. Springer Berlin Heidelberg. DOI : 10.1007/978-3-540-73560-1_15.
- BOHLOULI, R., ROSTAMI, B. et KEIGHOBADI, J. (2012). Application of Neuro-Wavelet Algorithm in Ultrasonic-Phased Array Nondestructive Testing of Polyethylene Pipelines. *Journal of Control Science and Engineering*, 2012. DOI : 10.1155/2012/389690, 10.1155/2012/389690.
- BOURLES, B., MOLINARI, R. L., JOHNS, E., WILSON, W. D. et LEAMAN, K. D. (1999). Upper layer currents in the western tropical North Atlantic (1989–1991). *Journal of Geophysical Research : Oceans*, 104(C1):1361–1375. DOI : 10.1029/1998JC900025.
- BOUTET, E. (2007). Onde électromagnétique. URL : https://commons.wikimedia.org/wiki/File:Onde_electromagnetique.svg.
- BOUTIN, J., MARTIN, N., REVERDIN, G., YIN, X. et GAILLARD, F. (2013). Sea surface freshening inferred from SMOS and ARGO salinity : impact of rain. *Ocean Science*, 9(1):183–192. DOI : 10.5194/os-9-183-2013.
- CARTON, J. A., CAO, X., GIESE, B. S. et DA SILVA, A. M. (1996). Decadal and Interannual SST Variability in the Tropical Atlantic Ocean. *Journal of Physical Oceanography*, 26(7):1165–1175. DOI : 10.1175/1520-0485(1996)026<1165 :DAISVI>2.0.CO ;2.
- CARTON, J. A. et GIESE, B. S. (2008). A Reanalysis of Ocean Climate Using Simple Ocean Data Assimilation (SODA). *Monthly Weather Review*, 136(8):2999–3017. DOI : 10.1175/2007MWR1978.1.
- CARTON, J. A. et HACKERT, E. C. (1989). Application of multi-variate statistical objective analysis to the circulation in the tropical Atlantic ocean. *Dynamics of Atmospheres and Oceans*, 13:491–515. DOI : 10.1016/0377-0265(89)90051-1.
- CARTON, J. A. et HACKERT, E. C. (1990). Data Assimilation Applied to the Temperature and Circulation in the Tropical Atlantic, 1983–84. *Journal of Physical Oceanography*, 20(8):1150–1165. DOI : 10.1175/1520-0485(1990)020<1150 :DAATTT>2.0.CO ;2.
- CARTON, J. A. et ZHOU, Z. (1997). Annual cycle of sea surface temperature in the tropical Atlantic Ocean. *Journal of Geophysical Research : Oceans*, 102(C13):27813–27824. DOI : 10.1029/97JC02197.
- CAVAZOS, T. (2000). Using Self-Organizing Maps to Investigate Extreme Climate Events : An Application to Wintertime Precipitation in the Balkans. *Journal of Climate*, 13(10):1718–1732. DOI : 10.1175/1520-0442(2000)013<1718 :USOMTI>2.0.CO ;2.
- CENTRE CANADIEN DE CARTOGRAPHIE ET D'OBSERVATION DE LA TERRE (2015). Notion fondamentales de télédétection. <http://www.rncan.gc.ca/sciences-terre/geomatique/imagerie-satellitaire-photos-aeriennes/>

- imagerie-satellitaire-produits/ressources-educatives/9310. Consulté le 11 décembre 2015.
- COPIN-MONTEGUT, C. (1988). A new formula for the effect of temperature on the partial pressure of CO₂ in seawater. *Marine Chemistry*, 25:29–37. DOI : 10.1016/0304-4203(88)90012-6.
- COPIN-MONTEGUT, C. (1989). A new formula for the effect of temperature on the partial pressure of carbon dioxide in seawater, corrigendum. *Marine Chemistry*, 27:143–144.
- COPIN-MONTÉGUT, C. et AVRIL, B. (1995). Continuous pCO₂ measurements in surface water of the Northeastern tropical Atlantic. *Tellus B*, 47(1-2):86–92. DOI : 10.1034/j.1600-0889.47.issue1.9.x.
- DA-ALLADA, C. Y., ALORY, G., du PENHOAT, Y., KESTENARE, E., DURAND, F. et HOUNKONNOU, N. M. (2013). Seasonal mixed-layer salinity balance in the tropical Atlantic Ocean : Mean state and seasonal cycle. *Journal of Geophysical Research : Oceans*, 118(1):332–345. DOI : 10.1029/2012JC008357.
- da SILVEIRA, I. C. A., de MIRANDA, L. B. et BROWN, W. S. (1994). On the origins of the North Brazil Current. *Journal of Geophysical Research : Oceans*, 99(C11):22501–22512. DOI : 10.1029/94JC01776.
- DEMASTER, D. J., SMITH JR, W. O., NELSON, D. M. et ALLER, J. Y. (1996). Biogeochemical processes in Amazon shelf waters : chemical distributions and uptake rates of silicon, carbon and nitrogen. *Continental Shelf Research*, 16(5–6):617–643. DOI : 10.1016/0278-4343(95)00048-8.
- DEMPSTER, A. P., LAIRD, N. M. et RUBIN, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *JOURNAL OF THE ROYAL STATISTICAL SOCIETY, SERIES B*, 39(1):1–38.
- DICKSON, A. G. (1984). pH scales and proton-transfer reactions in saline media such as sea water. *Geochimica et Cosmochimica Acta*, 48(11):2299–2308. DOI : 10.1016/0016-7037(84)90225-4.
- DICKSON, A. G. (1993). pH buffers for sea water media based on the total hydrogen ion concentration scale. *Deep Sea Research Part I : Oceanographic Research Papers*, 40(1):107–118. DOI : 10.1016/0967-0637(93)90055-8.
- DICKSON, A. G. et MILLERO, F. J. (1987). A comparison of the equilibrium constants for the dissociation of carbonic acid in seawater media. *Deep Sea Research Part A. Oceanographic Research Papers*, 34(10):1733–1743. DOI : 10.1016/0198-0149(87)90021-5.
- DUPONT, S. et PÖRTNER, H.-O. (2013). A snapshot of ocean acidification research. *Marine Biology*, 160(8):1765–1771. DOI : 10.1007/s00227-013-2282-9.

- ESA (2013). SMOS earth explorers. http://www.esa.int/Our_Activities/Observing_the_Earth/The_Living_Planet_Programme/Earth_Explorers/SMOS/Overview3.
- FOLTZ, G. R. et MCPHADEN, M. J. (2008). Seasonal mixed layer salinity balance of the tropical North Atlantic Ocean. *Journal of Geophysical Research : Oceans*, 113(C2): C02013. DOI : 10.1029/2007JC004178.
- FORGY, E. W. (1965). Cluster analysis of multivariate data : efficiency versus interpretability of classifications. *Biometrics*, 21:768–769.
- FRATANTONI, D. M. et GLICKSON, D. A. (2002). North Brazil Current Ring Generation and Evolution Observed with SeaWiFS*. *Journal of Physical Oceanography*, 32(3): 1058–1074. DOI : 10.1175/1520-0485(2002)032<1058 :NBCRGA>2.0.CO ;2.
- FRIEDLINGSTEIN, P., ANDREW, R. M., ROGELJ, J., PETERS, G. P., CANADELL, J. G., KNUTTI, R., LUDERER, G., RAUPACH, M. R., SCHAEFFER, M., van VUUREN, D. P. et LE QUÉRÉ, C. (2014). Persistent growth of CO₂ emissions and implications for reaching climate targets. *Nature Geoscience*, 7:709–715. DOI : 10.1038/ngeo2248.
- FRIEDRICH, T. et OSCHLIES, A. (2009). Neural network-based estimates of North Atlantic surface pCO₂ from satellite data : A methodological study. *Journal Of Geophysical Research-Oceans*, 114:C03020. DOI : 10.1029/2007JC004646.
- GATTUSO, J.-P. et LAVIGNE, H. (2009). Technical Note : Approaches and software tools to investigate the impact of ocean acidification. *Biogeosciences*, 6(10):2121–2133. DOI : 10.5194/bg-6-2121-2009.
- GORDON, L. I. et JONES, L. B. (1973). The effect of temperature on carbon dioxide partial pressures in seawater. *Marine Chemistry*, 1:317–322. DOI : 10.1016/0304-4203(73)90021-2.
- GORGUES, T. et PLANQUETTE, H. (2016). COMPLEXITÉ ET EFFICACITÉ DE LA "POMPE BIOLOGIQUE" DE CARBONE. URL : <https://www.labexmer.eu/fr/recherche/pompe-biologique>, Institut universitaire européen de la mer (IUEM)/LabexMER, labexMER frontiers in marine research.
- GOYET, C., ADAMS, R. et EISCHEID, G. (1998). Observations of the CO₂ system properties in the tropical Atlantic Ocean. *Marine Chemistry*, 60:49–61. DOI : 10.1016/S0304-4203(97)00081-9.
- GOYET, C., MILLERO, F. J., POISSON, A. et SHAFER, D. K. (1993). Temperature dependence of CO₂ fugacity in seawater. *Marine Chemistry*, 44:205–219. DOI : 10.1016/0304-4203(93)90203-Z.
- GOYET, C. et POISSON, A. (1989). New determination of carbonic acid dissociation constants in seawater as a function of temperature and salinity. *Deep-Sea Research Part I-Oceanographic Research Papers*, 36:1635–1654. DOI : 10.1016/0198-0149(89)90064-2.

- GRUBER, N., GLOOR, M., MIKALOFF FLETCHER, S. E., DONEY, S. C., DUTKIEWICZ, S., FOLLOWS, M. J., GERBER, M., JACOBSON, A. R., JOOS, F., LINDSAY, K., MENEMENLIS, D., MOUCHET, A., MÜLLER, S. A., SARMIENTO, J. L. et TAKAHASHI, T. (2009). Oceanic sources, sinks, and transport of atmospheric CO₂. *Global Biogeochemical Cycles*, 23(1):GB1005. DOI : 10.1029/2008GB003349.
- HALLDIN, M. (2006). Intertropical Convergence Zone (ITCZ). URL : https://commons.wikimedia.org/wiki/File:ITCZ_january-july.png.
- HANSSON, I. (1973a). A new set of acidity constants for carbonic acid and boric acid in sea water. *Deep Sea Research and Oceanographic Abstracts*, 20(5):461–478. DOI : 10.1016/0011-7471(73)90100-9.
- HANSSON, I. (1973b). A new set of pH-scales and standard buffers for sea water. *Deep Sea Research and Oceanographic Abstracts*, 20(5):479–491. DOI : 10.1016/0011-7471(73)90101-0.
- HARRY, C. (2009). English : potato. URL : <https://commons.wikimedia.org/wiki/File:Carbon-cycle-full.jpg>.
- HARTIGAN, J. A. (1975). *Clustering Algorithms (Probability & Mathematical Statistics)*. John Wiley & Sons Inc. ISBN : 0-471-35645-X.
- HARTIGAN, J. A. et WONG, M. A. (1979). Algorithm AS 136 : A K-Means Clustering Algorithm. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 28(1): 100–108. DOI : 10.2307/2346830.
- HASTENRATH, S. et LAMB, P. (1978). On the dynamics and climatology of surface flow over the Equatorial oceans. *Tellus*, 30:436–448. DOI : 10.1111/j.2153-3490.1978.tb00859.x.
- HEBB, D. O. (1949). *The Organization of Behavior : A Neuropsychological Theory*. New York : Wiley. ISBN : 978-1-4106-1240-3.
- HO, D. T., LAW, C. S., SMITH, M. J., SCHLOSSER, P., HARVEY, M. et HILL, P. (2006). Measurements of air-sea gas exchange at high wind speeds in the Southern Ocean : Implications for global parameterizations. *Geophysical Research Letters*, 33(16). DOI : 10.1029/2006GL026817.
- HO, D. T., WANNINKHOF, R., SCHLOSSER, P., ULLMAN, D. S., HEBERT, D. et SULLIVAN, K. F. (2011). Toward a universal relationship between wind speed and gas exchange : Gas transfer velocities measured with ³He/SF₆ during the Southern Ocean Gas Exchange Experiment. *Journal of Geophysical Research : Oceans*, 116(C4):C00F04. DOI : 10.1029/2010JC006854.
- HOPFIELD, J. J. (1982). Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences*, 79(8):2554–2558.

- INOUE, H. Y., ISHII, M., MATSUEDA, H., KAWANO, T., MURATA, A. et TAKASUGI, Y. (2003). Distribution of the partial pressure of CO₂ in surface water (pCO₂^{sw}) between Japan and the Hawaiian Islands : pCO₂^{sw}—SST relationship in the winter and summer. *Tellus Series B-Chemical And Physical Meteorology*, 55:456–465. DOI : 10.1034/j.1600-0889.2003.01482.x.
- IOCCP (2007). *Surface Ocean CO₂ Variability and Vulnerabilities Workshop*. UNESCO, Paris, France.
- IPCC (2014). *Climate Change 2013 - The Physical Science Basis : Working Group I Contribution to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA. DOI : 10.1017/CBO9781107415324. ISBN : 978-1-107-41532-4.
- JACOBSON, M. Z. (2005). Studying ocean acidification with conservative, stable numerical schemes for nonequilibrium air-ocean exchange and ocean equilibrium chemistry. *Journal of Geophysical Research : Atmospheres*, 110(D7). DOI : 10.1029/2004JD005220.
- JANCEY, R. C. (1966). Multidimensional group analysis. *Australian Journal of Botany*, 14(1):127–130.
- JEAN-BAPTISTE, P., FOURRÉ, E. et POISSON, A. (2002). Gas Transfer Velocities for ³He in a Lake at High Wind Speeds. In DONELAN, M. A., MENNAN, W., SALTZMAN, E. S. et WANNINKHOF, R., éditeurs : *Gas Transfer at Water Surfaces*, pages 233–238. American Geophysical Union. ISBN : 978-1-118-66863-4.
- JOHNSON, S. C. (1967). Hierarchical clustering schemes. *Psychometrika*, 32(3):241–254.
- JOOS, F., PLATTNER, G.-K., STOCKER, T. F., MARCHAL, O. et SCHMITTNER, A. (1999). Global warming and marine carbon cycle feedbacks on future atmospheric CO₂. *Science (New York, N. Y.)*, 284(5413):464–467.
- KHOO, K. H., RAMETTE, R. W., CULBERSON, C. H. et BATES, R. G. (1977). Determination of hydrogen ion concentrations in seawater from 5 to 40°C : standard potentials at salinities from 20 to 45‰. *Analytical Chemistry*, 49(1):29–34. DOI : 10.1021/ac50009a016.
- KOHONEN, T. (1972). Correlation Matrix Memories. *IEEE Transactions on Computers*, C-21(4):353–359. DOI : 10.1109/TC.1972.5008975.
- KOHONEN, T. (1982). Self-organized formation of topologically correct feature maps. *Biological Cybernetics*, 43(1):59–69. DOI : 10.1007/BF00337288.
- KOŁODZIEJCZYK, N., BOUTIN, J., VERGELY, J.-L., MARCHAND, S., MARTIN, N. et REVERDIN, G. (2016). Mitigation of systematic errors in SMOS sea surface salinity. *Remote Sensing of Environment*, 180:164 – 177. DOI : <http://dx.doi.org/10.1016/j.rse.2016.02.061>, Special Issue : ESA's Soil Moisture and Ocean Salinity Mission - Achievements and Applications.

- KOMHYR, W. D., GAMMON, R. H., HARRIS, T. B., WATERMAN, L. S., CONWAY, T. J., TAYLOR, W. R. et THONING, K. W. (1985). Global atmospheric CO₂ distribution and variations from 1968–1982 NOAA/GMCC CO₂ flask sample data. *Journal of Geophysical Research : Atmospheres*, 90(D3):5567–5596. DOI : 10.1029/JD090iD03p05567.
- KÖRTZINGER, A. (2003). A significant CO₂ sink in the tropical Atlantic Ocean associated with the Amazon River plume. *Geophysical Research Letters*, 30(24):2287. DOI : 10.1029/2003GL018841.
- KUSHNIR, Y., ROBINSON, W. A., CHANG, P. et ROBERTSON, A. W. (2006). The Physical Basis for Predicting Atlantic Sector Seasonal-to-Interannual Climate Variability*. *Journal of Climate*, 19(23):5949–5970. DOI : 10.1175/JCLI3943.1.
- LANDSCHÜTZER, P., GRUBER, N., BAKKER, D. C. E., SCHUSTER, U., NAKAOKA, S., PAYNE, M. R., SASSE, T. P. et ZENG, J. (2013). A neural network-based estimate of the seasonal to inter-annual variability of the Atlantic Ocean carbon sink. *Biogeosciences*, 10:7793–7815. DOI : 10.5194/bg-10-7793-2013.
- LE QUÉRÉ, C., ANDRES, R. J., BODEN, T., CONWAY, T., HOUGHTON, R. A., HOUSE, J. I., MARLAND, G., PETERS, G. P., van der WERF, G., AHLSTRÖM, A., ANDREW, R. M., BOPP, L., CANADELL, J. G., CIAIS, P., DONEY, S. C., ENRIGHT, C., FRIEDLINGSTEIN, P., HUNTINGFORD, C., JAIN, A. K., JOURDAIN, C., KATO, E., KEELING, R. F., KLEIN GOLDEWIJK, K., LEVIS, S., LEVY, P., LOMAS, M., POULTER, B., RAUPACH, M. R., SCHWINGER, J., SITCH, S., STOCKER, B. D., VIOVY, N., ZAEHLE, S. et ZENG, N. (2012). The global carbon budget 1959–2011. *Earth System Science Data Discussions*, 5(2):1107–1157. DOI : 10.5194/essdd-5-1107-2012.
- LE QUÉRÉ, C., MORIARTY, R., ANDREW, R. M., PETERS, G. P., CIAIS, P., FRIEDLINGSTEIN, P., JONES, S. D., SITCH, S., TANS, P., ARNETH, A., BODEN, T. A., BOPP, L., BOZEC, Y., CANADELL, J. G., CHINI, L. P., CHEVALLIER, F., COSCA, C. E., HARRIS, I., HOPPEMA, M., HOUGHTON, R. A., HOUSE, J. I., JAIN, A. K., JOHANNESSEN, T., KATO, E., KEELING, R. F., KITIDIS, V., KLEIN GOLDEWIJK, K., KOVEN, C., LANDA, C. S., LANDSCHÜTZER, P., LENTON, A., LIMA, I. D., MARLAND, G., MATTHIS, J. T., METZL, N., NOJIRI, Y., OLSEN, A., ONO, T., PENG, S., PETERS, W., PFEIL, B., POULTER, B., RAUPACH, M. R., REGNIER, P., RÖDENBECK, C., SAITO, S., SALISBURY, J. E., SCHUSTER, U., SCHWINGER, J., SÉFÉRIAN, R., SEGSCHNEIDER, J., STEINHOFF, T., STOCKER, B. D., SUTTON, A. J., TAKAHASHI, T., TILBROOK, B., van der WERF, G. R., VIOVY, N., WANG, Y.-P., WANNINKHOF, R., WILTSHIRE, A. et ZENG, N. (2015). Global carbon budget 2014. *Earth System Science Data*, 7:47–85. DOI : 10.5194/essd-7-47-2015.
- LE QUÉRÉ, C., RAUPACH, M. R., CANADELL, J. G. et MARLAND, G. (2009). Trends in the sources and sinks of carbon dioxide. *Nature Geoscience*, 2(12):831–836. DOI : 10.1038/ngeo689.
- LEE, K., WANNINKHOF, R., TAKAHASHI, T., DONEY, S. C. et FEELY, R. A. (1998). Low interannual variability in recent oceanic uptake of atmospheric carbon dioxide. *Nature*, 396(6707):155–159. DOI : 10.1038/24139.

- LEFÈVRE, N., J. WATSON, A. J. et R. WATSON, A. R. (2005). A comparison of multiple regression and neural network techniques for mapping in situ pCO₂ data. *Tellus Series B-Chemical And Physical Meteorology*, 57:375–384. DOI : 10.1111/j.1600-0889.2005.00164.x.
- LEFÈVRE, N., MOORE, G., AIKEN, J., WATSON, A., COOPER, D. et LING, R. (1998). Variability of pCO₂ in the tropical Atlantic in 1995. *Journal Of Geophysical Research-Oceans*, 103:5623–5634. DOI : 10.1029/97JC02303.
- LEFÈVRE, N. et TAYLOR, A. (2002). Estimating pCO₂ from sea surface temperatures in the Atlantic gyres. *Deep-Sea Research Part I-Oceanographic Research Papers*, 49:539–554. DOI : 10.1016/S0967-0637(01)00064-4.
- LEFÈVRE, N., URBANO, D. F., GALLOIS, F. et DIVERRÈS, D. (2014). Impact of physical processes on the seasonal distribution of the fugacity of CO₂ in the western tropical Atlantic. *Journal Of Geophysical Research-Oceans*, 119:646–663. DOI : 10.1002/2013JC009248.
- LEFÈVRE, N., DIVERRÈS, D. et GALLOIS, F. (2010). Origin of CO₂ undersaturation in the western tropical Atlantic. *Tellus B*, 62(5):595–607. DOI : 10.1111/j.1600-0889.2010.00475.x.
- LISS, P. S. et MERLIVAT, L. (1986). Air-Sea Gas Exchange Rates : Introduction and Synthesis. In BUAT-MÉNARD, P., éditeur : *The Role of Air-Sea Exchange in Geochemical Cycling*, pages 113–127. Springer Netherlands. ISBN : 978-94-010-8606-6 978-94-009-4738-2.
- LIU, Y., WEISBERG, R. H. et HE, R. (2006). Sea Surface Temperature Patterns on the West Florida Shelf Using Growing Hierarchical Self-Organizing Maps. *Journal of Atmospheric and Oceanic Technology*, 23(2):325–338. DOI : 10.1175/JTECH1848.1.
- LLOYD, S. P. (1957). Binary Block Coding. *Bell System Technical Journal*, 36(2):517–535. DOI : 10.1002/j.1538-7305.1957.tb02410.x.
- LUMPKIN, R. et GARZOLI, S. L. (2005). Near-surface circulation in the Tropical Atlantic Ocean. *Deep Sea Research Part I : Oceanographic Research*, 52:495–518. DOI : 10.1016/j.dsr.2004.09.001.
- MACQUEEN, J. (1967). Some methods for classification and analysis of multivariate observations. In LE CAM, L. M. et J., N., éditeurs : *Fifth Berkeley Symposium on Mathematical Statistics and Probability*, volume 1, Berkeley, Calif. : University of California Press. The Regents of the University of California.
- MATEAR, R. J. et HIRST, A. C. (1999). Climate change feedback on the future oceanic CO₂ uptake. *Tellus B*, 51(3):722—733. DOI : 10.3402/tellusb.v51i3.16472.
- MCCULLOCH, W. S. et PITTS, W. (1943). A logical calculus of the ideas immanent in nervous activity. *Bulletin Of Mathematical Biology*, 5:115–133.

- MEHRBACH, C., CULBERSON, C. H., HAWLEY, J. E. et PYTKOWICX, R. M. (1973). Measurement of the Apparent Dissociation Constants of Carbonic Acid in Seawater at Atmospheric Pressure. *Limnology and Oceanography*, 18(6):897–907. DOI : 10.4319/lo.1973.18.6.0897.
- MERLE, J., FIEUX, M. et HISARD, P. (1979). Annual signal and interannual anomalies of sea surface temperature in the Eastern Equatorial Atlantic ocean. *Deep Sea Research Gate Supplement*, 5(26):77–102.
- MIKALOFF FLETCHER, S. E., GRUBER, N., JACOBSON, A. R., DONEY, S. C., DUTKIEWICZ, S., GERBER, M., FOLLOWS, M., JOOS, F., LINDSAY, K., MENEMENLIS, D., MOUCHET, A., MÜLLER, S. A. et SARMIENTO, J. L. (2006). Inverse estimates of anthropogenic CO₂ uptake, transport, and storage by the ocean. *Global Biogeochemical Cycles*, 20(2):GB2002. DOI : 10.1029/2005GB002530.
- MIKALOFF FLETCHER, S. E., GRUBER, N., JACOBSON, A. R., GLOOR, M., DONEY, S. C., DUTKIEWICZ, S., GERBER, M., FOLLOWS, M., JOOS, F., LINDSAY, K., MENEMENLIS, D., MOUCHET, A., MÜLLER, S. A. et SARMIENTO, J. L. (2007). Inverse estimates of the oceanic sources and sinks of natural CO₂ and the implied oceanic carbon transport. *Global Biogeochemical Cycles*, 21(1):GB1010. DOI : 10.1029/2006GB002751.
- MINSKY, M. et PAPERT, S. (1969). *Perceptrons : an introduction to computational geometry*. MIT Press, Cambridge (Mass.) ; London.
- MITCHELL, T. P. et WALLACE, J. M. (1992). The Annual Cycle in Equatorial Convection and Sea Surface Temperature. *Journal of Climate*, 5(10):1140–1156. DOI : 10.1175/1520-0442(1992)005<1140 :TACIEC>2.0.CO ;2.
- MOORE, T. S., CAMPBELL, J. W. et DOWELL, M. D. (2009). A class-based approach to characterizing and mapping the uncertainty of the MODIS ocean chlorophyll product. *Remote Sensing of Environment*, 113(11):2424–2430. DOI : 10.1016/j.rse.2009.07.016.
- MOUSSA, H., BENALLAL, M. A., GOYET, C., LEFÈVRE, N., EL JAI, M. C., GUGLIELMI, V. et TOURATIER, F. (2015). A comparison of Multiple Non-linear regression and neural network techniques for sea surface salinity estimation in the tropical Atlantic ocean based on satellite data. In EL YACOUBI, S., AFIFI, L., ZERRIK, E. H. et TRIDANE, A., éditeurs : *CMACS 2014 - Colloque de Modélisation, Analyse et Contrôle des Systèmes*, volume 49, pages 65–78, Al-Akhawayn University, Morocco. ESAIM : Proceedings And Surveys. DOI : 10.1051/proc/201549006.
- MOUSSA, H., BENALLAL, M. A., GOYET, C. et LEFÈVRE, N. (2016). Satellite-derived CO₂ fugacity in surface seawater of the tropical Atlantic Ocean using a feedforward neural network. *International Journal of Remote Sensing*, 37(3):580–598. DOI : 10.1080/01431161.2015.1131872.
- NAKAOKA, S., TELSZEWSKI, M., NOJIRI, Y., YASUNAKA, S., MIYAZAKI, C., MUKAI, H. et USUI, N. (2013). Estimating temporal and spatial variation of ocean surface pCO₂ in the North Pacific using a self-organizing map neural network technique. *Biogeosciences*, 10:6093–6106. DOI : 10.5194/bg-10-6093-2013.

- NASA OCEAN BIOLOGY (OB) (2014a). Aquarius/sac-d. <http://oceancolor.gsfc.nasa.gov/cms/data/aquarius>.
- NASA OCEAN BIOLOGY (OB) (2014b). Moderate resolution imaging spectroradiometer-aqua (MODIS-A) ocean color data. <http://oceancolor.gsfc.nasa.gov/cms/data/aqua>.
- NASA OCEAN BIOLOGY (OB) (2014c). Sea-viewing wide field-of-view sensor (seawifs) ocean color data. <http://oceancolor.gsfc.nasa.gov/cms/data/seawifs>.
- NIANG, A., BADRAN, F., MOULIN, C., CRÉPON, M. et THIRIA, S. (2006). Retrieval of aerosol type and optical thickness over the Mediterranean from SeaWiFS images using an automatic neural classification method. *Remote Sensing of Environment*, 100(1):82–94. DOI : 10.1016/j.rse.2005.10.005.
- NIGHTINGALE, P. D., MALIN, G., LAW, C. S., WATSON, A. J., LISS, P. S., LIDDICOAT, M. I., BOUTIN, J. et UPSTILL-GODDARD, R. C. (2000). In situ evaluation of air-sea gas exchange parameterizations using novel conservative and volatile tracers. *Global Biogeochemical Cycles*, 14(1):373–387. DOI : 10.1029/1999GB900091.
- NITTROUER, C. A. et DEMASTER, D. J. (1996). The Amazon shelf setting : tropical, energetic, and influenced by a large river. *Continental Shelf Research*, 16(5–6):553–573. DOI : 10.1016/0278-4343(95)00069-0.
- OLSEN, A., BELLERBY, R. G. J., JOHANNESSEN, T., OMAR, A. M. et SKJELVAN, I. (2003). Interannual variability in the wintertime air–sea flux of carbon dioxide in the northern North Atlantic, 1981–2001. *Deep-Sea Research Part I-Oceanographic Research Papers*, 50:1323–1338. DOI : 10.1016/S0967-0637(03)00144-4.
- ORR, J. C., MAIER-REIMER, E., MIKOLAJEWICZ, U., MONFRAY, P., SARMIENTO, J. L., TOGGWEILER, J. R., TAYLOR, N. K., PALMER, J., GRUBER, N., SABINE, C. L., LE QUÉRE, C., KEY, R. M. et BOUTIN, J. (2001). Estimates of anthropogenic carbon uptake from four three-dimensional global ocean models. *Global Biogeochemical Cycles*, 15(1):43–60. DOI : 10.1029/2000GB001273.
- OUDOT, C. et ANDRIÉ, C. (1986). Variabilité des pressions partielles de CO₂ océanique et atmosphérique dans l’atlantique tropical. *Oceanologica Acta*, 9(2):169–177.
- PARARD, G. (2011). *Etude de la variabilité de la fugacité du CO₂ dans l’Atlantique tropicale : de l’échelle diurne à saisonnière*. Doctorat en océanographie, Université Pierre et Marie Curie, laboratoire LOCEAN, Ecole Doctorale des Sciences de l’Environnement d’Ile de France, Paris, France.
- PARK, G.-H. et WANNINKHOF, R. (2012). A large increase of the CO₂ sink in the western tropical North Atlantic from 2002 to 2009. *Journal Of Geophysical Research-Oceans*, 117:C08029. DOI : 10.1029/2011JC007803.
- PEREZ, V., FERNÁNDEZ, E., MARAÑÓN, E., SERRET, P. et GARCÍA-SOTO, C. (2005). Seasonal and interannual variability of chlorophyll-a and primary production in the

- Equatorial Atlantic : in situ and remote sensing observations. *Journal of Plankton Research*, 27(2):189–197. DOI : 10.1093/plankt/fbh159.
- PETERSON, R. G. et STRAMMA, L. (1991). Upper-level circulation in the South Atlantic Ocean. *Progress in Oceanography*, 26(1):1–73. DOI : 10.1016/0079-6611(91)90006-8.
- PFEIL, B., OLSEN, A., BAKKER, D. C. E., HANKIN, S., KOYUK, H., KOZYR, A., MALCZYK, J., MANKE, A., METZL, N., SABINE, C. L., AKL, J., ALIN, S. R., BATES, N., BELLERBY, R. G. J., BORGES, A., BOUTIN, J., BROWN, P. J., CAI, W.-J., CHAVEZ, F. P., CHEN, A., COSCA, C., FASSBENDER, A. J., FEELY, R. A., GONZÁLEZ-DÁVILA, M., GOYET, C., HALES, B., HARDMAN-MOUNTFORD, N., HEINZE, C., HOOD, M., HOPPEMA, M., HUNT, C. W., HYDES, D., ISHII, M., JOHANNESSEN, T., JONES, S. D., KEY, R. M., KÖRTZINGER, A., LANDSCHÜTZER, P., LAUVSET, S. K., LEFÈVRE, N., LENTON, A., LOURANTOU, A., MERLIVAT, L., MIDORIKAWA, T., MINTROP, L., MIYAZAKI, C., MURATA, A., NAKADATE, A., NAKANO, Y., NAKAOKA, S., NOJIRI, Y., OMAR, A. M., PADIN, X. A., PARK, G.-H., PATERSON, K., PEREZ, F. F., PIERROT, D., POISSON, A., RÍOS, A. F., SANTANA-CASIANO, J. M., SALISBURY, J., SARMA, V. V. S. S., SCHLITZER, R., SCHNEIDER, B., SCHUSTER, U., SIEGER, R., SKJELVAN, I., STEINHOFF, T., SUZUKI, T., TAKAHASHI, T., TEDESCO, K., TELSZEWSKI, M., THOMAS, H., TILBROOK, B., TJIPUTRA, J., VANDEMARK, D., VENESS, T., WANNINKHOF, R., WATSON, A. J., WEISS, R., WONG, C. S. et YOSHIKAWA-INOUE, H. (2013). A uniform, quality controlled Surface Ocean CO₂ Atlas (SOCAT). *Earth System Science Data*, 5:125–143. DOI : 10.5194/essd-5-125-2013.
- PHILANDER, S. G. H. et PACANOWSKI, R. C. (1986). A model of the seasonal cycle in the tropical Atlantic Ocean. *Journal of Geophysical Research : Oceans*, 91(C12):14192–14206. DOI : 10.1029/JC091iC12p14192.
- PMEL CARBON GROUP (2016). Hawaii Carbon Dioxide Time-Series. URL : <http://www.pmel.noaa.gov/co2/file/Hawaii+Carbon+Dioxide+Time-Series>, NOAA PMEL Carbon Program.
- RAGHUNATH, R. et CARTON, J. A. (1990). Four-dimensional Data Assimilation In The Ocean. In *OCEANS '90. Engineering in the Ocean Environment. Conference Proceedings*, pages 221–226. DOI : 10.1109/OCEANS.1990.584763.
- RAMETTE, R. W., CULBERSON, C. H. et BATES, R. G. (1977). Acid-base properties of tris(hydroxymethyl)aminomethane (Tris) buffers in sea water from 5 to 40°C. *Analytical Chemistry*, 49(6):867–870. DOI : 10.1021/ac50014a049.
- RÖDENBECK, C., KEELING, R. F., BAKKER, D. C. E., METZL, N., OLSEN, A., SABINE, C. et HEIMANN, M. (2013). Global surface-ocean pCO₂ and sea–air CO₂ flux variability from an observation-driven ocean mixed-layer scheme. *Ocean Science*, 9(2):193–216. DOI : 10.5194/os-9-193-2013.
- REUSCH, D. B., ALLEY, R. B. et HEWITSON, B. C. (2007). North Atlantic climate variability from a self-organizing map perspective. *Journal of Geophysical Research : Atmospheres*, 112(D2):D02104. DOI : 10.1029/2006JD007460.

- REVERDIN, G., KESTENARE, E., FRANKIGNOUL, C. et DELCROIX, T. (2007). Surface salinity in the Atlantic Ocean (30°S–50°N). *Progress in Oceanography*, 73(3–4):311–340. DOI : 10.1016/j.pocean.2006.11.004.
- RICHARDSON, A. J., RISIEN, C. et SHILLINGTON, F. A. (2003). Using self-organizing maps to identify patterns in satellite imagery. *Progress in Oceanography*, 59(2–3):223–239. DOI : 10.1016/j.pocean.2003.07.006.
- RICHARDSON, P. L. et WALSH, D. (1986). Mapping climatological seasonal variations of surface currents in the tropical Atlantic using ship drifts. *Journal of Geophysical Research : Oceans*, 91(C9):10537–10550. DOI : 10.1029/JC091iC09p10537.
- ROSENBLATT, F. (1958). The perceptron : A probabilistic model for information storage and organization in the brain. *Psychological Review*, 65:386–408. DOI : 10.1037/h0042519.
- ROY, R. N., ROY, L. N., VOGEL, K. M., PORTER-MOORE, C., PEARSON, T., GOOD, C. E., MILLERO, F. J. et CAMPBELL, D. M. (1993). The dissociation constants of carbonic acid in seawater at salinities 5 to 45 and temperatures 0 to 45°C. *Marine Chemistry*, 44(2–4):249–267. DOI : 10.1016/0304-4203(93)90207-5.
- RUMELHART, D. E. et MCCLELLAND, J. L., éditeurs (1986). *Parallel Distributed Processing : Explorations in the Microstructure of Cognition, Vol. 1 : Foundations*. MIT PRESS, Cambridge, MA, USA. ISBN : 0-262-68053-X.
- RYTHER, J. H. et YENTSCH, C. S. (1957). The Estimation of Phytoplankton Production in the Ocean from Chlorophyll and Light Data. *Limnology and Oceanography*, 2(3):281–286. DOI : 10.1002/lno.1957.2.3.0281.
- SABINE, C., DUCKLOW, H. et HOOD, M. (2010). International Carbon Coordination : Roger Revelle's Legacy in the Intergovernmental Oceanographic Commission. *Oceanography*, 23(3):48–61. DOI : 10.5670/oceanog.2010.23.
- SABINE, C. L., FEELY, R. A., GRUBER, N., KEY, R. M., LEE, K., BULLISTER, J. L., WANNINKHOF, R., WONG, C. S., WALLACE, D. W. R., TILBROOK, B., MILLERO, F. J., PENG, T.-H., KOZYR, A., ONO, T. et RÍOS, A. F. (2004). The Oceanic Sink for Anthropogenic CO₂. *Science*, 305(5682):367–371. DOI : 10.1126/science.1097403.
- SABINE, C. L., HANKIN, S., KOYUK, H., BAKKER, D. C. E., PFEIL, B., OLSEN, A., METZL, N., KOZYR, A., FASSBENDER, A., MANKE, A., MALCZYK, J., AKL, J., ALIN, S. R., BELLERBY, R. G. J., BORGES, A., BOUTIN, J., BROWN, P. J., CAI, W.-J., CHAVEZ, F. P., CHEN, A., COSCA, C., FEELY, R. A., GONZÁLEZ-DÁVILA, M., GOYET, C., HARDMAN-MOUNTFORD, N., HEINZE, C., HOPPEMA, M., HUNT, C. W., HYDES, D., ISHII, M., JOHANNESSEN, T., KEY, R. M., KÖRTZINGER, A., LANDSCHÜTZER, P., LAUVSET, S. K., LEFÈVRE, N., LENTON, A., LOURANTOU, A., MERLIVAT, L., MIDORIKAWA, T., MINTROP, L., MIYAZAKI, C., MURATA, A., NAKADATE, A., NAKANO, Y., NAKAOKA, S., NOJIRI, Y., OMAR, A. M., PADIN, X. A., PARK, G.-H., PATERSON, K., PEREZ, F. F., PIERROT, D., POISSON, A., RÍOS, A. F., SALISBURY, J., SANTANA-CASIANO, J. M., SARMA, V. V. S. S., SCHLITZER, R., SCHNEIDER, B., SCHUSTER, U.,

- SIEGER, R., SKJELVAN, I., STEINHOFF, T., SUZUKI, T., TAKAHASHI, T., TEDESCO, K., TELSZEWSKI, M., THOMAS, H., TILBROOK, B., VANDEMARK, D., VENESS, T., WATSON, A. J., WEISS, R., WONG, C. S. et YOSHIKAWA-INOUE, H. (2013). Surface Ocean CO₂ Atlas (SOCAT) gridded data products. *Earth System Science Data*, 5(1):145–153.
- SAMARASINGHE, S. (2006). *Neural Networks for Applied Sciences and Engineering : From Fundamentals to Complex Pattern Recognition*. CRC Press. ISBN : 978-1-4200-1306-1.
- SARMIENTO, J. L., GLOOR, M., GRUBER, N., BEAULIEU, C., JACOBSON, A. R., MIKALOFF FLETCHER, S. E., PACALA, S. et RODGERS, K. (2010). Trends and regional distributions of land and ocean carbon sinks. *Biogeosciences*, 7(8):2351–2367. DOI : 10.5194/bg-7-2351-2010.
- SARMIENTO, J. L. et LE QUÉRÉ, C. (1996). Oceanic Carbon Dioxide Uptake in a Model of Century-Scale Global Warming. *Science (New York, N.Y.)*, 274(5291):1346–1350.
- SOCAT et SOCOM (2015). SOLAS Open Science Conference 2015. Conférence, Université de Kiel, Allemagne.
- STEINHAUS, H. (1956). Sur la division des corps matériels en parties. *Bulletin de l'Académie Polonaise des Sciences*, 4(12):801–804.
- STEPHENS, M. P., SAMUELS, G., OLSON, D. B., FINE, R. A. et TAKAHASHI, T. (1995). Sea-air flux of CO₂ in the North Pacific using shipboard and satellite data. *Journal Of Geophysical Research-Oceans*, 100:13571–13583. DOI : 10.1029/95JC00901.
- STRAMMA, L. et SCHOTT, F. (1999). The mean flow field of the tropical Atlantic Ocean. *Deep-Sea Research Part II-Topical Studies In Oceanography*, 46:279–303. DOI : 10.1016/S0967-0645(98)00109-X.
- SWEENEY, C., GLOOR, E., JACOBSON, A. R., KEY, R. M., MCKINLEY, G., SARMIENTO, J. L. et WANNINKHOF, R. (2007). Constraining global air-sea gas exchange for CO₂ with recent bomb 14C measurements. *Global Biogeochemical Cycles*, 21(2):GB2015. DOI : 10.1029/2006GB002784.
- TAKAHASHI, T. (1961). Carbon dioxide in the atmosphere and in Atlantic Ocean water. *Journal Of Geophysical Research-Oceans*, 66:477–494. DOI : 10.1029/JZ066i002p00477.
- TAKAHASHI, T., FEELY, R. A., WEISS, R. F., WANNINKHOF, R. H., CHIPMAN, D. W., SUTHERLAND, S. C. et TAKAHASHI, T. T. (1997). Global air-sea flux of CO₂ : An estimate based on measurements of sea-air pCO₂ difference. *Proceedings of the National Academy of Sciences*, 94(16):8292–8299.
- TAKAHASHI, T., SUTHERLAND, S. C., CHIPMAN, D. W., GODDARD, J. G., HO, C., NEWBERGER, T., SWEENEY, C. et MUNRO, D. R. (2014). Climatological distributions of pH, pCO₂, total CO₂, alkalinity, and CaCO₃ saturation in the global surface ocean, and temporal changes at selected locations. *Marine Chemistry*, 164:95–125. DOI : 10.1016/j.marchem.2014.06.004.

- TAKAHASHI, T., SUTHERLAND, S. C., SWEENEY, C., POISSON, A., METZL, N., TILBROOK, B., BATES, N., WANNINKHOF, R., FEELY, R. A., SABINE, C., OLAFSSON, J. et NOJIRI, Y. (2002). Global sea–air CO₂ flux based on climatological surface ocean pCO₂, and seasonal biological and temperature effects. *Deep-Sea Research Part II-Topical Studies In Oceanography*, 49:1601–1622. DOI : 10.1016/S0967-0645(02)00003-6.
- TAKAHASHI, T., SUTHERLAND, S. C., WANNINKHOF, R., SWEENEY, C., FEELY, R. A., CHIPMAN, D. W., HALES, B., FRIEDERICH, G., CHAVEZ, F., SABINE, C., WATSON, A., BAKKER, D. C. E., SCHUSTER, U., METZL, N., YOSHIKAWA-INOUE, H., ISHII, M., MIDORIKAWA, T., NOJIRI, Y., KÖRTZINGER, A., STEINHOFF, T., HOPPEMA, M., OLAFSSON, J., ARNARSON, T. S., TILBROOK, B., JOHANNESSEN, T., OLSEN, A., BELLERBY, R., WONG, C. S., DELILLE, B., BATES, N. R. et de BAAR, H. J. W. (2009). Climatological mean and decadal change in surface ocean pCO₂, and net sea–air CO₂ flux over the global oceans. *Deep-Sea Research Part II-Topical Studies In Oceanography*, 56:554–577. DOI : 10.1016/j.dsr2.2008.12.009.
- TAKAHASHI, T., WANNINKHOF, R. H., FEELY, R. A., WEISS, R. F., CHIPMAN, D. W., BATES, N., OLAFSSON, J., SABINE, C. et SUTHERLAND, S. C. (1999). Net sea–air CO₂ flux over the global oceans : An improved estimate based on the sea–air pCO₂ difference. pages 9–14, Center for Global Environmental Research, National Institute for Environmental Studies, Tsukuba, Japan. Proceedings of the Second International Symposium, CO₂ in the Oceans.
- TELSZEWSKI, M., CHAZOTTES, A., SCHUSTER, U., WATSON, A. J., MOULIN, C., BAKKER, D. C. E., GONZÁLEZ-DÁVILA, M., JOHANNESSEN, T., KÖRTZINGER, A., LÜGER, H., OLSEN, A., OMAR, A., PADIN, X. A., RÍOS, A. F., STEINHOFF, T., SANTANA-CASIANO, M., WALLACE, D. W. R. et WANNINKHOF, R. (2009). Estimating the monthly pCO₂ distribution in the North Atlantic using a self-organizing neural network. *Biogeosciences*, 6:1405–1421. DOI : 10.5194/bg-6-1405-2009.
- TERNON, J. F., OUDOT, C., DESSIER, A. et DIVERRES, D. (2000). A seasonal tropical sink for atmospheric CO₂ in the Atlantic ocean : the role of the Amazon River discharge. *Marine Chemistry*, 68(3):183–201. DOI : 10.1016/S0304-4203(99)00077-8.
- THE MATHWORKS, INC. (2016). MATLAB Product Description - MATLAB & Simulink. http://fr.mathworks.com/help/matlab/learn_matlab/product-description.html. Consulté le 13 mars 2016.
- TZORTZI, E., JOSEY, S. A., SROKOSZ, M. et GOMMENGINGER, C. (2013). Tropical Atlantic salinity variability : New insights from SMOS. *Geophysical Research Letters*, 40(10):2143–2147. DOI : 10.1002/grl.50225.
- UNIDATA (2016). netcdf package overview. <http://www.unidata.ucar.edu/software/netcdf/docs/faq.html#whatisit>.
- VOITURIEZ, B. et HERBLAND, A. (1982). Comparaisons des systèmes productifs de l’atlantique tropical est : dômes thermiques, upwellings côtiers et upwelling équatorial.

Rapports et procès-verbaux des réunions / Conseil permanent international pour l'exploration de la mer, 180:107–123.

WANG, C., DESER, C., YU, J.-Y., DI NEZIO, P. et CLEMENT, A. (2012). El Niño and southern oscillation (ENSO) : a review. *Coral Reefs of the Eastern Pacific*, pages 3–19.

WANNINKHOF, R. (1992). Relationship between wind speed and gas exchange. *Journal of Geophysical Research*, 97(25):7373–7382.

WANNINKHOF, R. (2014). Relationship between wind speed and gas exchange over the ocean revisited : Gas exchange and wind speed over the ocean. *Limnology and Oceanography : Methods*, 12(6):351–362. DOI : 10.4319/lom.2014.12.351.

WANNINKHOF, R., ASHER, W. E., HO, D. T., SWEENEY, C. et MCGILLIS, W. R. (2009). Advances in Quantifying Air-Sea Gas Exchange and Environmental Forcing. *Annual Review of Marine Science*, 1(1):213–244. DOI : 10.1146/annurev.marine.010908.163742.

WANNINKHOF, R., DONEY, S. C., BULLISTER, J. L., LEVINE, N. M., WARNER, M. et GRUBER, N. (2010). Detecting anthropogenic CO₂ changes in the interior Atlantic Ocean between 1989 and 2005. *Journal of Geophysical Research : Oceans*, 115(C11): C11028. DOI : 10.1029/2010JC006251.

WANNINKHOF, R. et MCGILLIS, W. R. (1999). A cubic relationship between air-sea CO₂ exchange and wind speed. *Geophysical Research Letters*, 26(13):1889–1892. DOI : 10.1029/1999GL900363.

WEISS, R. F. (1974). Carbon dioxide in water and seawater : the solubility of a non-ideal gas. *Marine Chemistry*, 2(3):203–215. DOI : 10.1016/0304-4203(74)90015-2.

WIDROW, B. (1960). An Adaptive 'Adaline' Neuron Using Chemical 'Memistors'. Technical Report 1553–2, Stanford Electronics Laboratories.

WIDROW, B. et HOFF, M. E. (1960). Adaptive switching circuits. *IRE WESCON Convention Record*, 4:96–104.

WORLDATELAS (2016). Equator Map, Tropic of Cancer Map, Tropic of Capricorn Map, Prime Meridian. URL : <http://www.worldatlas.com/aatlas/imagee.htm>.

XIE, S.-P. et CARTON, J. A. (2004). Tropical Atlantic Variability : Patterns, Mechanisms, and Impacts. In WANG, C., XIE, S. P. et CARTON, J. A., éditeurs : *Earth's Climate*, pages 121–142. American Geophysical Union. ISBN : 978-1-118-66594-7.

ZENG, J., NOJIRI, Y., LANDSCHÜTZER, P., TELSZEWSKI, M. et NAKAOKA, S. (2014). A Global Surface Ocean fCO₂ Climatology Based on a Feed-Forward Neural Network. *Journal Of Atmospheric And Oceanic Technology*, 31:1838–1849. DOI : 10.1175/JTECH-D-13-00137.1.

Annexe A

CODES MATLAB

A.1 Spécifications techniques

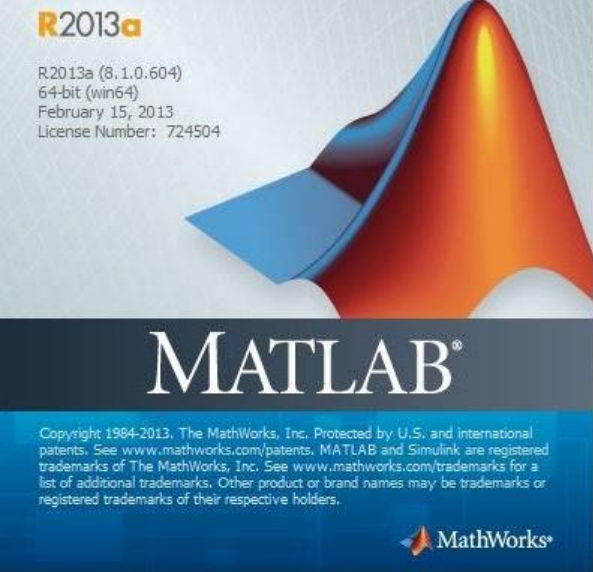
La version MATLAB® sur laquelle ont été développés les codes sources, ainsi que les spécifications techniques du PC sur lequel ces codes ont été exécutés, sont présentées dans la **Figure A.1**.

A.2 Lecture de fichier de type NetCDF

```
1 name_file = 'nomdufichier.nc'; % Le nom du fichier à traiter
2 % Ouvrir le fichier seulement en lecture
3 ncid = netcdf.open(name_file, 'NC_NOWRITE');
4 % Afin d'avoir des informations sur les variables que contient ce ...
   fichier, notamment les noms de ces variables, il faut exécuter la ...
   commande suivante en faisant varier le 'i' jusqu'au nombre total ...
   des variables:
5 % 'varname': nom de la variable
6 % 'xtype': le type de données de cette variable
7 [varname, xtype] = netcdf.inqVar(ncid,i);
8 % Récupérer les variables souhaitées , Exemples:
9 lon = double(netcdf.getVar(ncid,5));
10 lat = double(netcdf.getVar(ncid,6));
11 sst = double(netcdf.getVar(ncid,9));
12 fco2 = double(netcdf.getVar(ncid,10));
13 % Après l'extraction des variables, on peut les regrouper dans un ...
   seul tableau
```

A.3 Formatage des données *in situ*

```
1 % Charger la BD qui contient les données in situ
2 load('BD_globale_ETAPE_1.mat')
3 tab = BD_globale; clear BD_globale
```




R2013a
R2013a (8.1.0.604)
64-bit (win64)
February 15, 2013
License Number: 724504

MATLAB®

Copyright 1984-2013. The MathWorks, Inc. Protected by U.S. and International patents. See www.mathworks.com/patents. MATLAB and Simulink are registered trademarks of The MathWorks, Inc. See www.mathworks.com/trademarks for a list of additional trademarks. Other product or brand names may be trademarks or registered trademarks of their respective holders.

MathWorks®

CPU		Caches	Mainboard	Memory	SPD	Graphics	About		
Processor									
Name	Intel Xeon E5 1650 v2								
Code Name	Ivy Bridge-EP/EX	Max TDP	130 W						
Package	Socket 2011 LGA								
Technology	22 nm	Core VID	0.831 V						
Specification: Intel(R) Xeon(R) CPU E5-1650 v2 @ 3.50GHz									
Family	6	Model	E	Stepping	4				
Ext. Family	6	Ext. Model	3E	Revision	S0/S1				
Instructions	MMX, SSE, SSE2, SSE3, SSSE3, SSE4.1, SSE4.2, EM64T, VT-x, AES, AVX								
Clocks (Core #0)									
Core Speed	1196.71 MHz					Cache			
Multiplier	x 12.0 (12 - 39)					L1 Data	6 x 32 KBytes	8-way	
Bus Speed	99.73 MHz					L1 Inst.	6 x 32 KBytes	8-way	
QPI Link	3191.24 MHz					Level 2	6 x 256 KBytes	8-way	
						Level 3	12 MBytes	16-way	
Selection		Processor #1		Cores		6		Threads	12

CPU		Caches	Mainboard	Memory	SPD	Graphics	About
Motherboard							
Manufacturer	Dell Inc.						
Model	09M8Y8	A01					
Chipset	Intel	Ivy Bridge-E	Rev.	04			
Southbridge	Intel	X79	Rev.	06			
LPCIO	SMSC						
BIOS							
Brand	Dell Inc.						
Version	A06						
Date	02/28/2014						
Graphic Interface							
Version	PCI-Express						
Link Width	x16	Max. Supported	x16				
Side Band							

CPU		Caches	Mainboard	Memory	SPD	Graphics	About
L1 D-Cache							
Size	32 KBytes			x 6			
Descriptor	8-way set associative, 64-byte line size						
L1 I-Cache							
Size	32 KBytes			x 6			
Descriptor	8-way set associative, 64-byte line size						
L2 Cache							
Size	256 KBytes			x 6			
Descriptor	8-way set associative, 64-byte line size						
L3 Cache							
Size	12 MBytes						
Descriptor	16-way set associative, 64-byte line size						
Size							
Descriptor							

CPU		Caches	Mainboard	Memory	SPD	Graphics	About
General							
Type	DDR3			Channel #	Quad		
Size	32714 MBytes			DC Mode			
				NIB Frequency			
Timings							
DRAM Frequency	930.8 MHz						
FSB:DRAM	1:14						
CAS# Latency (CL)	13.0 clocks						
RAS# to CAS# Delay (tRCD)	13 clocks						
RAS# Precharge (tRP)	13 clocks						
Cycle Time (tRAS)	32 clocks						
Row Refresh Cycle Time (tRFC)	243 clocks						
Command Rate (CR)	1T						
DRAM Idle Timer							
Total CAS# (tRDRAM)							
Row To Column (tRCD)							


CPU		Caches	Mainboard	Memory	SPD	Graphics	About
Display Device Selection							
		NVIDIA Quadro K4000		Perf Level	Current		
GPU							
Name	NVIDIA Quadro K4000						
Board Manuf.	NVIDIA Corporation						
Code Name				Revision	A1		
Technology							
Clocks							
Core	324 MHz			Memory			
Shaders				Size	3072 MBytes		
Memory	324 MHz			Type			
				Bus Width			

FIGURE A.1 – Version MATLAB utilisée, et spécifications techniques du PC.

```

4 % Déclarer le nouveau tableau
5 table = cell(0);
6 table{1,1} = 'year';
7 table{1,2} = 'month';
8 table{1,3} = 'data';
9 %*****
10 year = [2000:1:2014]'; mon = [1:1:12]';
11 T1 = ones(size(mon,1),size(year,1));
12 for i = 1:size(year,1)
13     T1(:,i) = T1(:,i)*year(i);
14 end
15 T1 = T1(:); % Transformer la matrice T1 en un vecteur pour le ...
    concaténer avec T2
16 T2 = [];
17 for i = 1:size(year,1)
18     T2 = [T2; mon];
19 end
20 yearmon = [T1 T2];
21 %*****
22 % Remplir le tableau
23 k = 2;
24     for i = 1:size(yearmon,1)
25         id = find(tab(:,1)==yearmon(i,1) & tab(:,2)==yearmon(i,2));
26         if isempty(id)==0
27             table{k,1} = num2str(yearmon(i,1));
28             if length(num2str(yearmon(i,2))) == 1
29                 table{k,2} = ['0',num2str(yearmon(i,2))];
30             else
31                 table{k,2} = num2str(yearmon(i,2));
32             end
33             table{k,3} = tab(id,:);
34             k = k+1;
35         end
36     end
37 % Enregistrer le résultat
38 save('BD_globale_ETAPE_2.mat', 'table')

```

A.3.1 Diverses structurations des données *in situ*

Dans tous les codes qui suivent, on utilisera la tableau 'table' créé suivant le code MATLAB précédent.

A.3.1.1 Par saison :

```

1 sais = [12 1 2; 3 4 5; 6 7 8; 9 10 11]; % Les trois mois de chaque ...
    saison
2 hiver = []; printemps = []; ete = []; automne = [];
3 for i=1:4
4     % Pour chaque saison, chercher les mois dans le tableau 'table'
5     id = find( str2num(cell2mat(table(2:end,2))) == sais(i,1)...
6             | str2num(cell2mat(table(2:end,2))) == sais(i,2)...

```

```

7         | str2num(cell2mat(table(2:end,2))) == sais(i,3));
8     switch i
9         case 1
10            for j=1:size(id)
11                hiver = [hiver; table{id(j)+1,3}];
12            end
13        case 2
14            for j=1:size(id)
15                printemps = [printemps; table{id(j)+1,3}];
16            end
17        case 3
18            for j=1:size(id)
19                ete = [ete; table{id(j)+1,3}];
20            end
21        case 4
22            for j=1:size(id)
23                automne = [automne; table{id(j)+1,3}];
24            end
25    end
26 end

```

A.3.1.2 Par mois :

```

1 Jan = []; Fev = []; Mar = []; Avr = []; Mai = []; Jun = []; Jui = ...
  []; Aou = []; Sep = []; Oct = []; Nov = []; Dec = [];
2 tab = ['Jan'; 'Fev'; 'Mar'; 'Avr'; 'Mai'; 'Jun'; 'Jui'; 'Aou'; ...
  'Sep'; 'Oct'; 'Nov'; 'Dec'];
3 for i=1:12
4     id = find( str2num(cell2mat(table(2:end,3))) == i);
5     for j=1:size(id)
6         eval([ tab(i,:), ' = [' , tab(i,:), '; table{id(j)+1,4}]; ' ])
7     end
8 end

```

A.3.1.3 Par année :

```

1 for i=2000:2014
2     id = find( str2num(cell2mat(table(2:end,2))) == i);
3     eval([ 'a', num2str(i), ' = []; ' ])
4     for j=1:size(id)
5         eval([ 'a', num2str(i), ' = [a', num2str(i), '; ...
6             table{id(j)+1,4}]; ' ])
7     end
8 end

```

A.3.1.4 Par saison de chaque année :

```

1 sais = [12 1 2; 3 4 5; 6 7 8; 9 10 11];
2 for j = 1:4
3     for i=2000:2014
4         id = find( str2num(cell2mat(table(2:end,2))) == i & ( ...
                    str2num(cell2mat(table(2:end,3))) == sais(j,1) | ...
                    str2num(cell2mat(table(2:end,3))) == sais(j,2) | ...
                    str2num(cell2mat(table(2:end,3))) == sais(j,3)) );
5         if isempty(id) == 0
6             eval([ 'a',num2str(i),'sais',num2str(j),' = []; ' ])
7             for k=1:size(id)
8                 eval([ 'a',num2str(i),'sais',num2str(j),' = ...
                        [a',num2str(i),'sais',num2str(j),' ; ...
                        table{id(k)+1,4}]; ' ])
9             end
10            eval([ 'a',num2str(i),'sais',num2str(j),' = ...
                  a',num2str(i),'sais',num2str(j),' ( ...
                  a',num2str(i),'sais',num2str(j),' (:,10)>=32 ,:); '])
11        end
12    end
13 end

```

A.4 Prétraitement des données *in situ*

A.4.1 Bathymétrie et SSS

```

1 % Application sur les tableaux des saisons
2 a = find(hiver(:,6) >= 500 & hiver(:,10) >= 32); % 6: Bathy., 10: SSS
3 hiver = hiver(a,:);
4
5 a = find(printemps(:,6) >= 500 & printemps(:,10) >= 32);
6 printemps = printemps(a,:);
7
8 a = find(ete(:,6) >= 500 & ete(:,10) >= 32);
9 ete = ete(a,:);
10
11 a = find(automne(:,6) >= 500 & automne(:,10) >= 32);
12 automne = automne(a,:);

```

A.4.2 fCO₂ et distance à la terre

```

1 % Application sur le tableau de la saison d'été
2 a = find(ete(:,9) >= 140 | ete(:,12) <= 450); % 9: Dist., 12: fCO2
3 ete = ete(a,:);
4
5 a = find(ete(:,12) >= 320);
6 ete = ete(a,:);

```

A.5 Formatage des données de SST et Chl-*a* de MODIS-A

A.5.1 Lecture et structuration de fichiers HDF

Ce code doit être exécuté après avoir téléchargé les fichiers des données satellitaires. Prenons l'exemple des fichiers de SST, les noms des fichiers sont de la forme :

"A20070012007031.L3m_MO_SST_4"

1
2
3
4
5
6

1. jour du début de l'observation

2007 : année

001 : numéro du jour dans l'année

2. jour de fin de l'observation

2007 : année

031 : numéro du jour dans l'année (la période entre 2007001 et 2007031 représente le mois de janvier 2007)

3. format L3 SMI (**Section 3.4**)

4. nom du capteur (MO pour MODIS)

5. le paramètre observé (ici c'est la SST)

6. la résolution spatiale (4 pour 4 km × 4 km)

Après le téléchargement, il faut structurer les fichiers dans des dossiers par année (voir **Figure A.2**)

Nom	Modifié le	Type	Taille
2002	04/02/2015 09:35	Dossier de fichiers	
2003	04/02/2015 09:35	Dossier de fichiers	
2004	04/02/2015 09:36	Dossier de fichiers	
2005	04/02/2015 09:36	Dossier de fichiers	
2006	04/02/2015 09:36	Dossier de fichiers	
2007	04/02/2015 09:37	Dossier de fichiers	
2008	04/02/2015 09:37	Dossier de fichiers	
2009	04/02/2015 09:37	Dossier de fichiers	
2010	04/02/2015 09:38	Dossier de fichiers	
2011	04/02/2015 09:38	Dossier de fichiers	
2012	04/02/2015 09:38	Dossier de fichiers	
2013	04/02/2015 09:39	Dossier de fichiers	
2014	04/02/2015 09:39	Dossier de fichiers	

FIGURE A.2 – Exemple de structuration par année de fichiers satellitaires.


```

2 Slope = 7.17185E-4; Intercept = -2.0; % Pour la CHL; Slope = 1 et ...
   Intercept = 0
3 a1 = 1596; a2 = 2725; % Indices des latitudes -23.5 et 23.5
4 b1 = 2640; b2 = 4681; % Indices des longitudes -70 et 15
5 % L'adresse pour accéder aux fichiers sats
6 adr = 'D:/Documents/Data_Sat/SST/';
7 for i = 1:size(T,1) % Le tableau T contient: les années, jours de ...
   début et de fin d'observation, et le mois correspondant. Pour ...
   l'exemple du nom de fichier cité ci-dessus, on aura: | 2007 | 1 | ...
   31 | 1
8   annee = num2str(T(i, 1));
9   jour1 = num2str(T(i, 2));
10  jour2 = num2str(T(i, 3));
11  % Ajouter '0' aux numéros de jour composés de 2 chiffres, et ...
   '00' aux numéros composés d'un seul chiffre
12  if (length(jour1) == 2) jour1 = ['0', jour1];
13  elseif (length(jour1) == 1) jour1 = ['00', jour1];
14  end
15  if (length(jour2) == 2) jour2 = ['0', jour2];
16  elseif (length(jour2) == 1) jour2 = ['00', jour2];
17  end
18  % Lire le fichier HDF
19  s=[adr, annee, '/A', annee, jour1, annee, jour2, '.L3m_MO_SST_4'];
20  eval(['SST', annee, '_', num2str(T(i,4)), ' = ...
   hdfread(s, 'l3m_data');' ])
21  % Conversion des valeurs en 'double'
22  eval(['SST', annee, '_', num2str(T(i,4)), ' = double(SST', ...
   annee, '_', num2str(T(i,4)), ');' ])
23  % Extraction de la zone d'étude
24  eval(['SST', annee, '_', num2str(T(i,4)), ' = SST', annee, '_', ...
   num2str(T(i,4)), '(a1:a2,b1:b2);' ])
25  % Pour la SST, la valeur '65535' indique que le point ne ...
   contient pas de donnée (pour la chl-a c'est '-32767')
26  eval(['none = SST', annee, '_', num2str(T(i,4)), ' ~= 65535;' ])
27  eval(['SST', annee, '_', num2str(T(i,4)), ' = SST', annee, '_', ...
   num2str(T(i,4)), '.*none;' ])
28  eval(['temp = SST', annee, '_', num2str(T(i,4)), '';' ])
29  temp=temp(:);
30  % Combiner les valeurs de SST avec le tableau 'lat_lon' (Section ...
   A.4.1.1)
31  eval(['tab', annee, '_', num2str(T(i,4)), ' = [lat_lon temp];' ])
32  eval(['m = find(tab', annee, '_', num2str(T(i,4)), '(:,3) ~= 0);' ])
33  eval(['tab', annee, '_', num2str(T(i,4)), ' = tab', annee, '_', ...
   num2str(T(i,4)), '(m,:);' ])
34  % Applique la 'Scaling equation'
35  eval(['tab', annee, '_', num2str(T(i,4)), '(:,3) = (Slope*tab', ...
   annee, '_', num2str(T(i,4)), '(:,3)) + Intercept;' ])
36  % Sauvegarder le tableau créé
37  s = [adr, annee, '/SST', annee, '_', num2str(T(i,4)), '.mat'];
38  save([s , ['tab', annee, '_', num2str(T(i,4))]]);
39 end

```

À la fin de ce code, on aura pour chaque fichier traité, un fichier matlab (.mat) contenant les attributs : latitude | longitude | SST.

A.5.1.1 Création du tableau 'lat_lon'

```

1  %**** Tableau Latitude_Longitude (lat_lon) pour 4 km x 4 km ****
2  lat=[89.979164:-0.041666668:-90]; % Le pas est de 0.041666668
3  lon=[-179.97917:0.041666668:180];
4  % Extraire la partie de l'Atlantique tropical
5  a1 = 1596; a2 = 2725; % Indices des latitudes -23.5 et 23.5
6  b1 = 2640; b2 = 4681; % Indices des longitudes -70 et 15
7  lat=lat(a1:a2); lon=lon(b1:b2);
8  T1 = ones(size(lon,2),size(lat,2));
9  for i = 1:size(lat,2)
10     T1(:,i) = T1(:,i)*lat(i);
11 end
12 T1 = T1(:); % Transformer la matrice T1 en un vecteur pour le ...
    concaténer avec T2
13 T2 = [];
14 for i = 1:size(lat,2)
15     T2 = [T2; lon'];
16 end
17 lat_lon = [T1 T2];

```

A.6 Regroupement de SST et de Chl-a satellitaires

```

1  % Récupérer la liste des années et des mois depuis le tableau 'T' ...
    (comme dans l'étape précédente)
2  M = T(:, [1,4]);
3
4  for i = 1:size(M,1)
5     annee = num2str(M(i, 1));
6     mois = num2str(M(i, 2));
7
8     % Charger le fichier SST
9     load(['D:/Documents/Data_Sat/SST/', annee, '/SST', annee, '_', ...
        mois, '.mat'])
10    eval(['sst = tab', annee, '_', mois, ';' ])
11    eval(['clear tab', annee, '_', mois, ';' ])
12
13    % Charger le fichier CHL
14    load(['D:/Documents/Data_Sat/CHL/', annee, '/CHL', annee, '_', ...
        mois, '.mat'])
15    eval(['chl = tab', annee, '_', mois, ';' ])
16    eval(['clear tab', annee, '_', mois, ';' ])
17
18    % Rechercher les points appartenant à la fois au tableau de SST ...
        et au tableau de CHL
19    [a,b] = ismember(chl(:, [1,2]), sst(:, [1,2]), 'rows');
20
21    % Récupération des indices de ces points
22    c = [1:size(b)]';
23    b = [b c];

```

```

24     b = b(b(:,1)~=0, :);
25
26     % Joindre les deux tableaux
27     sst = sst(b(:,1), :);
28     chl = chl(b(:,2), :);
29     sst_chl = [sst chl(:,end)];
30
31     % Créer le dossier s'il n'existe pas
32     if ~exist(['D:/Documents/Data_Sat/SST_CHL/',annee], 'dir')
33         mkdir(['D:/Documents/Data_Sat/SST_CHL/',annee]);
34     end
35
36     % Enregistrer le nouveau tableau
37     eval(['tab', annee, '_',mois, '= sst_chl;'] )
38     save(['D:/Documents/Data_Sat/SST_CHL/', annee, '\SST_CHL', ...
39         annee, '_', mois, '.mat'], ['tab', annee, '_',mois]);
40     eval(['clear tab', annee, '_', mois, ';'] )
41 end

```

A.7 Formatage des données de SSS de SMOS

```

1 year = [2010:1:2013]'; % Déclaration des années pour lesquelles SSS ...
   de SMOS est disponible
2 mon = [1:1:12]'; % Déclaration des mois
3 % Création du tableau 'yearmon' qui va contenir toutes les ...
   combinaisons (année,mois)
4 T1 = ones(size(mon,1),size(year,1));
5 for i = 1:size(year,1)
6     T1(:,i) = T1(:,i)*year(i);
7 end
8 T1 = T1(:); % Transformer la matrice T1 en un vecteur pour le ...
   concaténer avec T2
9 T2 = [];
10 for i = 1:size(year,1)
11     T2 = [T2; mon];
12 end
13 yearmon = [T1 T2];
14 % Déclaration du chemin qui va contenir les fichiers structurés
15 chem = 'D:\Documents\Data_Sat\SMOS\SSS\';
16 % Lecture du fichier NetCDF de SSS SMOS
17 ncid = netcdf.open( ...
18     'D:\Documents\Data_Sat\SMOS\_REPBIN001M025B_ALL.nc', 'NC_NOWRITE');
19 % Récupération des données
20 [varname, xtype, varDimIDs, varAtts] = netcdf.inqVar(ncid,2);
21 lat = flipud(double(netcdf.getVar(ncid,2)));
22 [varname, xtype, varDimIDs, varAtts] = netcdf.inqVar(ncid,3);
23 lon = double(netcdf.getVar(ncid,3));
24 [varname, xtype, varDimIDs, varAtts] = netcdf.inqVar(ncid,4);
25 sss = double(netcdf.getVar(ncid,4));
26 a1 = 266; a2 = 455; % Indices de latitude
27 b1 = 440; b2 = 781; % Indices de longitude

```

```

28 % Création du tableau 'latlon' contenant toutes les latitudes et les ...
    longitudes
29 lat=lat(a1:a2); lon=lon(b1:b2);
30 T1 = ones(size(lat,1),size(lon,1));
31 for i = 1:size(lon,1)
32     T1(:,i) = T1(:,i)*lon(i);
33 end
34 T1 = T1(:); % Transformer la matrice T1 en un vecteur pour le ...
    concaténer avec T2
35 T2 = [];
36 for i = 1:size(lon,1)
37     T2 = [T2; lat];
38 end
39 latlon = [T1 T2];
40 % Structuration et sauvegarde des 48 fichiers (48 est le nombre ...
    total des mois; 4 années x 12 mois)
41 for i=1:48
42     eval(['sss_',num2str(i),' = flipud(sss(:, :, i))']; ])
43     eval(['none = sss_',num2str(i),'(:, :) < 0; ' ])
44     eval(['sss_',num2str(i),' (none) = NaN; ' ])
45     eval(['sss_',num2str(i),' = sss_',num2str(i),' (a1:a2,b1:b2); ' ])
46     eval(['tab_',num2str(yearmon(i,1)),'_',num2str(yearmon(i,2)),' ...
        = [latlon sss_',num2str(i),'(:)];'])
47
48     s = ...
        [chem,num2str(yearmon(i,1)),'\SMOS_SSS_',num2str(yearmon(i,1)), ...
        ...
        '_,num2str(yearmon(i,2))];
49     save(s, ['tab_',num2str(yearmon(i,1)),'_',num2str(yearmon(i,2))])
51 end

```

A.8 Formatage des données de SSS SODA

```

1 year = [2002:1:2009]'; % Déclaration des années pour lesquelles SSS ...
    de SMOS est disponible
2 mon = [1:1:12]'; % Déclaration des mois
3 % Déclaration du chemin contenant qui va contenir les fichiers ...
    structurés
4 chem = 'D:\Documents\Data_Climato\SODA\SSS\';
5 % Création du tableau 'latlon' contenant toutes les latitudes et les ...
    longitudes
6 lat_ = [89.25:-0.5:-75.25];
7 lon_ = [-180:0.5:179.5];
8 a1 = 132; a2 = 227; % Indices de latitude
9 b1 = 221; b2 = 391; % Indices de longitude
10 lat_=lat_(a1:a2); lon_=lon_(b1:b2);
11 T1 = ones(size(lat_,2),size(lon_,2));
12 for i = 1:size(lon_,2)
13     T1(:,i) = T1(:,i)*lon_(i);
14 end

```

```

15 T1 = T1(:); % Transformer la matrice T1 en un vecteur pour le ...
    concaténer avec T2
16 T2 = [];
17 for i = 1:size(lon_,2)
18     T2 = [T2; lat_'];
19 end
20 lat_lon = [T1 T2];
21 % de 2002 à 2008
22 i_sst = 0; i_sss = 1; i_depth = 10;
23 % pour 2009: i_sst = 6; i_sss = 7; i_depth = 4;
24 for i=1:size(year,1)
25     for j=1:size(mon,1)
26         mois=num2str(mon(j));
27         if (length(mois)==1)
28             mois=['0',mois];
29         end
30         disp([num2str(year(i)), '\SODA_',mois, '.cdf']);
31         % Lecture du fichier netcdf
32         ncid = ...
            netcdf.open([chem,num2str(year(i)), '\SODA_',mois, '.cdf'], ...
33             'NC_NOWRITE');
34         % Récupération des paramètres
35         sst = double(netcdf.getVar(ncid,i_sst));
36         sss = double(netcdf.getVar(ncid,i_sss));
37         depth = double(netcdf.getVar(ncid,i_depth));
38         sst=sst(:, :, 1);
39         sst=(flipud(sst));
40         sss=sss(:, :, 1);
41         sss=(flipud(sss));
42         % Inversion de la matrice
43         sss = [sss(:, 361:720) sss(:, 1:360)];
44         none=sss(:, :)<0;
45         sss(none)=NaN;
46         sst = [sst(:, 361:720) sst(:, 1:360)];
47         none=sst(:, :)<0;
48         sst(none)=NaN;
49         sss = sss(a1:a2,b1:b2);
50         sss = [lat_lon sss(:)];
51         sst = sst(a1:a2,b1:b2);
52         sst = sst(:);
53         depth = ones(size(lat_lon,1),1)*depth(1);
54         eval(['tab_', num2str(year(i)), '_',mois, ' = [sss, sst, ...
            depth];'])
55         s = ...
            [chem,num2str(year(i)), '\SODA_SSS_SST_', num2str(year(i)), ...
56             '_',mois];
57         % Sauvegarder le fichier
58         save(s, ['tab_', num2str(year(i)), '_',mois])
59     end
60 end

```

A.9 Regroupement des données *in situ* et SST et Chl-a de MODIS-A

```

1  %***** Pour 2002 à partir de juillet *****
2  for k = 2002
3      pt = [];
4      % Classer le tableau de données in situ par ordre croissant des mois
5      eval([ 'a', num2str(k), ' = sortrows(a', num2str(k), ',2);' ])
6      eval([ 'mois = unique(a', num2str(k), '(:,2));' ])
7      % Commencer par le mois de juillet car les données sat de MODISA ne
8      % sont disponibles qu'à partir de ce mois
9      for j = find(mois==7):size(mois,1)
10         eval([ ...
11             'load(''D:\Documents\Data_Sat\SST_CHL\MONTHLY\'', num2str(k), ...
12             '\SST_CHL', num2str(k), '_' , num2str(mois(j)), ''')' ])
13         % Récupérer la partie de données in situ correspondante au mois
14         % courant
15         eval([ 'temp = a', num2str(k), '(a', num2str(k), '(:,2) == ...
16             mois(j), :);' ])
17         for i = 1:size(temp,1)
18             % Affecter le tableau de données sat à la variable 'tab'
19             eval([ 'tab = tab', num2str(k), '_' , num2str(mois(j)), ';' ])
20             % Calcul de la distance entre chaque point du tableau de
21             % données in situ avec tous les points du tableau de données
22             % sat correspondant
23             dist = Distance_eq( temp(i,8), temp(i,7), tab(:,1), ...
24                 tab(:,2) );
25             [minval,ind] = min(dist(:));
26             % Récupération du point le plus proche (s'il existe) sinon
27             % mettre un zéro
28             if minval<0.0416666669
29                 pt = [pt; tab(ind,:)];
30             else
31                 pt = [pt; 0, 0, 0, 0];
32             end
33         end
34     end
35 end
36 % Ajouter les données sat aux données in situ
37 eval([ 'a', num2str(k), ' = [a', num2str(k), ', pt];' ])
38 %***** De 2003 à 2014 *****
39 for k = 2003:2014
40     pt = [];
41     % Classer le tableau de données in situ par ordre croissant des mois
42     eval([ 'a', num2str(k), ' = sortrows(a', num2str(k), ',2);' ])
43     eval([ 'mois = unique(a', num2str(k), '(:,2));' ])
44     for j = 1:size(mois,1)
45         eval([ ...
46             'load(''D:\Documents\Data_Sat\SST_CHL\MONTHLY\'', num2str(k), ...
47             '\SST_CHL', num2str(k), '_' , num2str(mois(j)), ''')' ])

```

```

45     % Récupérer la partie de données in situ correspondant au ...
        mois courant
46     eval([ 'temp = a', num2str(k), ' (a', num2str(k), '(:,2) == ...
        mois(j), :);' ])
47     for i = 1:size(temp,1)
48         % Affecter le tableau de données sat à la variable 'tab'
49         eval([ 'tab = tab', num2str(k), '_', num2str(mois(j)), ';' ])
50         % Calcul de la distance entre chaque point du tableau de ...
            données in situ avec tous les points du tableau de ...
            données sat correspondant en utilisant la fonction ...
            'Distance_eq' (définie dans la section A.8.1)
51         dist = Distance_eq( temp(i,8), temp(i,7), tab(:,1), ...
            tab(:,2) );
52         [minval,ind] = min(dist(:));
53         % Récupération du point le plus proche (s'il existe) ...
            sinon mettre des zéros
54         if minval<0.0416666669
55             pt = [pt; tab(ind,:)];
56         else
57             pt = [pt; 0, 0, 0, 0];
58         end
59     end
60 end
61 % Ajouter les données sat de CHL aux données in situ
62 eval([ 'a', num2str(k), ' = [a', num2str(k), ', pt];' ])
63 end

```

A.9.1 Fonction de calcul de la distance euclidienne

```

1 function D = Distance_eq( x1, y1, x2, y2 )
2
3     D = sqrt( (x2-x1).^2 + (y2-y1).^2 );

```

A.10 Regroupement des données *in situ* et SSS de SODA

```

1 % Déclarer le chemin des données de SSS SODA
2 chem = 'D:\Documents\Data_Climato\SODA\SSS\';
3 for j=2:size(table,1)
4     pt = [];
5     % Récupérer année mois et data
6     ann = table{j,2};
7     mois = table{j,3};
8     data = table{j,4};
9     if (length(mois)==1)
10         mois = ['0',mois];
11 end

```

```

12 % vérifier si le dossier SODA correspondant à cette date existe
13 if exist([chem,num2str(ann)], 'dir') ~= 0
14     % vérifier si le fichier existe
15     if exist([chem,num2str(ann),'\SODA_SSS_SST_',num2str(ann), ...
16         '_',num2str(mois),'.mat'], 'file') ~= 0
17         load([chem,num2str(ann),'\SODA_SSS_SST_',num2str(ann),' ...
18             '_',num2str(mois),'.mat']);
19         eval([ 'tab = ...
20                 tab_',num2str(ann),'_',num2str(mois),';' ] )
21         eval([ 'clear tab_',num2str(ann),'_',num2str(mois) ])
22     % Rechercher le points de climato correspondant
23     for k=1:size(data,1)
24         dist = Distance_eq( data(k,7), data(k,8), tab(:,1), ...
25             tab(:,2) );
26         [minval,ind] = min(dist(:));
27         if minval <= 0.5
28             pt = [pt; tab(ind,:)];
29         else
30             pt = [pt; 0 0 0 0 0];
31         end
32     end
33     data = [data pt];
34     table{j,4} = data;
35 end
36 else
37     data = [data zeros(size(data,1),5)];
38     table{j,4} = data;
39 end
end

```

A.11 Regroupement des données *in situ* et SSS de SMOS

```

1 % Déclarer le chemin des données de SSS SMOS
2 chem = 'D:\Documents\Data_Sat\SMOS\SSS\';
3 for j=2:size(table,1) % La première ligne contient l'entête
4     pt = [];
5     % Récupérer année mois et data
6     ann = str2num(table{j,2});
7     mois = str2num(table{j,3});
8     data = table{j,4};
9     % vérifier si le dossier sat et le fichier correspondant à ...
10    cette date existe
11    if (exist([chem,num2str(ann)], 'dir') ~= 0) & ...
12        (exist([chem,num2str(ann),'\SMOS_SSS_',num2str(ann), ...
13            '_',num2str(mois),'.mat'], 'file') ~= 0)
14        load([chem,num2str(ann),'\SMOS_SSS_',num2str(ann), ...
15            '_',num2str(mois),'.mat']);
16        eval([ 'tab = ...
17                tab_',num2str(ann),'_',num2str(mois),';' ] )

```



```

15         eval([ 'clear tab_', num2str(ann), '_', num2str(mois) ])
16     % Rechercher le points sat correspondant
17     for k=1:size(data,1)
18         dist = Distance_eq( data(k,7), data(k,8), tab(:,1), ...
19                             tab(:,2) );
20         [minval,ind] = min(dist(:));
21         if minval <= 0.25 & (isnan(tab(ind,3))==0)
22             pt = [pt; tab(ind,:)];
23         else
24             pt = [pt; 0 0 0];
25         end
26     end
27     data = [data pt];
28     table{j,4} = data;
29 else
30     data = [data zeros(size(data,1),3)];
31     table{j,4} = data;
32 end
end

```

A.12 Classification des données satellitaires de SST

```

1 % Pour chaque saison, construire un tableau global qui contient les ...
2   données de la saison dans toutes les années (ici de 2002 -> 2014)
3 tab_printemps = [];
4 tab_ete       = [];
5 tab_automne   = [];
6 tab_hiver     = [];
7
8 a = [2002; 2003; 2004; 2005; 2006; 2007; 2008; 2009; 2010; 2011; ...
9       2012; 2013; 2014];
10 adr = 'D:/Documents/Data_Sat/SST/';
11 % Récupérer le mois de Décembre 2002 pour le combiner avec la saison ...
12   d'hiver de 2003
13 load 'D:/Documents/Data_Sat/SST/2002/SST2002_12.mat'
14 for i=2:size(a,1)
15     % Aller dans le dossier de l'année courante
16     cd([adr,num2str(a(i))])
17     % Récupérer et charger les fichiers .mat (créés suivant le code ...
18     dans la section A.4.1)
19     file = dir('*.mat');
20     for j = 1:size(file,1)
21         load(file(j,1).name)
22     end
23     eval([ 'tab_printemps = [tab_printemps; tab', num2str(a(i)), '_3 ...
24             ones(size(tab', num2str(a(i)), '_3,1),1)*', num2str(a(i)), ' ...
25             ones(size(tab', num2str(a(i)), '_3,1),1)*3; ...
26             tab', num2str(a(i)), '_4 ...
27             ones(size(tab', num2str(a(i)), '_4,1),1)*', num2str(a(i)), ' ...
28             ones(size(tab', num2str(a(i)), '_4,1),1)*4; ...
29             tab', num2str(a(i)), '_5 ...

```

```

    ones(size(tab', num2str(a(i)), '_5', 1), 1) *', num2str(a(i)), ' ...
    ones(size(tab', num2str(a(i)), '_5', 1), 1) *5];' ] )
20 eval([ 'clear tab', num2str(a(i)), '_3 tab', num2str(a(i)), '_4 ...
    tab', num2str(a(i)), '_5' ] )
21 eval([ 'tab_ete = [tab_ete; tab', num2str(a(i)), '_6 ...
    ones(size(tab', num2str(a(i)), '_6', 1), 1) *', num2str(a(i)), ' ...
    'ones(size(tab', num2str(a(i)), '_6', 1), 1) *6; ...
    tab', num2str(a(i)), '_7 ...
    ones(size(tab', num2str(a(i)), '_7', 1), 1) *', num2str(a(i)), ' ...
    ones(size(tab', num2str(a(i)), '_7', 1), 1) *7; ...
    tab', num2str(a(i)), '_8 ...
    ones(size(tab', num2str(a(i)), '_8', 1), 1) *', num2str(a(i)), ' ...
    ones(size(tab', num2str(a(i)), '_8', 1), 1) *8];' ] )
22 eval([ 'clear tab', num2str(a(i)), '_6 tab', num2str(a(i)), '_7 ...
    tab', num2str(a(i)), '_8' ] )
23 eval([ 'tab_automne = [tab_automne; tab', num2str(a(i)), '_9 ...
    ones(size(tab', num2str(a(i)), '_9', 1), 1) *', num2str(a(i)), ' ...
    ones(size(tab', num2str(a(i)), '_9', 1), 1) *9; ...
    tab', num2str(a(i)), '_10 ...
    ones(size(tab', num2str(a(i)), '_10', 1), 1) *', num2str(a(i)), ' ...
    ones(size(tab', num2str(a(i)), '_10', 1), 1) *10; ...
    tab', num2str(a(i)), '_11 ...
    ones(size(tab', num2str(a(i)), '_11', 1), 1) *', num2str(a(i)), ' ...
    ones(size(tab', num2str(a(i)), '_11', 1), 1) *11];' ] )
24 eval([ 'clear tab', num2str(a(i)), '_9 tab', num2str(a(i)), '_10 ...
    tab', num2str(a(i)), '_11' ] )
25 eval([ 'tab_hiver = [tab_hiver; tab', num2str(a(i-1)), '_12 ...
    ones(size(tab', num2str(a(i-1)), '_12', 1), 1) *', num2str(a(i-1)), ' ...
    'ones(size(tab', num2str(a(i-1)), '_12', 1), 1) *12; ...
    tab', num2str(a(i)), '_1 ...
    ones(size(tab', num2str(a(i)), '_1', 1), 1) *', num2str(a(i)), ' ...
    ones(size(tab', num2str(a(i)), '_1', 1), 1) *1; ...
    tab', num2str(a(i)), '_2 ...
    ones(size(tab', num2str(a(i)), '_2', 1), 1) *', num2str(a(i)), ' ...
    ones(size(tab', num2str(a(i)), '_2', 1), 1) *2];' ] )
26 eval([ 'clear tab', num2str(a(i-1)), '_12 tab', num2str(a(i)), '_1 ...
    tab', num2str(a(i)), '_2' ] )
27 end
28 % Classification en 3 classes, en utilisant la fonction prédéfinie ...
    'kmeans'
29 % Appliquer un seuil de 10 sur la SST
30 tab_printemps = tab_printemps(tab_printemps(:, 3) >= 10, :);
31 tab_ete = tab_ete(tab_ete(:, 3) >= 10, :);
32 tab_automne = tab_automne(tab_automne(:, 3) >= 10, :);
33 tab_hiver = tab_hiver(tab_hiver(:, 3) >= 10, :);
34 % Les variable 'B_...' contiennent les centroides des classes
35 [A_printemps, B_printemps] = kmeans(tab_printemps(:, 3), 3);
36 [A_ete, B_ete] = kmeans(tab_ete(:, 3), 3);
37 [A_automne, B_automne] = kmeans(tab_automne(:, 3), 3);
38 [A_hiver, B_hiver] = kmeans(tab_hiver(:, 3), 3);
39 % Classer les centroides par ordres croissant
40 B_hiver = sortrows(B_hiver, 1);
41 B_printemps = sortrows(B_printemps, 1);
42 B_ete = sortrows(B_ete, 1);
43 B_automne = sortrows(B_automne, 1);

```

```

44 % Sauvegarder les centroides
45 save('D:\Documents\Kmeans\centroides.mat', 'B_printemps', 'B_ete', ...
      'B_automne', 'B_hiver')

```

A.13 Classification des données *in situ*

```

1 % Charger les centroides des classes préalablement sauvegardés ...
  (Section A.11)
2 load('D:\Documents\Kmeans\centroides.mat')
3 % Structurer les données in situ (de la BD de construction des RNS) ...
  par saison (suivant le code MATLAB dans Section A.2.1.1)
4 % Diviser les tableaux en trois parties suivant les trois classes
5 % Hiver
6 h11 = B_hiver(1,2); h12 = B_hiver(1,3); h1=find(hiver(:,11)>=h11 & ...
  hiver(:,11)<=h12);
7 h21 = B_hiver(2,2); h22 = B_hiver(2,3); h2=find(hiver(:,11)>=h21 & ...
  hiver(:,11)<=h22);
8 h31 = B_hiver(3,2); h32 = B_hiver(3,3); h3=find(hiver(:,11)>=h31 & ...
  hiver(:,11)<=h32);
9 hiver_C1 = hiver(h1,:); hiver_C2 = hiver(h2,:); hiver_C3 = hiver(h3,:);
10 % Printemps
11 p11 = B_printemps(1,2); p12 = B_printemps(1,3); ...
  p1=find(printemps(:,11)>=p11 & printemps(:,11)<=p12);
12 p21 = B_printemps(2,2); p22 = B_printemps(2,3); ...
  p2=find(printemps(:,11)>=p21 & printemps(:,11)<=p22);
13 p31 = B_printemps(3,2); p32 = B_printemps(3,3); ...
  p3=find(printemps(:,11)>=p31 & printemps(:,11)<=p32);
14 printemps_C1 = printemps(p1,:); printemps_C2 = printemps(p2,:); ...
  printemps_C3 = printemps(p3,:);
15 %Été
16 e11 = B_ete(1,2); e12 = B_ete(1,3); e1=find(ete(:,11)>=e11 & ...
  ete(:,11)<=e12);
17 e21 = B_ete(2,2); e22 = B_ete(2,3); e2=find(ete(:,11)>=e21 & ...
  ete(:,11)<=e22);
18 e31 = B_ete(3,2); e32 = B_ete(3,3); e3=find(ete(:,11)>=e31 & ...
  ete(:,11)<=e32);
19 ete_C1 = ete(e1,:); ete_C2 = ete(e2,:); ete_C3 = ete(e3,:);
20 % Regrouper les deux premières classes
21 ete_C1C2 = [ete_C1; ete_C2];
22 % Automne
23 a11 = B_automne(1,2); a12 = B_automne(1,3); ...
  a1=find(automne(:,11)>=a11 & automne(:,11)<=a12);
24 a21 = B_automne(2,2); a22 = B_automne(2,3); ...
  a2=find(automne(:,11)>=a21 & automne(:,11)<=a22);
25 a31 = B_automne(3,2); a32 = B_automne(3,3); ...
  a3=find(automne(:,11)>=a31 & automne(:,11)<=a32);
26 automne_C1 = automne(a1,:); automne_C2 = automne(a2,:); automne_C3 = ...
  automne(a3,:);
27 % Regrouper les deux premières classes
28 automne_C1C2 = [automne_C1; automne_C2];

```

A.13.1 Domaines de validation

```

1 % Pour chaque matrice créé dans la section précédente, récupérer la ...
  valeur minimale et la valeur maximale de chaque paramètre. ...
  Exemple pour la saison du printemps
2 % C1
3 interv_sst_C1 = [min(printemps_C1(:,11)) max(printemps_C1(:,11))];
4 interv_sss_C1 = [min(printemps_C1(:,10)) max(printemps_C1(:,10))];
5 interv_chl_C1 = [min(printemps_C1(:,15)) max(printemps_C1(:,15))];
6 % C2
7 interv_sst_C2 = [min(printemps_C2(:,11)) max(printemps_C2(:,11))];
8 interv_sss_C2 = [min(printemps_C2(:,10)) max(printemps_C2(:,10))];
9 interv_chl_C2 = [min(printemps_C2(:,15)) max(printemps_C2(:,15))];
10 % C3
11 interv_sst_C3 = [min(printemps_C3(:,11)) max(printemps_C3(:,11))];
12 interv_sss_C3 = [min(printemps_C3(:,10)) max(printemps_C3(:,10))];
13 interv_chl_C3 = [min(printemps_C3(:,15)) max(printemps_C3(:,15))];
14 % Sauvegarder
15 % C1
16 save('Domaines_Validation_printemps_c1', 'interv_sst_C1', ...
      'interv_sss_C1', 'interv_chl_C1')
17 % C2
18 save('Domaines_Validation_printemps_c2', 'interv_sst_C2', ...
      'interv_sss_C2', 'interv_chl_C2')
19 % C3
20 save('Domaines_Validation_printemps_c3', 'interv_sst_C3', ...
      'interv_sss_C3', 'interv_chl_C3')

```

A.14 Répartitions des BDs pour la construction des RNs

```

1 % Ceci est un exemple de la saison d'Été
2 % Charger les données de la classe souhaitée, ici c'est la C3
3 load('D:\Documents\Data_Insitu\Ete', 'ete_C3')
4 ete_C3_BD_Appr = []; % Tableau pour l'apprentissage
5 ete_C3_BD_Valid = []; % Tableau pour la validation
6 % Tableaux pour le test
7 ete_C3_BD_Test_1 = []; % de 2000 à 2009, utilisation de SSS SODA
8 ete_C3_BD_Test_2 = []; % de 2010 à 2014, utilisation de SSS SMOS
9 % Base de données d'apprentissage
10 i = 1; id = [];
11 while i <= size(ete_C3,1)
12     for j=i:i+3
13         id = [id; j];
14     end
15     i = i+5;
16 end
17 if id(end,1) == (size(ete_C3,1)+1)
18     id = id(1:end-1,:);

```

```

19 end
20 ete_C3_BD_Appr = ete_C3(id,:);
21 % Base de données de validation (1 point pour la validation, pour ...
    chaque 4 points pour l'apprentissage)
22 i = 5; id = [];
23 while i <= size(ete_C3,1)
24     id = [id; i];
25     i = i+5;
26 end
27 if id(end,1) == (size(ete_C3,1)+1)
28     id = id(1:end-1,:);
29 end
30 ete_C3_BD_Valid = ete_C3(id,:);
31 % Bases de données de test
32 a = find(ete_C3(:,1)<2010); % utilisation de SSS SODA
33 ete_C3_BD_Test_1 = ete_C3(a,:);
34 ete_C3_BD_Test_1 = ete_C3_BD_Test_1(((ete_C3_BD_Test_1(:,13)~=0 | ...
    ete_C3_BD_Test_1(:,14)~=0 | ete_C3_BD_Test_1(:,15)~=0) & ...
    ete_C3_BD_Test_1(:,16)~=0 & ete_C3_BD_Test_1(:,19)~=0), :);
35 a = find(ete_C3(:,1)>=2010); % utilisation de SSS SMOS
36 ete_C3_BD_Test_2 = ete_C3(a,:);
37 ete_C3_BD_Test_2 = ete_C3_BD_Test_2(((ete_C3_BD_Test_2(:,13)~=0 | ...
    ete_C3_BD_Test_2(:,14)~=0 | ete_C3_BD_Test_2(:,15)~=0) & ...
    ete_C3_BD_Test_2(:,16)~=0 & ete_C3_BD_Test_2(:,24)~=0), :);
38
39 % Eliminer les points ne contenant pas de données de Chl [13: ...
    Latitude sat, 14: Longitude sat, 15: Chl-a sat]
40 ete_C3_BD_Appr = ete_C3_BD_Appr((ete_C3_BD_Appr(:,13)~=0 | ...
    ete_C3_BD_Appr(:,14)~=0 | ete_C3_BD_Appr(:,15)~=0), :);
41 ete_C3_BD_Valid = ete_C3_BD_Valid((ete_C3_BD_Valid(:,13)~=0 | ...
    ete_C3_BD_Valid(:,14)~=0 | ete_C3_BD_Valid(:,15)~=0), :);
42 % Transformation log10 des données de chl-a
43 ete_C3_BD_Appr(:,15) = log10(ete_C3_BD_Appr(:,15));
44 ete_C3_BD_Valid(:,15) = log10(ete_C3_BD_Valid(:,15));
45 ete_C3_BD_Test_1(:,15) = log10(ete_C3_BD_Test_1(:,15));
46 ete_C3_BD_Test_2(:,15) = log10(ete_C3_BD_Test_2(:,15));

```

A.15 Construction des réseaux de neurones

```

1 % Ceci est un exemple de la classe 3 de l'Été
2 % Charger les BDs construites suivant le code précédent
3 load('D:\Documents\Data_Insitu\Ete\C3\BDs_App_Val_Tst.mat')
4
5 % ajout d'un paramètre pour distinguer entre le Nord et le sud [8: ...
    Latitude in situ, 13: Latitude sat]
6 lat_app = ete_C3_BD_Appr(:,8)>=0;
7 lat_valid = ete_C3_BD_Valid(:,8)>=0;
8 lat_test_1 = ete_C3_BD_Test_1(:,13)>=0;
9 lat_test_2 = ete_C3_BD_Test_2(:,13)>=0;
10 % Définir les entrées et les sorties désirées du RN
11 % Apprentissage

```

```

12 input = [lat_app, ete_C3_BD_Appr(:, [11,10,15])]; % SST, SSS, et Chl-a
13 target= ete_C3_BD_Appr(:, 12); % fCO2
14 x = input';
15 t = target';
16 %
17 Validation
18 x1 = [lat_valid, ete_C3_BD_Valid(:, [11,10,15])]' ;
19 t1 = ete_C3_BD_Valid(:, 12)' ;
20 % Test
21 x11 = [lat_test_1, ete_C3_BD_Test_1(:, [16,19,15])]' ;
22 t11 = ete_C3_BD_Test_1(:, 12)' ;
23 x12 = [lat_test_2, ete_C3_BD_Test_2(:, [16,24,15])]' ;
24 t12 = ete_C3_BD_Test_2(:, 12)' ;
25 % Choisir le nombre minimum et maximum des neurones cachés
26 nb1=1;
27 nb2=20;
28 % Création des tableaux qui vont contenir les résultats
29 % Apprentissage
30 appr_all = cell(nb2,4);
31 % Validation
32 valid_all = zeros(nb2,2);
33 % Test
34 test_all = zeros(nb2,2);
35 % Construction des RNs
36 for i=nb1:nb2
37     % Définir le nombre de neurones cachés
38     hiddenLayerSize = i;
39     % Définir l'algorithmme d'apprentissage 'trainlm'
40     net = feedforwardnet(hiddenLayerSize, 'trainlm');
41     % Proportions de la répartition de la base d'apprentissage en ...
42     3 sous-bases (train, val, et test)
43     net.divideParam.trainRatio = 80/100;
44     net.divideParam.valRatio = 10/100;
45     net.divideParam.testRatio = 10/100;
46     % Définir le nombre maximum d'itération
47     net.trainParam.epochs = 1000;
48     % Définir la fonction de performance
49     net.performFcn = 'mse';
50     %+++++
51     % Étape d'apprentissage
52     %+++++
53     % Lancer l'apprentissage du RN en cours
54     [net,tr] = train(net,x,t);
55     % Calculer la performance du RN
56     y = net(x);
57     e = gsubtract(t,y);
58     performance = perform(net,t,y);
59     % Enregistrer le RN construit et ses paramètres
60     appr_all{i,1} = net;
61     appr_all{i,2} = tr;
62     appr_all{i,3} = sqrt(performance);
63     appr_all{i,4} = mean(e);
64     %+++++
65     % Étape de validation
66     %+++++

```

```

66     y1 = net(x1);
67     e1 = gsubtract(t1,y1);
68     performance = perform(net,t1,y1);
69     valid_all(i,1) = sqrt(performance);
70     valid_all(i,2) = mean(e1);
71     %+++++
72     % Etape de test
73     %+++++
74     x2 = [x11 x12];
75     t2 = [t11 t12];
76     y2 = net(x2);
77     e2 = gsubtract(t2,y2);
78     performance = perform(net,t2,y2);
79     test_all(i,1) = sqrt(performance);
80     test_all(i,2) = mean(e2);
81 end
82 % Sauvegarder
83 save('RES_RNs.mat', 'appr_all', 'valid_all', 'test_all')

```

A.15.1 Choix du RN optimal

```

1 % Charger les 20 RNs créés dans la section précédente
2 load('RES_RNs.mat')
3 % Récupérer le meilleur RN
4 [m,id] = min(test_all(:,1));
5 net = appr_all{id,1};
6 % Sauvegarder ce RN
7 save('RN_Optimal.mat', net)

```

A.16 Élaboration des fichiers CSV

A.16.1 Regroupement de SST-Chl-*a* avec SSS

```

1 % Ceci est un exemple de la saison du printemps
2 %% ***** SODA *****
3 % Chemin des data SSS SODA et SST/CHL MODIS
4 chem_soda = 'D:\Documents\Data_Climato\SODA\SSS\';
5 chem_modis = 'D:\Documents\Data_Sat\MODISA\SST_CHL\';
6 % Pour le printemps la data sont disponibles à partir de 2003
7 for i=2003:2009
8     for j=3:5
9         % Créer un nouveau tableau
10        eval(['tab_SST_CHL_SSS_',num2str(i),'_',num2str(j),' = [];'])
11        % Récupérer les data SSS
12        if (length(j)==1)
13            jj = ['0',num2str(j)];
14        end
15        load([chem_soda,num2str(i),'\SODA_SSS_SST_',num2str(i), ...

```

```

16     '_' , jj, '.mat' ]]);
17     eval([ 'data = tab_', num2str(i), '_', jj, ';' ] )
18     eval([ 'clear tab_', num2str(i), '_', jj ] )
19     % Supprimer les lignes contenant NaN
20     data = data(~any(isnan(data(:,1:3)),2),:);
21     % Récupérer les data SST_CHL, vérifier si le dossier et le ...
        fichier existent
22     if (exist([chem_modis,num2str(i)], 'dir') ~= 0) & ...
        (exist([chem_modis,num2str(i), '\SST_CHL', num2str(i), '_', ...
23     num2str(j), '.mat'], 'file') ~= 0)
24         load([chem_modis,num2str(i), '\SST_CHL', num2str(i), ...
25         '_', num2str(j), '.mat' ]]);
26         eval([ 'tab = tab', num2str(i), '_', num2str(j), ';' ] )
27         eval([ 'clear tab', num2str(i), '_', num2str(j) ] )
28         % Rechercher les points SSS correspondants, pour ceci, le ...
            tableau est divisé en 20 parties pour accélérer le ...
            traitement
29         s_10_init = size(tab,1)/20;
30         s_10_init = str2num(sprintf('%16.f',s_10_init));
31         s_10 = s_10_init;
32         k = 1; l=1;
33         while s_10<=size(tab,1)
34             eval([ 'tab', num2str(l), '= tab(k:s_10, :);' ] )
35             % Calculer la distance euclidienne
36             eval([ 'D', num2str(l), ' = ...
                pdist2(tab', num2str(l), '(:, [2,1]), data(:,1:2));' ] )
37             % Rechercher le minimum de chaque ligne
38             eval([ '[dist, id] = min(D', num2str(l), ', [], 2);' ] )
39             % Eliminer les distance > 0.5
40             a = find(dist<=0.5);
41             eval([ 'tab', num2str(l), ' = ...
                tab', num2str(l), '(a, :);' ] )
42             dist = dist(a);
43             id = id(a);
44             % Combiner les valeurs
45             eval([ ...
                'tab_SST_CHL_SSS_', num2str(i), '_', num2str(j), ' = ...
                [tab_SST_CHL_SSS_', num2str(i), '_', num2str(j), ';' ...
                tab', num2str(l), '(:, [2,1,3,4]) data(id,1:3)];' ] )
46             eval([ 'clear D', num2str(l), ' dist id tab', num2str(l) ] )
47             k = s_10+1;
48             s_10 = s_10+s_10_init;
49             l = l+1;
50         end
51         if l == 20
52             d = s_10 - size(tab,1);
53             s_10 = s_10 - d;
54             eval([ 'tab', num2str(l), '= tab(k:s_10, :);' ] )
55             eval([ 'D', num2str(l), ' = ...
                pdist2(tab', num2str(l), '(:, [2,1]), data(:,1:2));' ] )
56             eval([ '[dist, id] = min(D', num2str(l), ', [], 2);' ] )
57             a=find(dist<=0.5);
58             eval([ 'tab', num2str(l), ' = tab', num2str(l), '(a, :);' ] )
59             dist = dist(a);
60             id = id(a);

```



```

61         eval([ 'tab_SST_CHL_SSS_', num2str(i), '_', num2str(j), ' ...
              = [tab_SST_CHL_SSS_', num2str(i), '_', num2str(j), '; ...
              tab', num2str(l), '(:, [2,1,3,4]) data(id,1:3)];' ])
62         eval([ 'clear D', num2str(l), ' dist id tab', num2str(l)])
63     end
64 end
65 % Sauvegarde
66 save(['SST_CHL_SSS_', num2str(i), '_', num2str(j)], ...
      ['tab_SST_CHL_SSS_', num2str(i), '_', num2str(j)])
67 eval(['clear tab_SST_CHL_SSS_', num2str(i), '_', num2str(j)])
68 end
69 end
70 %% ***** SMOS *****
71 %Chemin des data SSS SMOS et SST/CHL MODIS
72 chem_smos = 'D:\Documents\Data_Sat\SMOS\SSS\';
73 chem_modis = 'D:\Documents\Data_Sat\MODISA\SST_CHL\';
74 for i=2010:2013
75     for j=3:5
76         % Créer un nouveau tableau
77         eval(['tab_SST_CHL_SSS_', num2str(i), '_', num2str(j), ' = [];'])
78         % Récupérer les data SSS
79         load([chem_smos, num2str(i), '\SMOS_SSS_', num2str(i), ...
80              '_', num2str(j)]);
81         eval(['data = tab_', num2str(i), '_', num2str(j), ';'])
82         eval(['clear tab_', num2str(i), '_', num2str(j) ])
83         % Supprimer les lignes contenant NaN
84         data = data(~any(isnan(data),2),:);
85         % Récupérer les data SST_CHL, vérifier si le dossier et le ...
            fichier existent
86         if (exist([chem_modis, num2str(i)], 'dir') ~= 0) & ...
            (exist([chem_modis, num2str(i), '\SST_CHL', num2str(i), '_', ...
87                 num2str(j), '.mat'], 'file') ~= 0)
88             load([chem_modis, num2str(i), '\SST_CHL', num2str(i), ...
89                  '_', num2str(j), '.mat']);
90             eval(['tab = tab', num2str(i), '_', num2str(j), ';'])
91             eval(['clear tab', num2str(i), '_', num2str(j) ])
92             % Rechercher les points SSS correspondants, pour ceci, le ...
                tableau est divisé en 20 parties pour accélérer le ...
                traitement
93             s_10_init = size(tab,1)/20;
94             s_10_init = str2num(sprintf('%16.f', s_10_init));
95             s_10 = s_10_init;
96             k = 1; l=1;
97             while s_10<=size(tab,1)
98                 eval(['tab', num2str(l), '= tab(k:s_10, :);'])
99                 % Calculer la distance euclidienne
100                eval(['D', num2str(l), ' = ...
                    pdist2(tab', num2str(l), '(:, [2,1]), data(:,1:2));' ])
101                % Rechercher le minimum de chaque ligne
102                eval(['[dist, id] = min(D', num2str(l), ', [],2);'])
103                % Eliminer les distance > 0.25
104                a = find(dist<=0.25);
105                eval(['tab', num2str(l), ' = ...
                    tab', num2str(l), '(a,:);'])
106                dist = dist(a);

```

```

107         id = id(a);
108         % Combiner les valeurs
109         eval([ ...
                'tab_SST_CHL_SSS_', num2str(i), '_', num2str(j), ' = ...
                [tab_SST_CHL_SSS_', num2str(i), '_', num2str(j), '; ...
                tab', num2str(l), '(:, [2,1,3,4]) data(id,:);' ])
110         eval(['clear D', num2str(l), ' dist id tab', num2str(l)])
111         k = s_10+1;
112         s_10 = s_10+s_10_init;
113         l = l+1;
114     end
115     if l == 20
116         d = s_10 - size(tab,1);
117         s_10 = s_10 - d;
118         eval(['tab', num2str(l), ' = tab(k:s_10, :);' ])
119         eval(['D', num2str(l), ' = ...
                pdist2(tab', num2str(l), '(:, [2,1]), data(:,1:2));' ])
120         eval(['[dist, id] = min(D', num2str(l), ', [], 2);' ])
121         a=find(dist<=0.25);
122         eval(['tab', num2str(l), ' = tab', num2str(l), '(a,:);' ])
123         dist = dist(a);
124         id = id(a);
125         eval(['tab_SST_CHL_SSS_', num2str(i), '_', num2str(j), ' ...
                = [tab_SST_CHL_SSS_', num2str(i), '_', num2str(j), '; ...
                tab', num2str(l), '(:, [2,1,3,4]) data(id,:);' ])
126         eval(['clear D', num2str(l), ' dist id tab', num2str(l)])
127     end
128 end
129 % Sauvegarde
130 save(['SST_CHL_SSS_', num2str(i), '_', num2str(j)], ...
        ['tab_SST_CHL_SSS_', num2str(i), '_', num2str(j)])
131 end
132 end

```

A.16.2 Calcul de la fCO₂

```

1 % Charger les RNs des trois classes
2 chem_rn = 'D:\Documents\Data_Insitu\PRINTEMPS\';
3 % C1
4 load([chem_rn, 'C1\RN_Optimal.mat'], 'net') % Récupérer la structure ...
        contenant le RN
5 net1 = net;
6 % C2
7 load([chem_rn, 'C2\RN_Optimal.mat'], 'net')
8 net2 = net;
9 % C3
10 load([chem_rn, 'C3\RN_Optimal.mat'], 'net')
11 net3 = net; clear net
12 % Charger les intervalles (domaines) de validation
13 % C1
14 load([chem_rn, 'C1\Domaines_Validation_printemps_c1'])
15 % C2

```

```

16 load([chem_rn, 'C2\Domaines_Validation_printemps_c2'])
17 % C3
18 load([chem_rn, 'C3\Domaines_Validation_printemps_c3'])
19 % Calculer la fCO2
20 for i=2003:2013
21     for j=3:5
22         % Charger la matrice créée dans la section précédente
23         eval([ ...
24             'load(''SST_CHL_SSS_', num2str(i), '_', num2str(j), '.mat'')' ])
25         eval([ 'tab = ...
26             tab_SST_CHL_SSS_', num2str(i), '_', num2str(j), ';' ])
27         eval([ 'clear ...
28             tab_SST_CHL_SSS_', num2str(i), '_', num2str(j), ';' ])
29         % Diviser la matrice en 3 suivant les domaines de validation ...
30         de chaque classe
31         % C1
32         a = find( (tab(:,3)>=interv_sst_C1(1,1) & ...
33             tab(:,3)<=interv_sst_C1(1,2)) & ...
34             (tab(:,4)>=interv_chl_C1(1,1) & ...
35             tab(:,4)<=interv_chl_C1(1,2)) & ...
36             (tab(:,7)>=interv_sss_C1(1,1) & ...
37             tab(:,7)<=interv_sss_C1(1,2)));
38         % C2
39         b = find( (tab(:,3)>=interv_sst_C2(1,1) & ...
40             tab(:,3)<=interv_sst_C2(1,2)) & ...
41             (tab(:,4)>=interv_chl_C2(1,1) & ...
42             tab(:,4)<=interv_chl_C2(1,2)) & ...
43             (tab(:,7)>=interv_sss_C2(1,1) & ...
44             tab(:,7)<=interv_sss_C2(1,2)));
45         % C3
46         c = find( (tab(:,3)>=interv_sst_C3(1,1) & ...
47             tab(:,3)<=interv_sst_C3(1,2)) & ...
48             (tab(:,4)>=interv_chl_C3(1,1) & ...
49             tab(:,4)<=interv_chl_C3(1,2)) & ...
50             (tab(:,7)>=interv_sss_C3(1,1) & ...
51             tab(:,7)<=interv_sss_C3(1,2)));
52         % Ajouter une colonne contenant les identifiants des lignes ...
53         afin de conserver l'ordre des lignes après la division
54         tab = [[1:1:size(tab,1)]' tab];
55         % Division
56         tab_C1 = tab(a,:); tab_C2 = tab(b,:); tab_C3 = tab(c,:);
57         % Vecteurs des latitudes pour les RNs
58         lat_C1 = tab_C1(:,3) >= 0;
59         lat_C2 = tab_C2(:,3) >= 0;
60         lat_C3 = tab_C3(:,3) >= 0;
61         % Appliquer le log10 sur la CHL
62         tab_C1(:,5) = log10(tab_C1(:,5)); tab_C2(:,5) = ...
63             log10(tab_C2(:,5)); tab_C3(:,5) = log10(tab_C3(:,5));
64         % Calcul de la fco2
65         % C1
66         tab_C1(:,9) = net1( [lat_C1, tab_C1(:,[4,8,5])] );
67         % C2
68         tab_C2(:,9) = net2( [lat_C2, tab_C2(:,[4,8,5])] );
69         % C3
70         tab_C3(:,9) = net3( [lat_C3, tab_C3(:,[4,8,5])] );

```

```

56     % Combiner les trois matrices
57     tab = [tab_C1; tab_C2; tab_C3];
58     % Réordonner les lignes
59     eval([ 'tab_SST_CHL_SSS_', num2str(i), '_', num2str(j), ' = ...
            sortrows(tab,1);' ])
60     % Sauvegarder la matrice
61     save(['SST_CHL_SSS_', num2str(i), '_', num2str(j)], ...
          ['tab_SST_CHL_SSS_', num2str(i), '_', num2str(j)])
62     eval(['clear tab_SST_CHL_SSS_', num2str(i), '_', num2str(j)])
63     end
64 end

```

A.16.3 Production des fichiers CSV

```

1  chem = 'D:\Documents\Data_Insitu\PRINTEMPS\';
2  % Charger les domaines de validation
3  %C1
4  load([chem, 'C1\Domaines_Validation_printemps_c1'])
5  %C2
6  load([chem, 'C2\Domaines_Validation_printemps_c2'])
7  %C3
8  load([chem, 'C3\Domaines_Validation_printemps_c3'])
9
10 for i=2003:2013
11     for j=3:5
12         % Charger le fichier des données compilées, créé dans la ...
13             section précédente
14         eval([ ...
15             'load(''SST_CHL_SSS_', num2str(i), '_', num2str(j), '.mat'')' ])
16         eval([ 'tab = ...
17             tab_SST_CHL_SSS_', num2str(i), '_', num2str(j), ';' ])
18         eval([ 'clear ...
19             tab_SST_CHL_SSS_', num2str(i), '_', num2str(j), ';' ])
20
21         tab(:,5) = 10.^tab(:,5);
22         % Appliquer les domaines de validation
23         a = find( (tab(:,4)>=interv_sst_C1(1,1) & ...
24                 tab(:,4)<=interv_sst_C1(1,2)) & ...
25                 (tab(:,5)>=interv_chl_C1(1,1) & ...
26                 tab(:,5)<=interv_chl_C1(1,2)) & ...
27                 (tab(:,8)>=interv_sss_C1(1,1) & ...
28                 tab(:,8)<=interv_sss_C1(1,2)) & ...
29                 (tab(:,9)>=interv_fco2_C1(1,1) & ...
30                 tab(:,9)<=interv_fco2_C1(1,2)));
31         b = find( (tab(:,4)>=interv_sst_C2(1,1) & ...
32                 tab(:,4)<=interv_sst_C2(1,2)) & ...
33                 (tab(:,5)>=interv_chl_C2(1,1) & ...
34                 tab(:,5)<=interv_chl_C2(1,2)) & ...
35                 (tab(:,8)>=interv_sss_C2(1,1) & ...
36                 tab(:,8)<=interv_sss_C2(1,2)) & ...
37                 (tab(:,9)>=interv_fco2_C2(1,1) & ...
38                 tab(:,9)<=interv_fco2_C2(1,2)));

```

```
27     c = find( (tab(:,4)>=interv_sst_C3(1,1) & ...
28             tab(:,4)<=interv_sst_C3(1,2)) & ...
29             (tab(:,5)>=interv_chl_C3(1,1) & ...
30             tab(:,5)<=interv_chl_C3(1,2)) & ...
31             (tab(:,8)>=interv_sss_C3(1,1) & ...
32             tab(:,8)<=interv_sss_C3(1,2)) & ...
33             (tab(:,9)>=interv_fco2_C3(1,1) & ...
34             tab(:,9)<=interv_fco2_C3(1,2)));
35     tab_C1 = tab(a,:); tab_C2 = tab(b,:); tab_C3 = tab(c,:);
36     tab = [tab_C1; tab_C2; tab_C3];
37     tab = sortrows(tab,1);
38     % Créer une colonne 'depth' pour l'affichage sur ODV
39     tab = [tab ones(size(tab,1),1)];
40     % Créer l'entête du fichier CSV
41     tab = [[{'id'}, {'longitude'}, {'latitude'}, {'SST'}, {'Chl-a'}, ...
42           {'lon'}, {'lat'}, {'SSS'}, {'fCO2'}, {'depth'}]; num2cell(tab)];
43     % Créer et sauvegarder le fichier CSV
44     fid = fopen([chem, ...
45               'fCO2_SAT\CSV_flies\CSV_fCO2_', num2str(i), '_', ...
46               num2str(j), '.csv'], 'w') ;
47     fprintf(fid, '%s', tab{1,1:end-1}) ;
48     fprintf(fid, '%s\n', tab{1,end}) ;
49     fclose(fid) ;
50     dlmwrite([chem, ...
51             'fCO2_SAT\CSV_flies\CSV_fCO2_', num2str(i), '_', ...
52             num2str(j), '.csv'], tab(2:end,:), '-append') ;
53     end
54 end
```


Annexe B

VARIABILITÉS DES PARAMÈTRES : été, automne

B.1 Été

B.1.1 Tableaux

B.1.1.1 SST

Année	Minimum (°C)	Maximum (°C)	Moyenne (°C)	Écart-type (°C)
2002	26,79	29,26	28,18	0,42
2003	24,51	29,17	27,6	1,18
2004	25,94	29,43	28,5	0,43
2005	17,36	30	25,68	4,16
2006	22	29,78	26,02	1,6
2007	23,82	29,51	27,18	1,58
2008	27,21	29,86	28,51	0,68
2009	24	29,57	27,12	1,05
2010	19,46	30,69	26,6	2,58
2011	17,9	29,47	25,55	2,3
2012	27,58	29,25	28,51	0,26
2013	27,3	29,25	28,35	0,59

TABLEAU B.1 – Paramètres statistiques de SST de toutes les années pendant la saison de l'été.

B.1.1.2 SSS

Année	Minimum	Maximum	Moyenne	Écart-type
2002	33,38	37,629	36,167	0,392
2003	32,848	37,401	35,907	0,548
2004	32,462	37,491	35,783	0,589
2005	33,811	37,257	35,936	0,457
2006	33,466	37,6	35,991	0,413
2007	32	37,644	36,096	0,822
2008	32,594	37,617	35,992	1,14
2009	32,001	37,725	35,954	0,728
2010	32,009	37,671	36,105	0,756
2011	32,004	37,49	36,212	0,663
2012	34,657	36,39	35,92	0,346
2013	32,321	37,6	35,282	1,759

TABLEAU B.2 – Paramètres statistiques de SSS de toutes les années pendant la saison de l'été.

B.1.1.3 fCO₂

Année	Minimum (μatm)	Maximum (μatm)	Moyenne (μatm)	Écart-type (μatm)
2002	355,08	393,69	373,84	6,29
2003	337,27	410,67	376,64	12,3
2004	348,34	399,03	383,49	5,85
2005	342,27	467,9	386	16,9
2006	332,86	493,5	389,07	17,36
2007	332,5	431,1	378,79	14,14
2008	354	408,16	369,8	14,56
2009	320,03	453,87	379,55	19,46
2010	333,2	455,52	390,93	18,08
2011	339,93	441,99	380,56	16,83
2012	377,42	400,36	389,59	3,89
2013	360,08	402,75	377,28	7,52

TABLEAU B.3 – Paramètres statistiques de fCO₂ de toutes les années pendant la saison de l'été.

B.1.1.4 Chl-*a*

Année	Minimum (mg.m ⁻³)	Maximum (mg.m ⁻³)	Moyenne (mg.m ⁻³)	Écart-type (mg.m ⁻³)
2002	-1,91	-0,6	-1,25	0,15
2003	-1,99	-0,12	-1,06	0,25
2004	-1,97	-0,4	-1,08	0,29
2005	-2,42	0,38	-0,9	0,47
2006	-1,98	0,82	-0,92	0,35
2007	-2,15	0,45	-1,14	0,33
2008	-2,63	-0,53	-1,33	0,26
2009	-2,41	0,87	-1,07	0,38
2010	-2,33	1,21	-0,84	0,51
2011	-2,3	0,3	-1,05	0,31
2012	-2,17	-0,74	-1,3	0,24
2013	-2,04	-0,05	-1,07	0,4

TABLEAU B.4 – Paramètres statistiques de Log₁₀(Chl-*a* satellitaire) de toutes les années pendant la saison de l'été.

Année	Minimum (mg.m ⁻³)	Maximum (mg.m ⁻³)	Moyenne (mg.m ⁻³)	Écart-type (mg.m ⁻³)
2002	0,01	0,25	0,06	0,02
2003	0,01	0,77	0,1	0,07
2004	0,01	0,4	0,1	0,06
2005	0,004	2,41	0,21	0,25
2006	0,01	6,68	0,17	0,26
2007	0,01	2,79	0,11	0,23
2008	0,002	0,29	0,06	0,04
2009	0,004	7,44	0,16	0,44
2010	0,005	16,19	0,35	0,74
2011	0,005	1,97	0,12	0,11
2012	0,01	0,18	0,06	0,03
2013	0,01	0,88	0,13	0,12

TABLEAU B.5 – Paramètres statistiques de Chl-*a* satellitaire de toutes les années pendant la saison de l'été.

B.1.2 Figures

B.1.2.1 SST

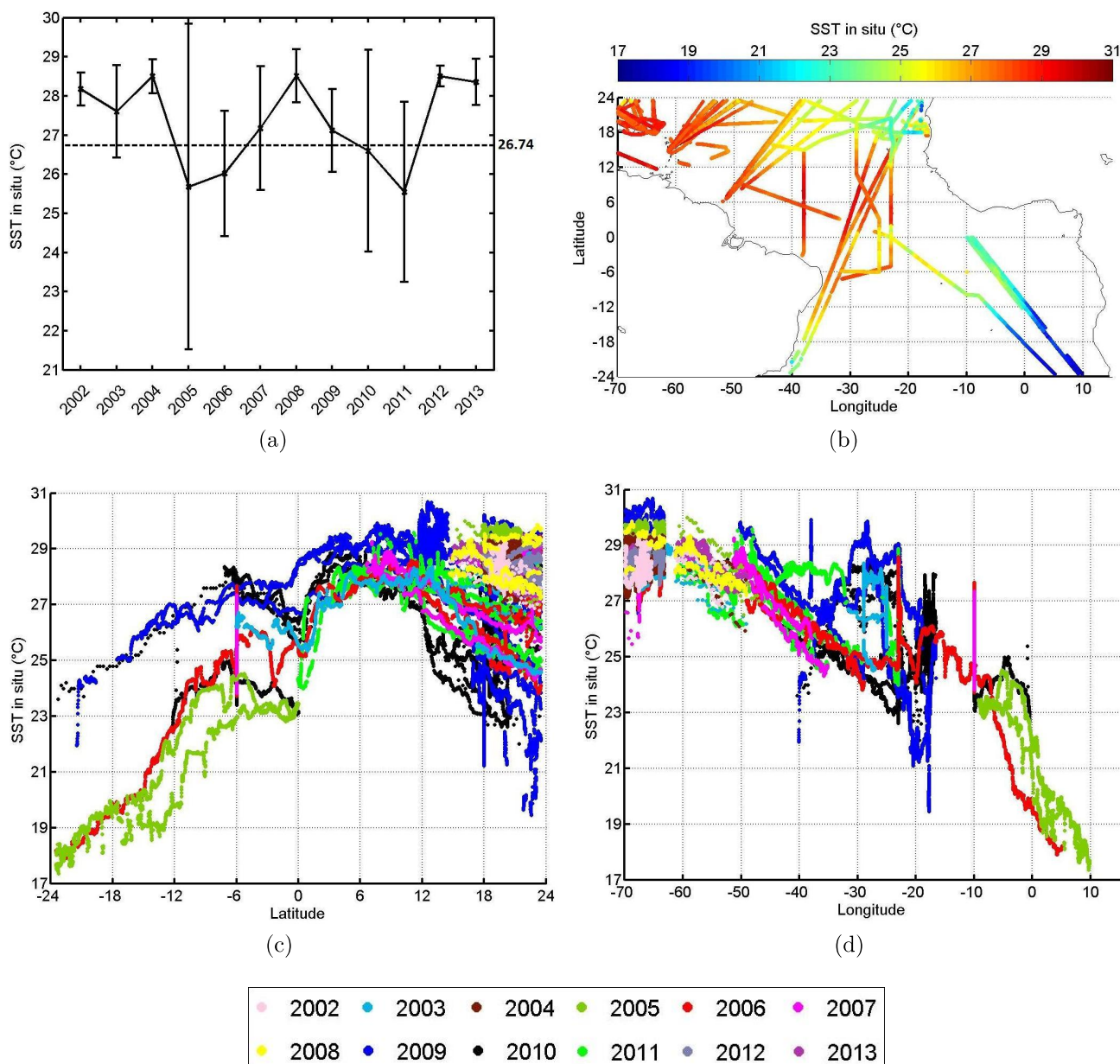


FIGURE B.1 – Représentation de la variabilité de SST *in situ* pendant la saison de l'été. (a) Moyenne et écarts-type de SST de chaque année, (b) Distribution spatiale de SST, (c) Variation de SST en fonction de la latitude, (d) Variation de SST en fonction de la longitude

B.1.2.2 SSS

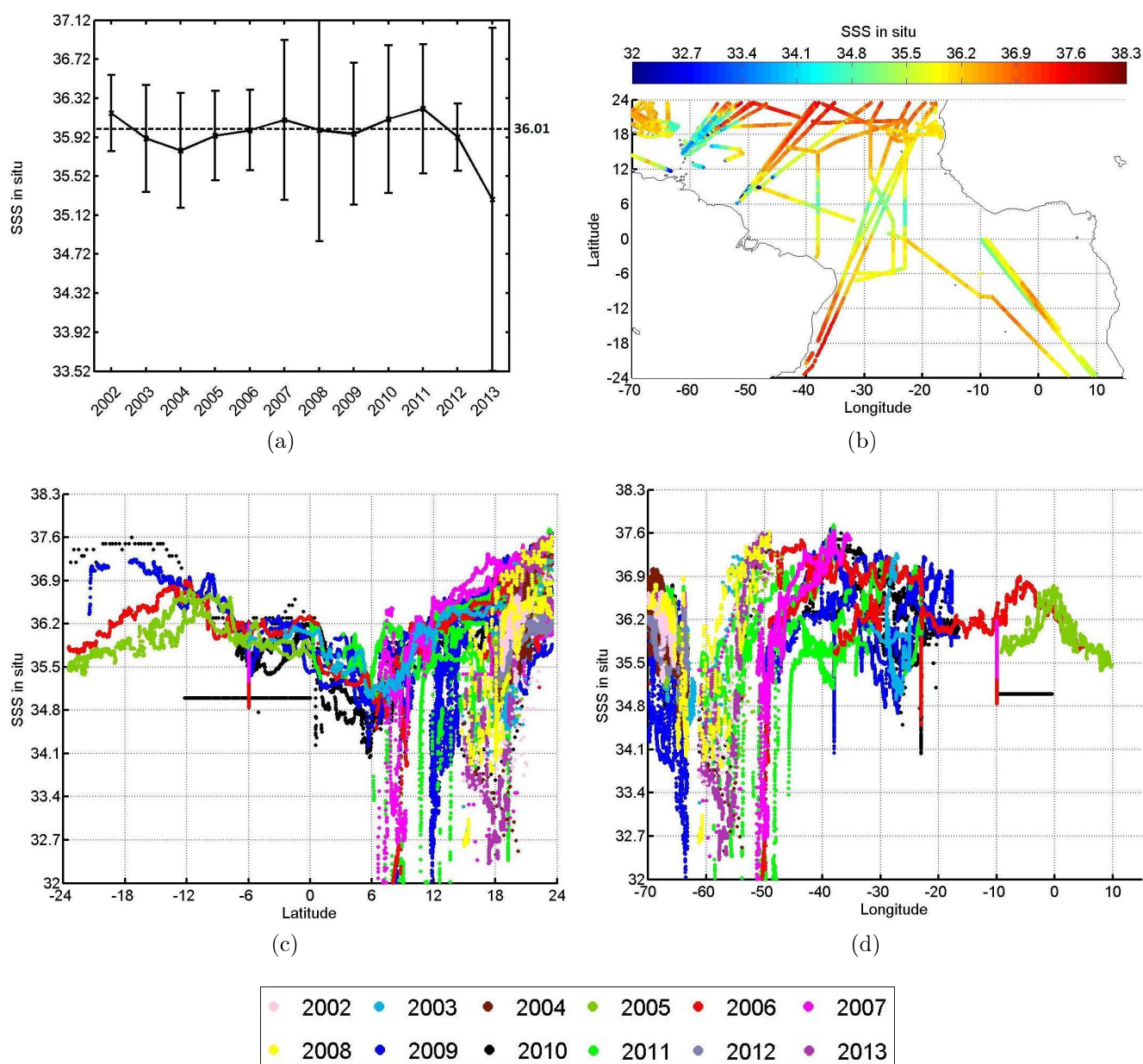


FIGURE B.2 – Représentation de la variabilité de SSS *in situ* pendant la saison de l’été. (a) Moyenne et écarts-type de SSS de chaque année, (b) Distribution spatiale de SSS, (c) Variation de SSS en fonction de la latitude, (d) Variation de SSS en fonction de la longitude

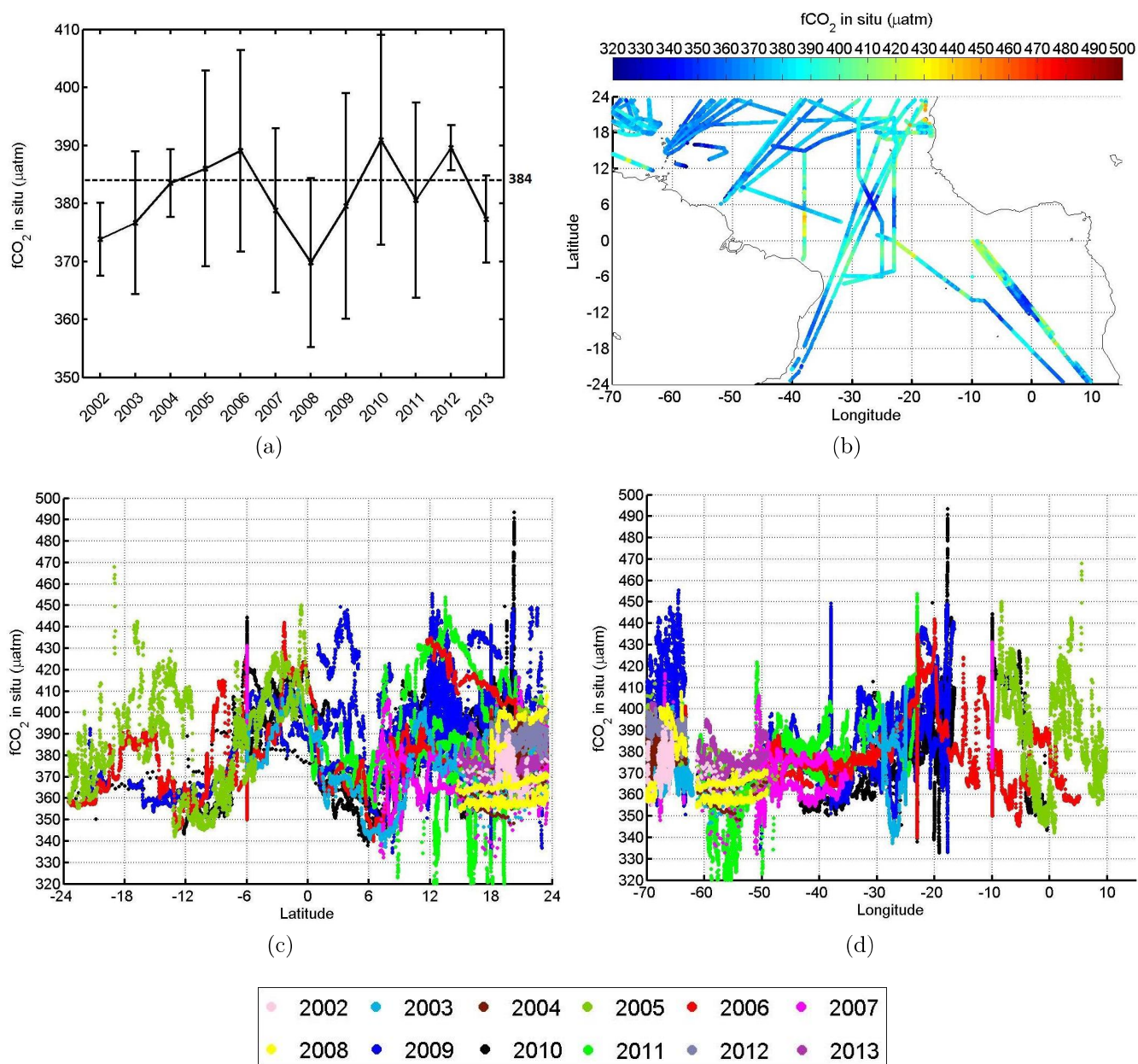
B.1.2.3 $f\text{CO}_2$ 

FIGURE B.3 – Représentation de la variabilité de $f\text{CO}_2$ *in situ* pendant la saison de l'été. (a) Moyenne et écarts-type de $f\text{CO}_2$ de chaque année, (b) Distribution spatiale de $f\text{CO}_2$, (c) Variation de $f\text{CO}_2$ en fonction de la latitude, (d) Variation de $f\text{CO}_2$ en fonction de la longitude

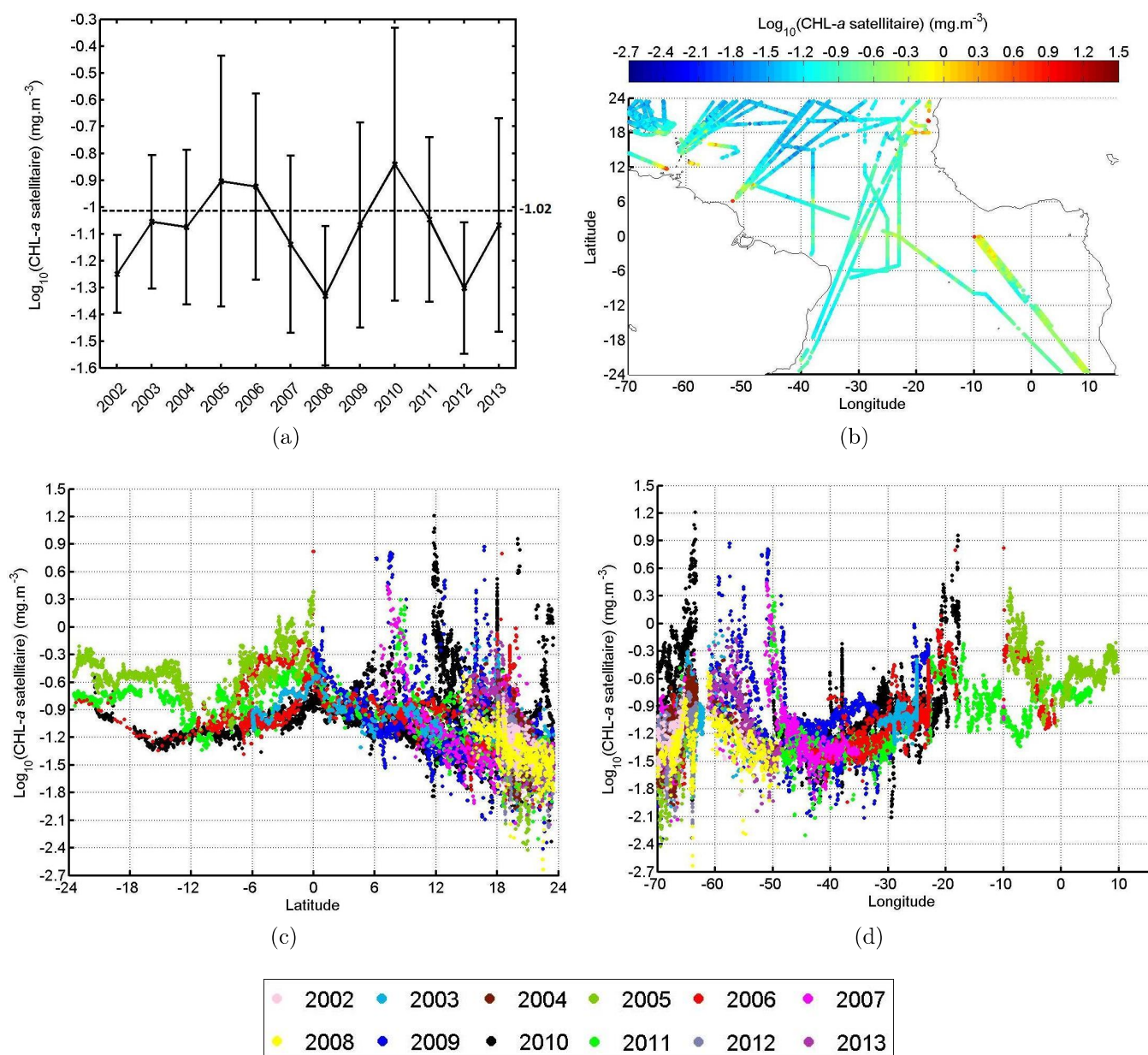
B.1.2.4 Chl-*a*

FIGURE B.4 – Représentation de la variabilité de $\text{Log}_{10}(\text{Chl-}a \text{ satellitaire})$ pendant la saison de l'été. (a) *Moyenne et écarts-type de Chl-a de chaque année*, (b) *Distribution spatiale de Chl-a*, (c) *Variation de Chl-a en fonction de la latitude*, (d) *Variation de Chl-a en fonction de la longitude*

B.2 Automne

B.2.1 Tableaux

B.2.1.1 SST

Année	Minimum (°C)	Maximum (°C)	Moyenne (°C)	Écart-type (°C)
2000	16,6	29,1	24,19	3,23
2001	26,23	28,77	27,58	0,69
2002	16,6	29,9	28,42	0,79
2003	27,25	30,03	29,22	0,26
2004	24,08	29,74	28,08	0,97
2005	17,62	30,72	26,93	3,3
2006	17,04	30,33	25,68	3,48
2007	22,48	30,1	26,9	1,75
2008	22,47	29,43	27,32	1,42
2009	24,39	29,42	26,7	1,52
2010	19,39	31,04	26,47	2,91
2011	22,55	29,1	26,82	1,59
2012	26,63	29,41	28,34	0,65
2013	25,02	29,31	27,14	1,13

TABLEAU B.6 – Paramètres statistiques de SST de toutes les années pendant la saison de l'automne.

B.2.1.2 SSS

Année	Minimum	Maximum	Moyenne	Écart-type
2000	33,9	37,458	35,917	0,561
2001	34,516	37,015	35,993	0,531
2002	33,9	37,538	35,411	0,831
2003	33,927	37,426	35,598	0,556
2004	33,852	37,52	35,706	0,667
2005	33,021	37,294	35,868	0,717
2006	33,025	37,358	35,697	0,635
2007	33,7	37,394	35,809	0,71
2008	33,378	37,22	35,992	0,857
2009	34,223	37,155	36,052	0,696
2010	32,026	37,754	35,825	0,873
2011	32,443	37,399	35,824	0,738
2012	34,308	36,988	35,561	0,692
2013	34,084	37,078	35,958	0,588

TABLEAU B.7 – Paramètres statistiques de SSS de toutes les années pendant la saison de l'automne.

B.2.1.3 fCO₂

Année	Minimum (μatm)	Maximum (μatm)	Moyenne (μatm)	Écart-type (μatm)
2000	320,04	444,1	370,32	22,61
2001	336,39	395,26	376,62	16,1
2002	320,04	411,23	355,96	17,29
2003	350,22	416,77	383,24	5,92
2004	339,18	418,89	376,11	14,43
2005	335,36	447,83	382,77	17,32
2006	331,29	452,7	394,6	26,62
2007	343,36	452,44	384,42	28,6
2008	325,34	394,73	360,68	11,79
2009	322,22	422,15	383,69	20,54
2010	328,64	441,4	380,91	21,63
2011	320,26	421,93	373,51	21,27
2012	353,23	399,42	371,56	10,9
2013	345,98	423,64	393,3	18,07

TABLEAU B.8 – Paramètres statistiques de fCO₂ de toutes les années pendant la saison de l'automne.

B.2.1.4 Chl-*a*

Année	Minimum (mg.m ⁻³)	Maximum (mg.m ⁻³)	Moyenne (mg.m ⁻³)	Écart-type (mg.m ⁻³)
2000	-1,45	-0,25	-0,94	0,17
2001	-1,53	-0,48	-0,92	0,2
2002	-1,45	-0,09	-0,95	0,21
2003	-1,62	-0,2	-1,17	0,15
2004	-1,77	-0,07	-1,1	0,18
2005	-1,93	0,14	-1,11	0,31
2006	-1,8	-0,08	-0,93	0,29
2007	-1,63	0,77	-1,04	0,28
2008	-2	0,69	-1,15	0,38
2009	-1,73	-0,24	-1,08	0,29
2010	-1,74	0,54	-0,95	0,29
2011	-1,99	0,6	-1,11	0,2
2012	-1,75	0,04	-1,26	0,13
2013	-1,76	-0,32	-1	0,18

TABLEAU B.9 – Paramètres statistiques de Log₁₀(Chl-*a* satellitaire) de toutes les années pendant la saison de l'automne.

Année	Minimum (mg.m ⁻³)	Maximum (mg.m ⁻³)	Moyenne (mg.m ⁻³)	Écart-type (mg.m ⁻³)
2000	0,04	0,57	0,12	0,06
2001	0,03	0,33	0,13	0,06
2002	0,04	0,81	0,13	0,07
2003	0,02	0,63	0,07	0,04
2004	0,02	0,85	0,09	0,05
2005	0,01	1,38	0,11	0,12
2006	0,02	0,84	0,15	0,11
2007	0,02	5,85	0,14	0,37
2008	0,01	4,89	0,14	0,32
2009	0,02	0,58	0,1	0,07
2010	0,02	3,46	0,15	0,15
2011	0,01	3,99	0,09	0,13
2012	0,02	1,11	0,06	0,03
2013	0,02	0,48	0,11	0,06

TABLEAU B.10 – Paramètres statistiques de Chl-*a* satellitaire de toutes les années pendant la saison de l'automne.

B.2.2 Figures

B.2.2.1 SST

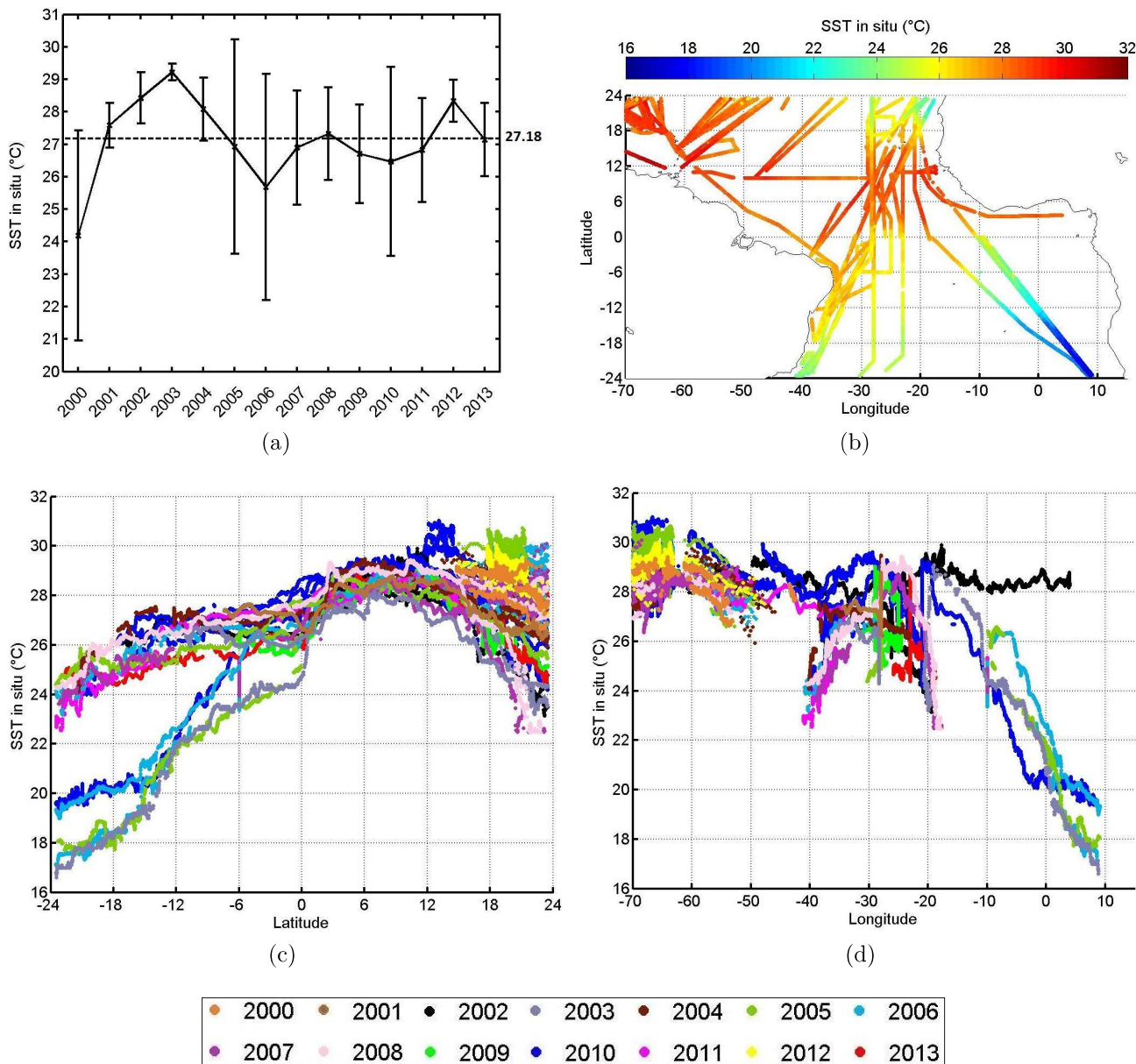


FIGURE B.5 – Représentation de la variabilité de SST *in situ* pendant la saison de l’automne. (a) Moyenne et écarts-type de SST de chaque année, (b) Distribution spatiale de SST, (c) Variation de SST en fonction de la latitude, (d) Variation de SST en fonction de la longitude

B.2.2.2 SSS

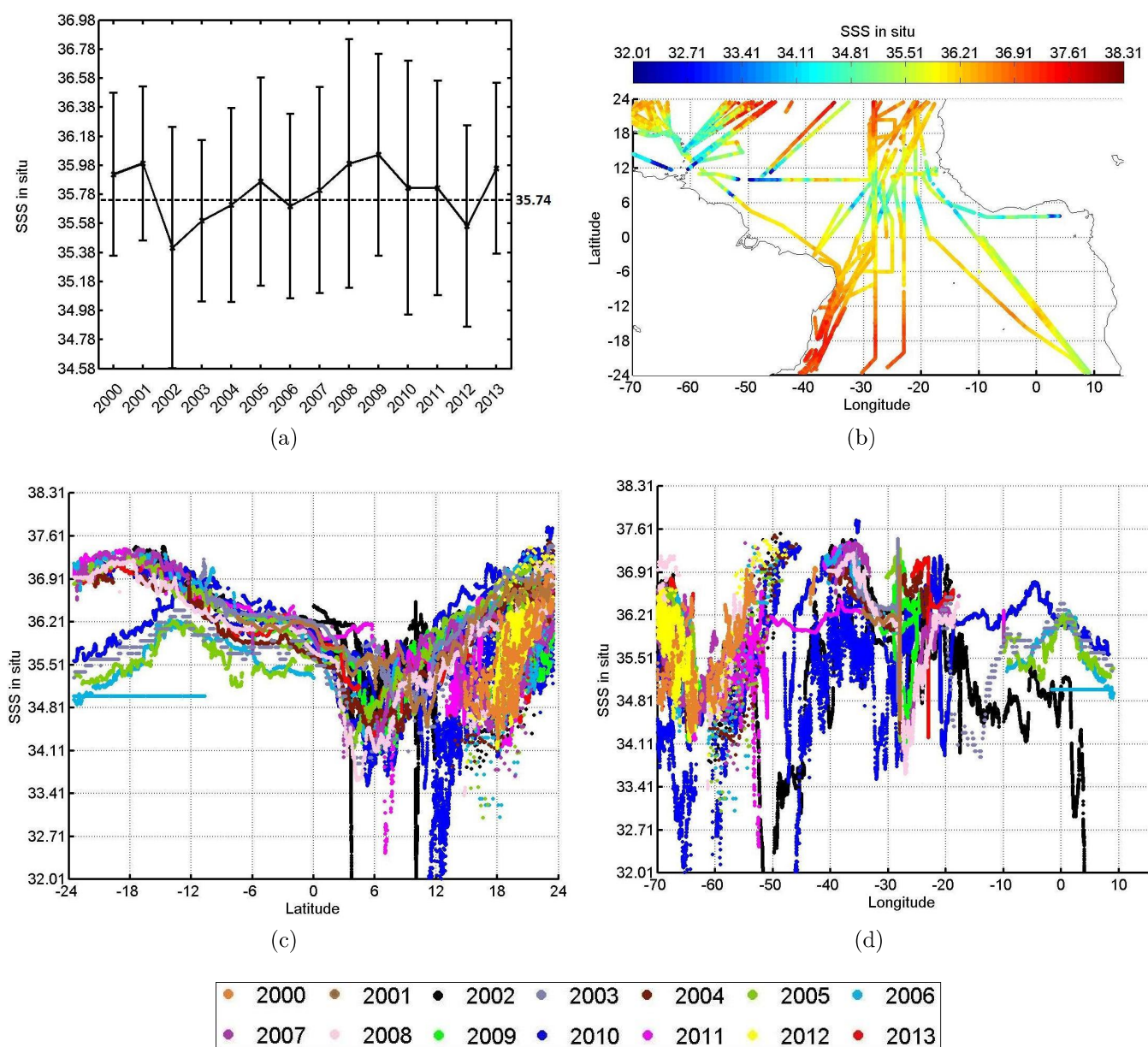


FIGURE B.6 – Représentation de la variabilité de SSS *in situ* pendant la saison de l'automne. (a) Moyenne et écarts-type de SSS de chaque année, (b) Distribution spatiale de SSS, (c) Variation de SSS en fonction de la latitude, (d) Variation de SSS en fonction de la longitude

B.2.2.3 fCO₂

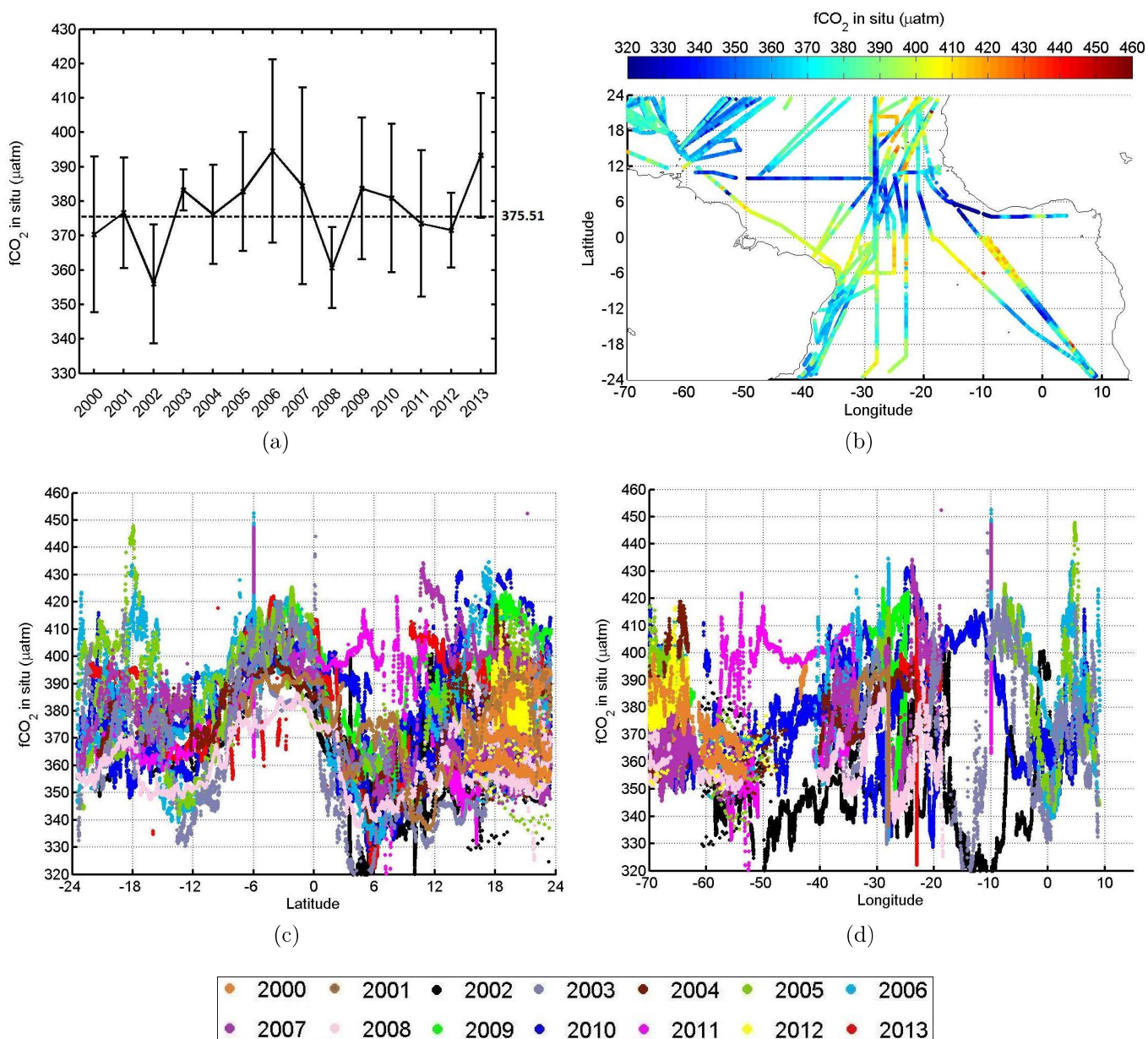


FIGURE B.7 – Représentation de la variabilité de fCO₂ *in situ* pendant la saison de l’automne. (a) Moyenne et écarts-type de fCO₂ de chaque année, (b) Distribution spatiale de fCO₂, (c) Variation de fCO₂ en fonction de la latitude, (d) Variation de fCO₂ en fonction de la longitude

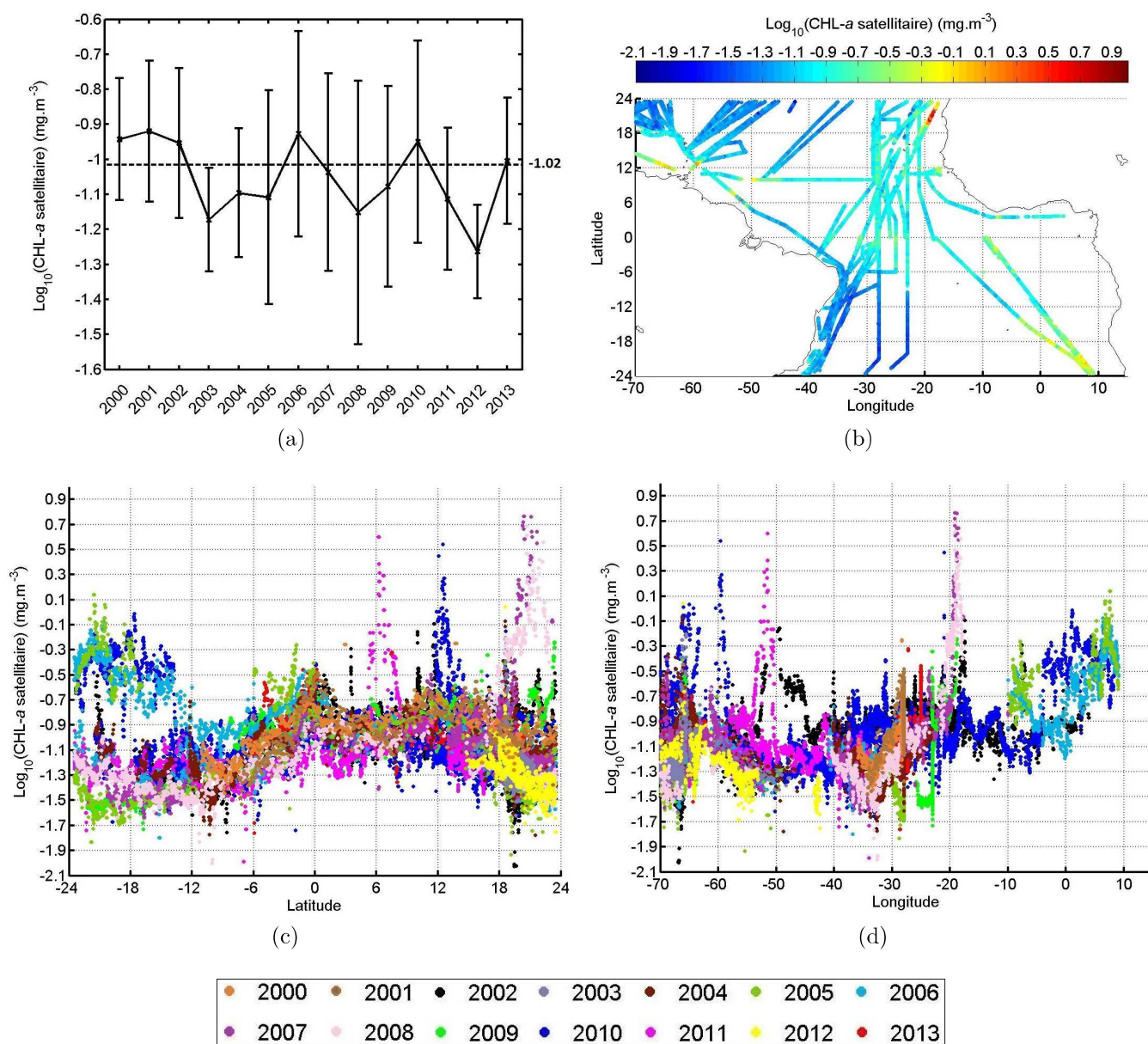
B.2.2.4 Chl-*a*

FIGURE B.8 – Représentation de la variabilité de $\text{Log}_{10}(\text{Chl-}a \text{ satellitaire})$ pendant la saison de l'automne. (a) Moyenne et écarts-type de Chl-*a* de chaque année, (b) Distribution spatiale de Chl-*a*, (c) Variation de Chl-*a* en fonction de la latitude, (d) Variation de Chl-*a* en fonction de la longitude

Annexe C

INTERPOLATION DE LA fCO₂ : été, automne

Après l'étape de classification des données *in situ* (**Section 5.5.2**), pour les deux saisons été et automne, la première classe obtenue (C1) contenait un nombre de points de données assez suffisant pour faire un apprentissage d'un RN (6127 et 8403, respectivement ; **Tableau 5.15**). Cependant, ces données étaient distribuées sur un nombre d'années très réduit (été : 2005, 2010, et 2011, automne : 2005, 2006, et 2010), et l'écart entre les années était trop important (été : 5 années entre 2005 et 2010, automne : 4 années entre 2006 et 2010). Ainsi, ces deux groupes de données ne peuvent pas être représentatifs de cette zone sur une grande échelle temporelle. Afin de palier à ce problème, pour les deux saisons, les deux premières classes ont été combinées pour former un seul set de données d'apprentissage nommé "classe C1C2".

C.1 Été

C.1.1 Classe 1 et 2

C.1.1.1 Apprentissage

Année	RMSE (μatm)	Biais (μatm)	NP *	PD (%) **
2003	7,4	0,01	575	2,68
2004	1,18	-1,18	1	0
2005	17,29	-3,4	2322	10,82
2006	12,79	0,21	3408	15,89
2007	13,5	-2,59	1959	9,13
2009	11,68	5,1	2385	11,12
2010	12,25	1,06	4835	22,54
2011	11,77	-0,98	5967	27,82
Total	12,81	-0,04	21452	/

TABEAU C.1 – Apprentissage de la classe C1C2 de la saison de l’été : RMSE et biais correspondants à chaque année. * nombre de points, ** pourcentage de données par rapport à la BD d’apprentissage

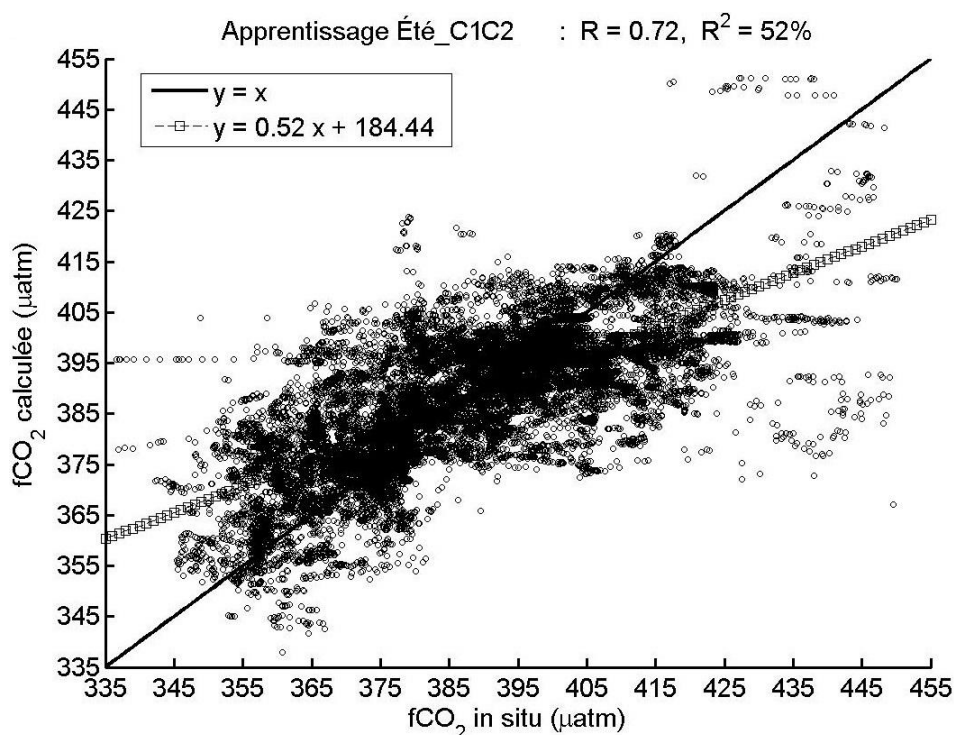


FIGURE C.1 – Apprentissage de la classe C1C2 de la saison de l’été : fCO₂ estimée par le RN en fonction de fCO₂ *in situ*.

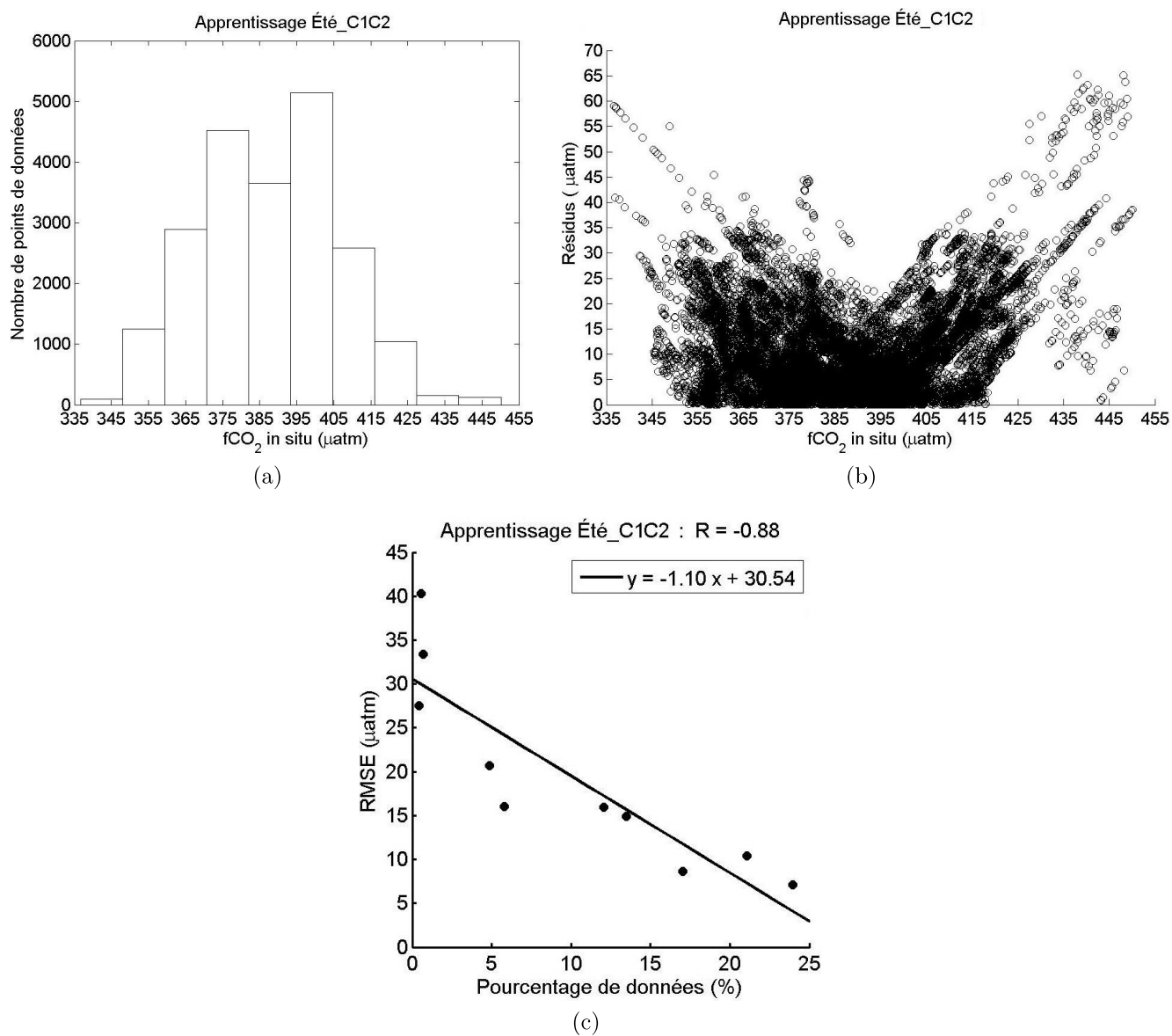


FIGURE C.2 – Apprentissage de la classe C1C2 de la saison de l'été : (a) *histogramme de fCO_2* , (b) *relation entre les résidus et les valeurs de fCO_2* , (c) *relation entre le pourcentage de données (par rapport à la BD d'apprentissage) de chaque intervalle de l'histogramme, et les RMSE correspondants.*

N°	Borne inf. (μatm)	Borne sup. (μatm)	NP *	PD (%) **	RMSE (μatm)
1	337,36	347,36	90	0,42	27,53
2	348,71	358,71	1246	5,81	16
3	360,05	370,05	2892	13,48	14,89
4	371,4	381,4	4523	21,08	10,39
5	382,75	392,75	3657	17,05	8,6
6	394,09	404,09	5142	23,97	7,15
7	405,44	415,44	2586	12,05	15,97
8	416,78	426,78	1045	4,87	20,7
9	428,13	438,13	150	0,7	33,4
10	439,47	449,47	121	0,56	40,33

TABLEAU C.2 – Apprentissage de la classe C1C2 de la saison de l'été : relation entre le nombre de points et le RMSE, pour chaque intervalle de l'histogramme de fCO_2 . * nombre de points ** pourcentage de données par rapport au nombre total de données de la BD d'apprentissage

C.1.1.2 Validation

Année	RMSE (μatm)	Biais (μatm)	NP *	PD (%) **
2003	7,51	-0,14	142	2,65
2005	17,49	-3,36	583	10,87
2006	12,88	0,22	852	15,89
2007	13,49	-2,61	489	9,12
2009	11,66	5,11	594	11,08
2010	12,35	1,11	1209	22,54
2011	11,71	-0,94	1494	27,86
Total	12,86	-0,02	5363	/

TABLEAU C.3 – Validation de la classe C1C2 de la saison de l'été : RMSE est biais correspondant à chaque année. * nombre de points, ** pourcentage de données par rapport à la BD de validation

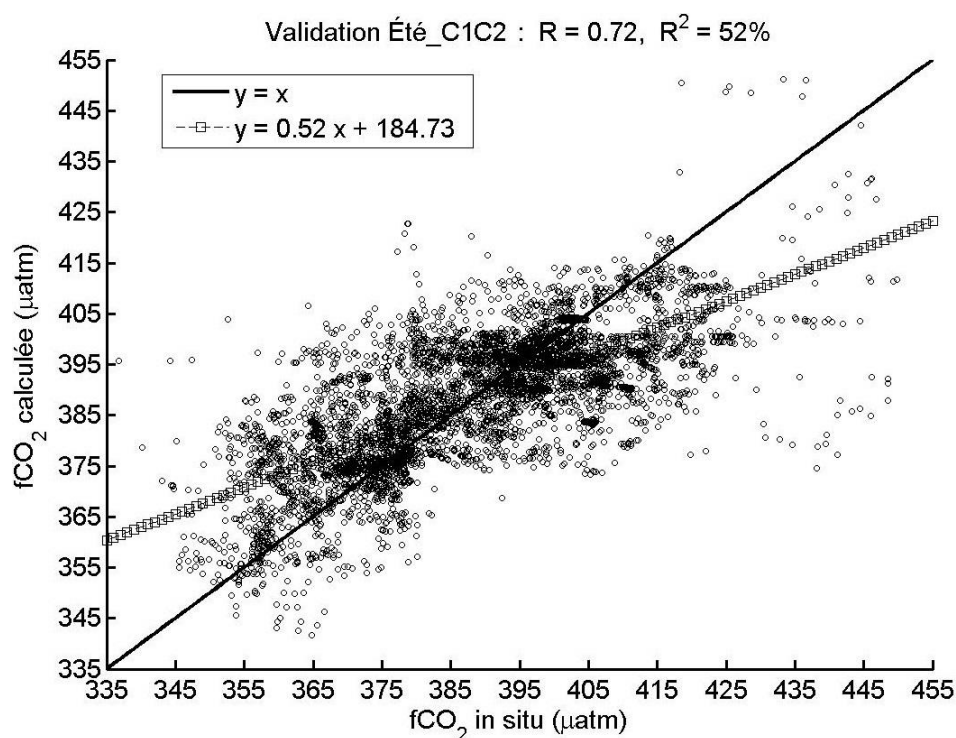


FIGURE C.3 – Validation de la classe C1C2 de la saison de l'été : fCO₂ estimée par le RN en fonction de fCO₂ *in situ*.

C.1.1.3 Test

Année	fCO_2 (μatm)		SST ($^{\circ}C$)		SSS		NP *	PD (%) **
	RMSE	Biais	RMSE	Biais	RMSE	Biais		
2003	7,14	1,12	0,54	0,07	0,18	-0,054	717	5,39
2004	0,9	0,9	0,66	-0,66	0,171	0,171	1	0,01
2005	18,61	-5,95	0,74	0,27	0,191	-0,026	2903	21,82
2006	14,09	-0,96	0,76	0,33	0,54	-0,299	4257	32
2007	14,24	-4,24	0,4	0,23	0,236	-0,081	2448	18,4
2009	14,62	3,45	0,73	0,49	0,704	-0,554	2979	22,39
2010	17,1	9,51	0,59	0,27	0,978	0,701	5675	43,32
2011	14,36	-2,62	0,5	0,27	0,403	-0,054	7424	56,68
Total	15,34	0,52	0,62	0,3	0,604	0,013	26404	0

TABLEAU C.4 – Test de la classe C1C2 de la saison de l'été : RMSE et biais (de fCO_2 , SST, et SSS) correspondants à chaque année. * nombre de points, ** pourcentage de données par rapport à la BD de test

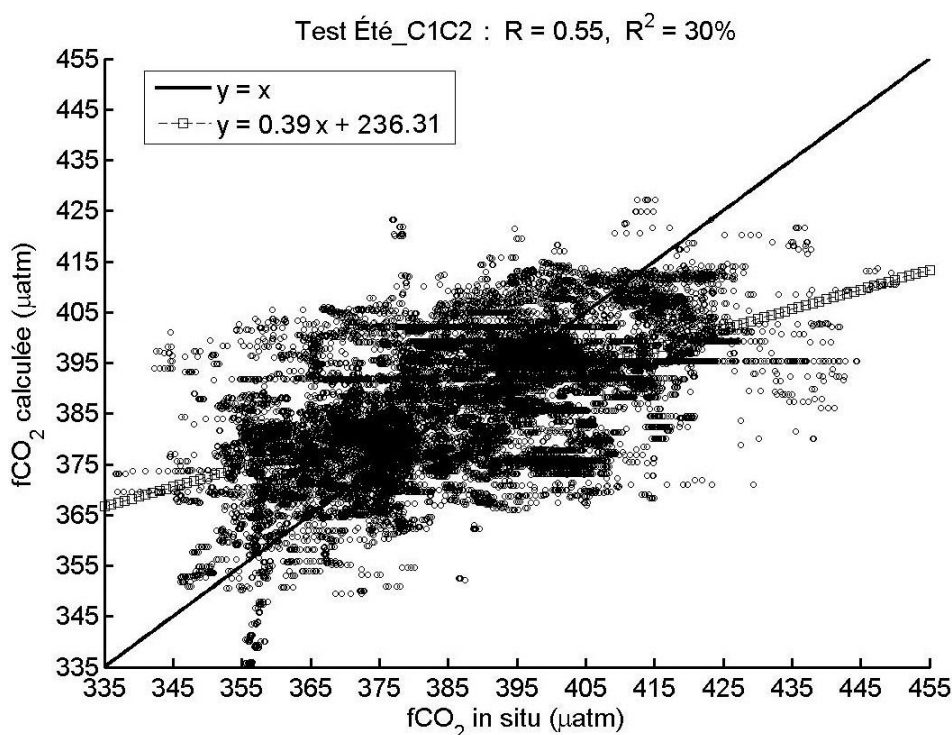


FIGURE C.4 – Test de la classe C1C2 de la saison de l'été : fCO_2 calculée par le RN à partir des données satellitaire, en fonction de fCO_2 *in situ*.

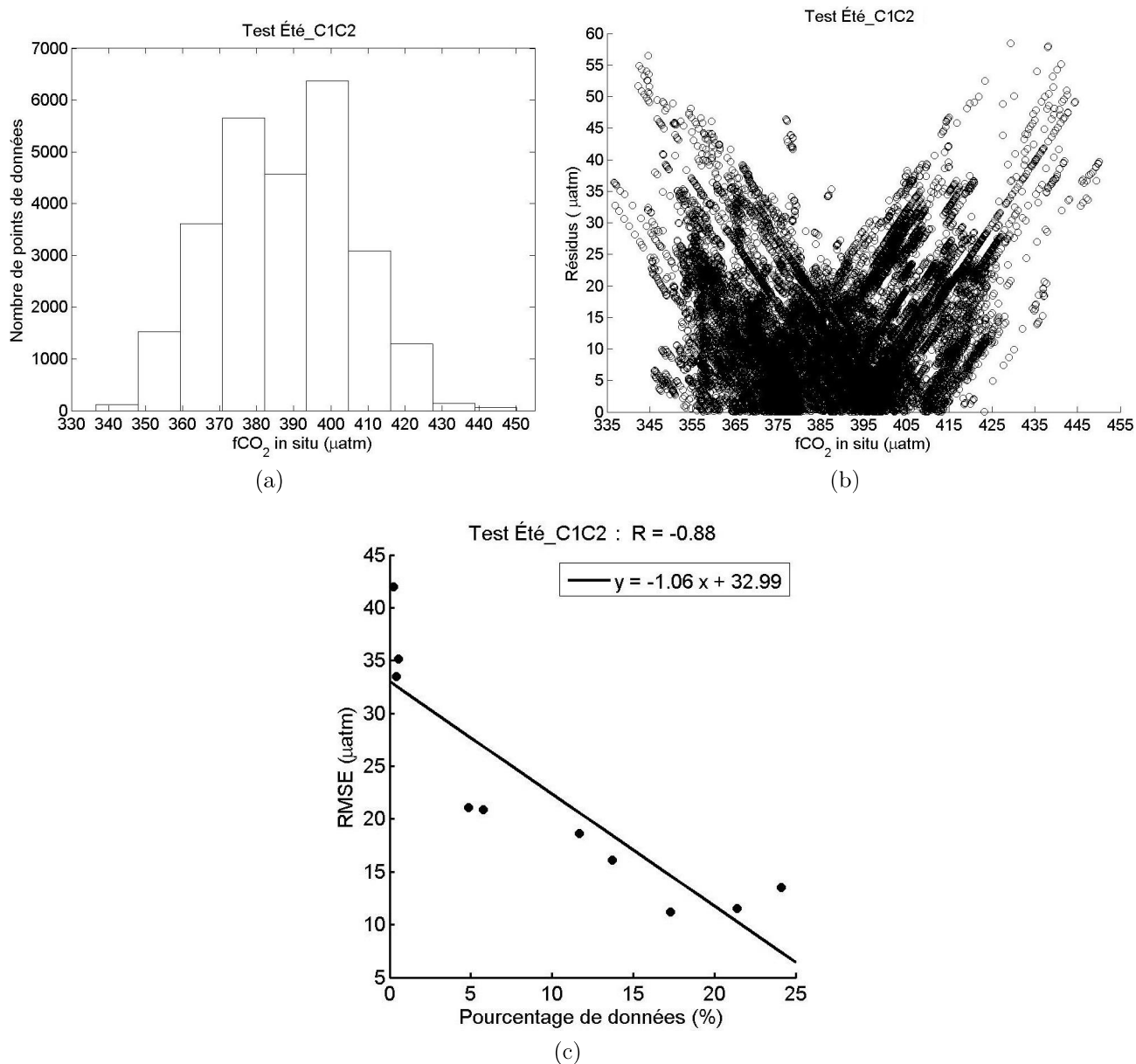


FIGURE C.5 – Test de la classe C1C2 de la saison de l'été : (a) histogramme de fCO_2 , (b) relation entre les résidus et les valeurs de fCO_2 , (c) relation entre le pourcentage de données (par rapport à la BD d'apprentissage) de chaque intervalle de l'histogramme, et les RMSE correspondants.

N°	Borne inf. (μatm)	Borne sup. (μatm)	NP *	PD (%) **	RMSE (μatm)
1	337,36	347,36	113	0,43	33,51
2	348,71	358,71	1519	5,75	20,89
3	360,05	370,05	3617	13,7	16,13
4	371,4	381,4	5653	21,41	11,57
5	382,75	392,75	4568	17,3	11,23
6	394,09	404,09	6365	24,11	13,51
7	405,44	415,44	3085	11,68	18,64
8	416,78	426,78	1285	4,87	21,09
9	428,13	438,13	139	0,53	35,17
10	439,47	449,47	60	0,23	41,95

TABLEAU C.5 – Test de la classe C1C2 de la saison de l'été : relation entre le nombre de points et le RMSE, pour chaque intervalle de l'histogramme de fCO₂. * nombre de points ** pourcentage de données par rapport au nombre total de données de la BD d'apprentissage

C.1.2 Classe 3

C.1.2.1 Apprentissage

Année	RMSE (μatm)	Biais (μatm)	NP *	PD (%) **
2002	7,76	-5,48	2505	4,09
2003	10,43	-3,26	5529	9,03
2004	5,91	1,65	4254	6,95
2005	5,7	-2,86	3282	5,36
2006	10,21	-0,55	5979	9,77
2007	8,42	-3,87	5584	9,12
2008	13,27	-7,6	1770	2,89
2009	15,11	1,29	10571	17,27
2010	14,66	2,29	11920	19,47
2011	11,27	-0,26	4624	7,55
2012	9,16	8,22	2568	4,2
2013	10,79	4,3	2626	4,29
Total	11,73	0	61212	/

TABLEAU C.6 – Apprentissage de la classe C3 de la saison de l'été : RMSE et biais correspondants à chaque année. * nombre de points, ** pourcentage de données par rapport à la BD d'apprentissage

N°	Borne inf. (μatm)	Borne sup. (μatm)	NP *	PD (%) **	RMSE (μatm)
1	320,84	332,84	206	0,34	30,35
2	334,38	346,38	578	0,94	19,7
3	347,93	359,93	3055	4,99	17,69
4	361,48	373,48	14119	23,07	10,59
5	375,02	387,02	27648	45,17	6,58
6	388,57	400,57	11219	18,33	9,63
7	402,11	414,11	2408	3,93	18,57
8	415,66	427,66	1016	1,66	28,06
9	429,21	441,21	887	1,45	43,03
10	442,75	454,75	76	0,12	46,49

TABLEAU C.7 – Apprentissage de la classe C3 de la saison de l'été : relation entre le nombre de points et le RMSE, pour chaque intervalle de l'histogramme de fCO_2 . * nombre de points ** pourcentage de données par rapport au nombre total de données de la BD d'apprentissage

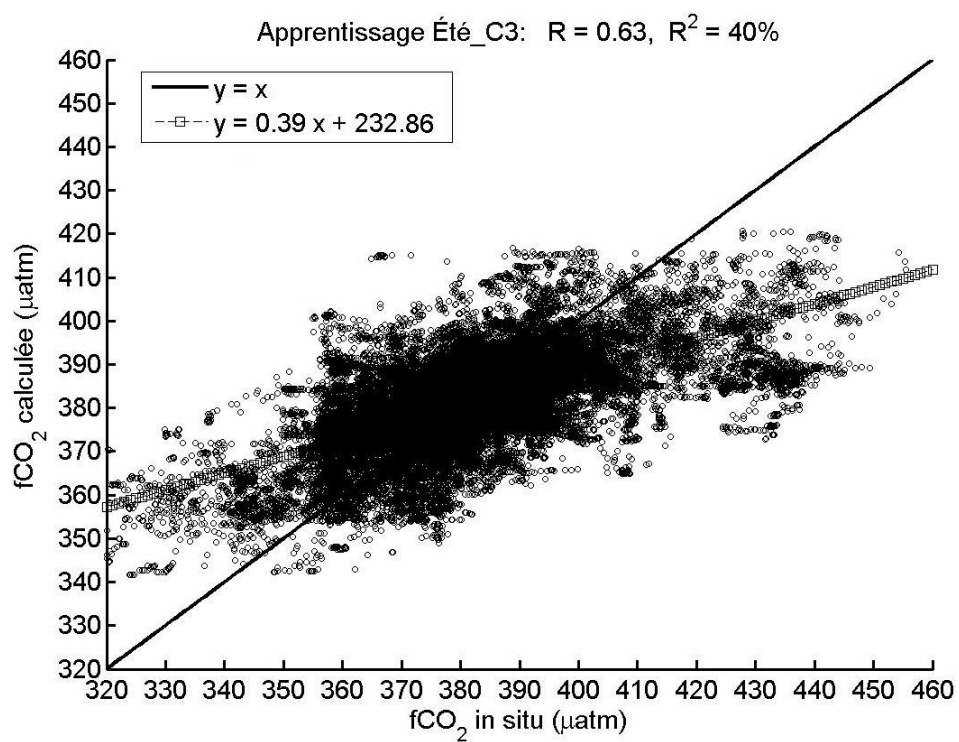


FIGURE C.6 – Apprentissage de la classe C3 de la saison de l'été : $f\text{CO}_2$ estimée par le RN en fonction de $f\text{CO}_2$ *in situ*.

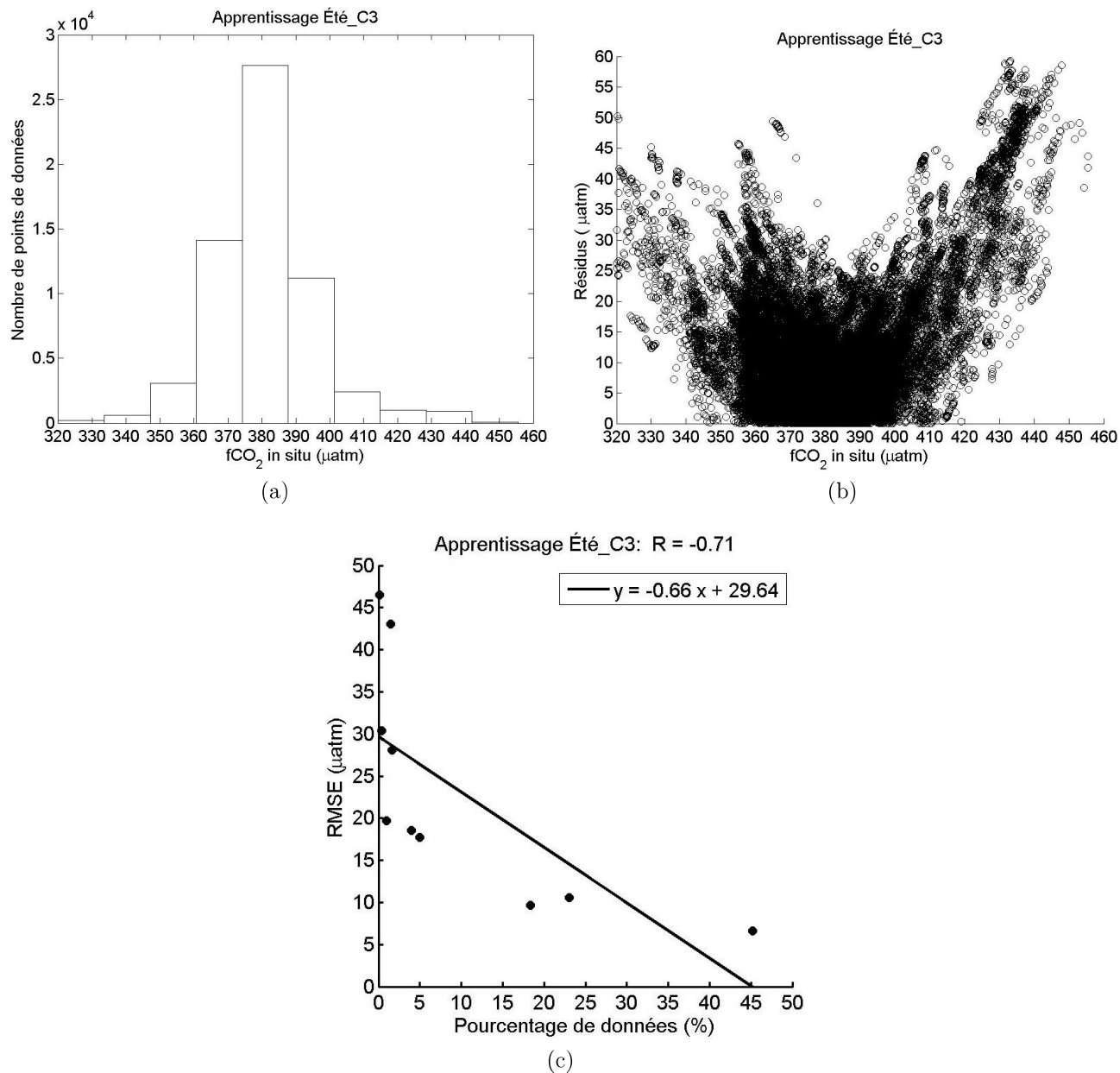


FIGURE C.7 – Apprentissage de la classe C3 de la saison de l'été : (a) *histogramme de fCO_2* , (b) *relation entre les résidus et les valeurs de fCO_2* , (c) *relation entre le pourcentage de données (par rapport à la BD d'apprentissage) de chaque intervalle de l'histogramme, et les RMSE correspondants.*

C.1.2.2 Validation

Année	RMSE (μatm)	Biais (μatm)	NP *	PD (%) **
2002	7,82	-5,6	496	4,05
2003	10,48	-3,45	1113	9,08
2004	5,94	1,65	848	6,92
2005	5,61	-2,84	656	5,35
2006	10,23	-0,48	1201	9,8
2007	8,48	-3,87	1118	9,13
2008	13,17	-7,48	355	2,9
2009	15,12	1,31	2117	17,28
2010	14,66	2,35	2385	19,47
2011	11,28	-0,21	924	7,54
2012	9,16	8,19	512	4,18
2013	10,71	4,22	526	4,29
Total	11,74	0	12251	/

TABLEAU C.8 – Validation de la classe C3 de la saison de l'été : RMSE est biais correspondant à chaque année. * nombre de points, ** pourcentage de données par rapport à la BD de validation

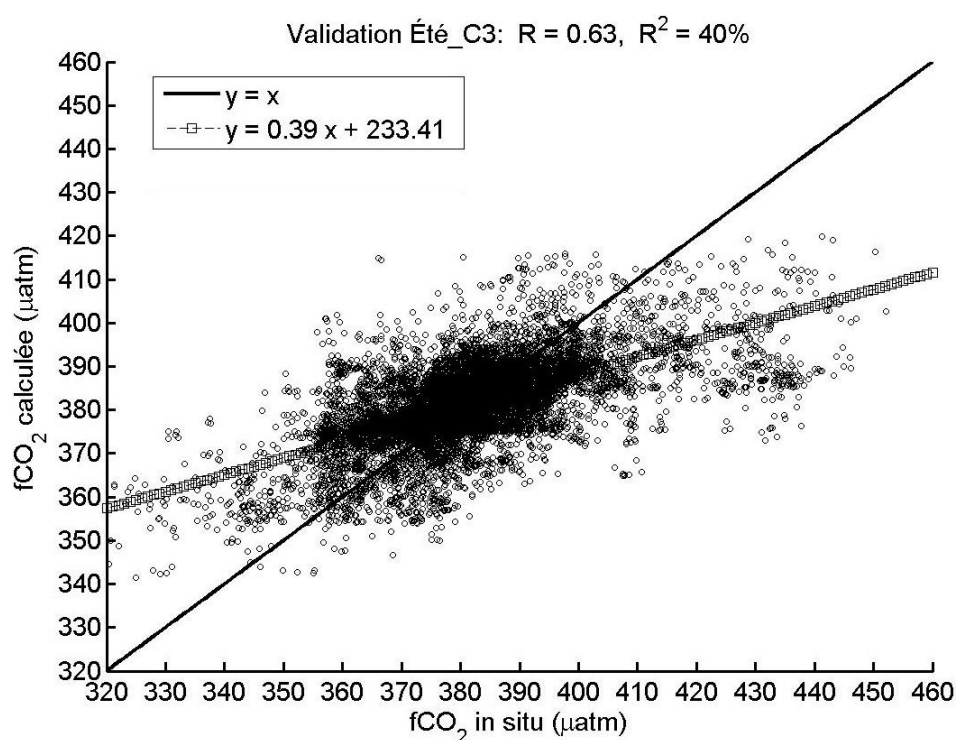


FIGURE C.8 – Validation de la classe C3 de la saison de l'été : fCO_2 estimée par le RN en fonction de fCO_2 *in situ*.

C.1.2.3 Test

Année	fCO ₂ (µatm)		SST (°C)		SSS		NP *	PD (%) **
	RMSE	Biais	RMSE	Biais	RMSE	Biais		
2002	7.22	-3,16	0,37	0,24	0,519	-0,125	1290	2,82
2003	13.84	-7,21	0,64	0,51	0,751	-0,54	6642	14,54
2004	7.65	0,46	0,5	0,44	0,55	-0,327	5102	11,17
2005	7.43	-3,08	0,42	0,29	0,631	-0,461	3938	8,62
2006	12.11	-2,01	0,48	0,26	0,388	-0,198	7180	15,72
2007	11.3	-2,79	0,58	0,38	0,621	0,016	6702	14,68
2008	14.57	-8,56	0,7	0,62	0,885	-0,494	2125	4,65
2009	17.29	-0,44	0,72	0,46	0,491	-0,142	12688	27,78
2010	17.53	4,98	0,55	0,26	0,668	0,238	13523	56,53
2011	13.47	0,05	0,69	0,46	0,353	0,102	5368	22,44
2012	18.05	16,1	0,57	0,5	0,647	0,412	2066	8,64
2013	13.84	10,09	0,64	0,59	0,6	0,27	2966	12,4
Total	14.37	0,17	0,6	0,4	0,59	-0,086	69590	/

TABLEAU C.9 – Test de la classe C3 de la saison de l'été : RMSE et biais (de fCO₂, SST, et SSS) correspondants à chaque année. * nombre de points, ** pourcentage de données par rapport à la BD de test

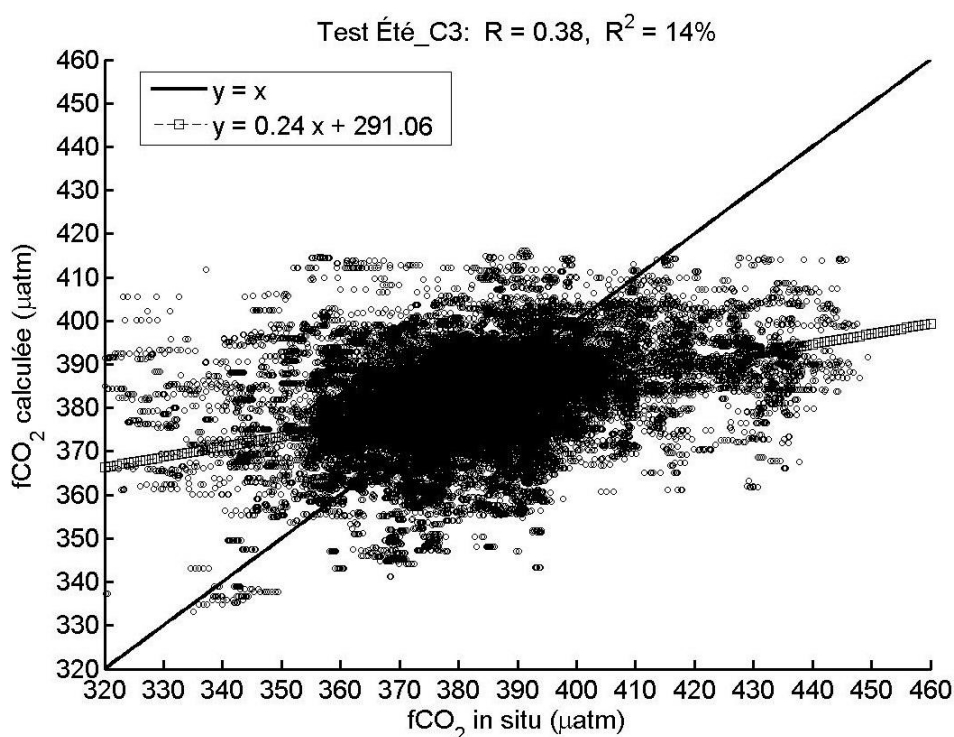


FIGURE C.9 – Test de la classe C3 de la saison de l'été : fCO₂ calculée par le RN à partir des données satellitaire, en fonction de fCO₂ *in situ*.

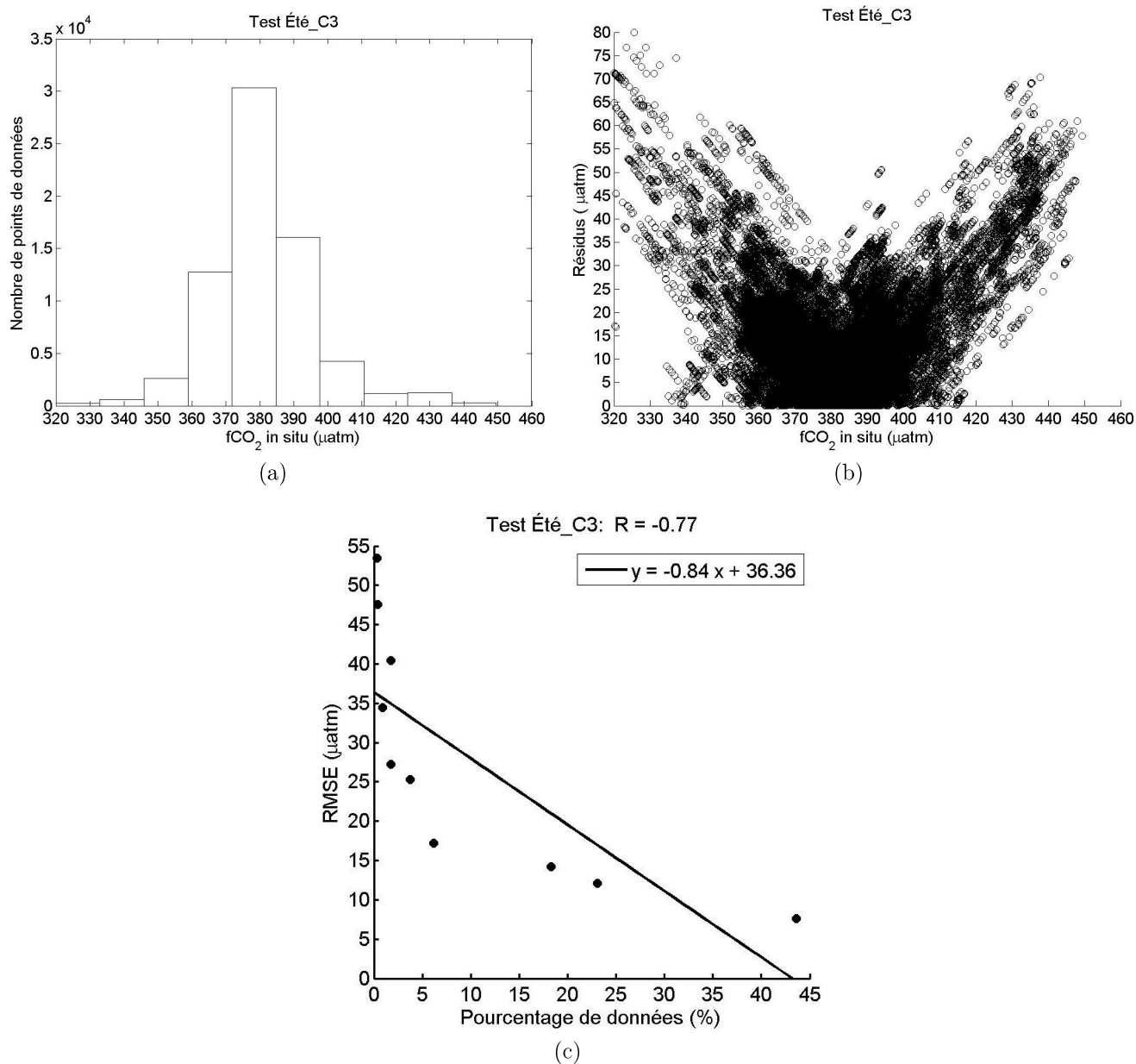


FIGURE C.10 – Test de la classe C3 de la saison de l'été : (a) *histogramme de fCO_2* , (b) *relation entre les résidus et les valeurs de fCO_2* , (c) *relation entre le pourcentage de données (par rapport à la BD d'apprentissage) de chaque intervalle de l'histogramme, et les RMSE correspondants.*

N°	Borne inf. (μatm)	Borne sup. (μatm)	NP *	PD (%) **	RMSE (μatm)
1	320,53	332,53	240	0,34	53,44
2	333,47	345,47	630	0,91	34,47
3	346,41	358,41	2608	3,75	25,29
4	359,35	371,35	12746	18,32	14,17
5	372,28	384,28	30324	43,58	7,59
6	385,22	397,22	16071	23,09	12,07
7	398,16	410,16	4273	6,14	17,18
8	411,1	423,1	1200	1,72	27,2
9	424,03	436,03	1221	1,75	40,43
10	436,97	448,97	277	0,4	47,56

TABLEAU C.10 – Test de la classe C3 de la saison de l'été : relation entre le nombre de points et le RMSE, pour chaque intervalle de l'histogramme de fCO_2 . * nombre de points ** pourcentage de données par rapport au nombre total de données de la BD d'apprentissage

C.2 Automne

C.2.1 Classe 1 et 2

C.2.1.1 Apprentissage

Année	RMSE (μatm)	Biais (μatm)	NP *	PD (%) **
2000	20,78	-15,87	853	2,98
2001	12,13	-11,39	16	0,06
2002	16,59	-10,3	925	3,23
2004	12,65	-3,98	183	0,64
2005	17,36	6,48	1742	6,09
2006	12,66	3,26	3828	13,38
2007	13,57	8,98	1511	5,28
2008	15,94	-12,37	728	2,55
2009	12,66	-1,44	5106	17,85
2010	8,54	0,47	7835	27,39
2011	17,23	-11,85	2056	7,19
2013	9,44	7,03	3818	13,35
Total	12,68	0,11	28601	/

TABLEAU C.11 – Apprentissage de la classe C1C2 de la saison de l'automne : RMSE et biais correspondants à chaque année. * nombre de points, ** pourcentage de données par rapport à la BD d'apprentissage

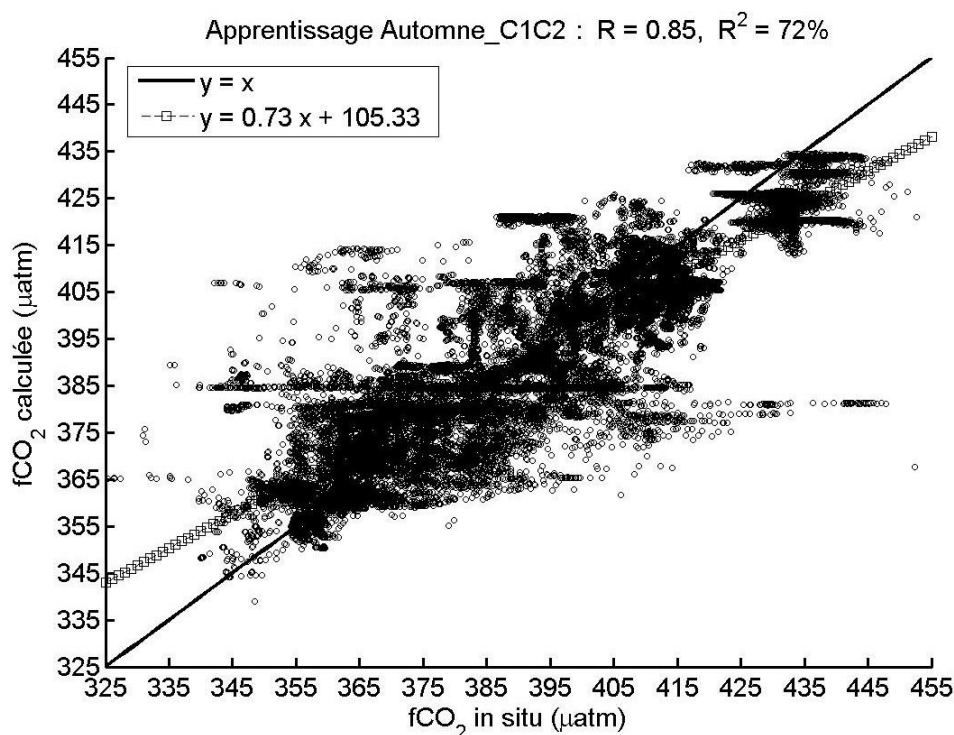


FIGURE C.11 – Apprentissage de la classe C1C2 de la saison de l’automne : fCO_2 estimée par le RN en fonction de fCO_2 *in situ*.

N°	Borne inf. (μatm)	Borne sup. (μatm)	NP *	PD (%) **	RMSE (μatm)
1	325,71	337,71	17	0,06	40,3
2	338,44	350,44	493	1,72	28,56
3	351,18	363,18	3462	12,1	12,78
4	363,92	375,92	4945	17,29	11,73
5	376,65	388,65	4004	14	11,96
6	389,39	401,39	5064	17,71	12,88
7	402,12	414,12	5154	18,02	11,36
8	414,86	426,86	2314	8,09	11,88
9	427,6	439,6	2870	10,03	10,64
10	440,33	452,33	278	0,97	27,77

TABLEAU C.12 – Apprentissage de la classe C1C2 de la saison de l’automne : relation entre le nombre de points et le RMSE, pour chaque intervalle de l’histogramme de fCO_2 . * nombre de points ** pourcentage de données par rapport au nombre total de données de la BD d’apprentissage

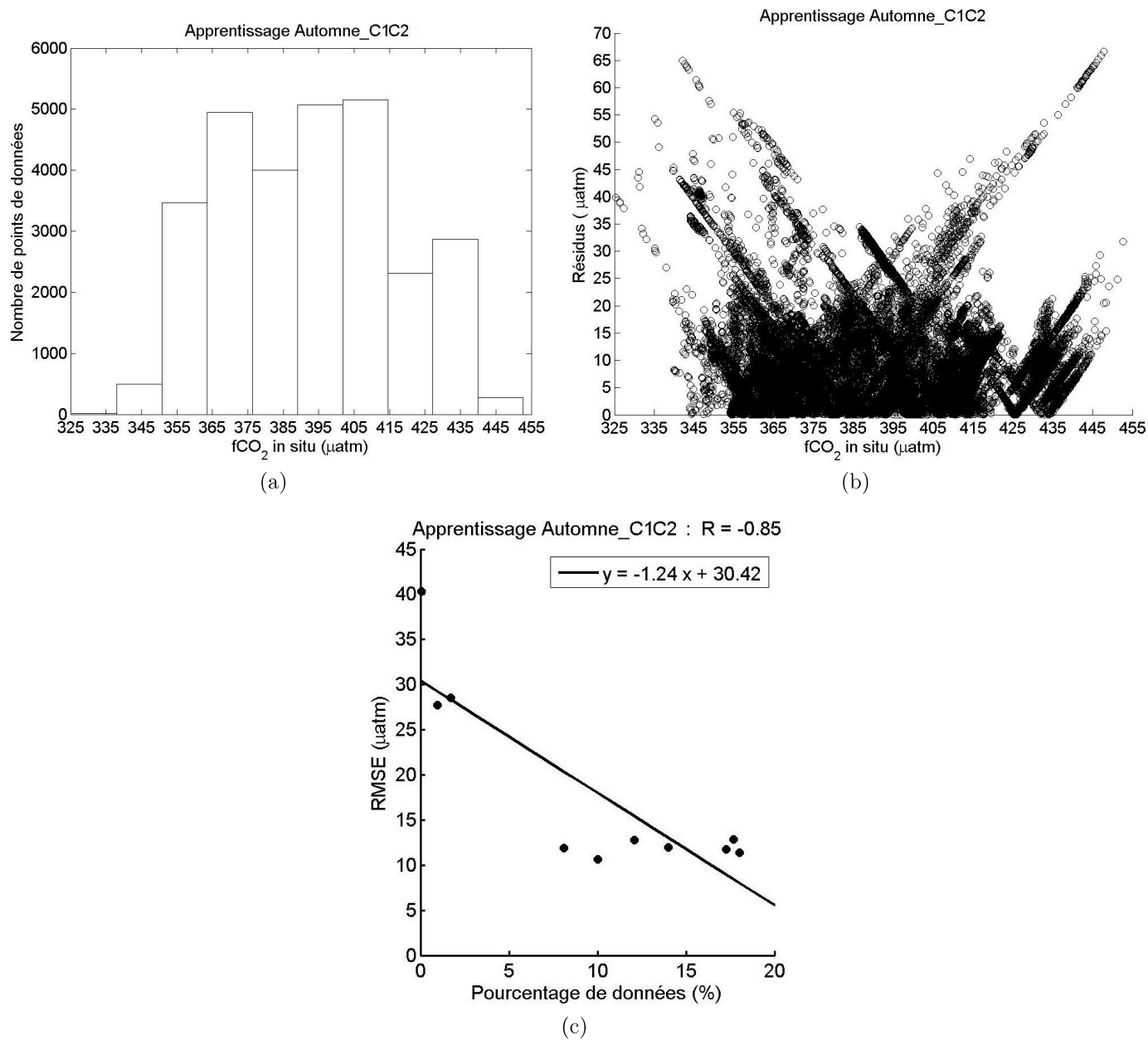


FIGURE C.12 – Apprentissage de la classe C1C2 de la saison de l'automne : (a) *histogramme de fCO_2* , (b) *relation entre les résidus et les valeurs de fCO_2* , (c) *relation entre le pourcentage de données (par rapport à la BD d'apprentissage) de chaque intervalle de l'histogramme, et les RMSE correspondants.*

C.2.1.2 Validation

Année	RMSE (μatm)	Biais (μatm)	NP *	PD (%) **
2000	20,84	-15,93	171	2,99
2001	13,01	-12,99	3	0,05
2002	16,38	-10,14	185	3,24
2004	12,54	-4,06	36	0,63
2005	17,43	6,94	346	6,06
2006	12,76	3,2	764	13,37
2007	13,24	8,95	302	5,29
2008	15,96	-12,24	146	2,56
2009	12,63	-1,44	1020	17,85
2010	8,56	0,45	1565	27,39
2011	17,19	-11,8	411	7,19
2013	9,39	6,95	765	13,39
Total	12,66	0,12	5714	/

TABLEAU C.13 – Validation de la classe C1C2 de la saison de l'automne : RMSE est biais correspondant à chaque année. * nombre de points, ** pourcentage de données par rapport à la BD de validation

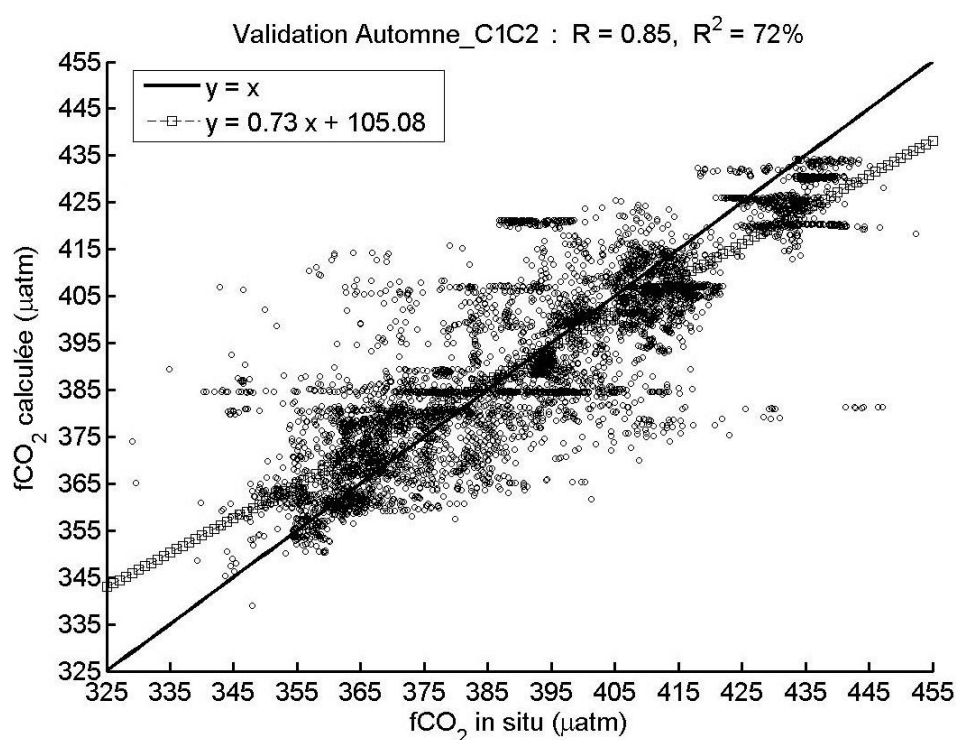


FIGURE C.13 – Validation de la classe C1C2 de la saison de l'automne : $f\text{CO}_2$ estimée par le RN en fonction de $f\text{CO}_2$ *in situ*.

C.2.1.3 Test

Année	fCO ₂ (µatm)		SST (°C)		SSS		NP *	PD (%) **
	RMSE	Biais	RMSE	Biais	RMSE	Biais		
2002	14,38	-9,1	0,77	0,57	0,476	0,138	1110	6,6
2004	10,17	-1,1	0,59	0,43	0,511	-0,417	219	1,3
2005	18,63	3,37	0,47	0,18	0,288	-0,164	2088	12,41
2006	14,16	4,46	0,69	0,45	0,514	-0,327	4592	27,3
2007	13,48	9,01	0,4	0,22	0,234	0,032	1813	10,78
2008	18,26	-14,19	0,81	0,72	0,143	-0,042	874	5,2
2009	13,62	1,74	0,28	0,04	0,121	-0,034	6126	36,42
2010	14,45	4,13	0,37	0,1	0,692	0,074	9297	57,01
2011	18,4	-10,26	0,57	-0,1	1,064	-0,027	2429	14,89
Total	11,61	7,8	0,37	0,26	0,174	-0,003	4583	28,1
0	14,57	2,44	0,47	0,19	0,528	-0,041	33131	/

TABLEAU C.14 – Test de la classe C1C2 de la saison de l'automne : RMSE et biais (de fCO₂, SST, et SSS) correspondants à chaque année. * nombre de points, ** pourcentage de données par rapport à la BD de test

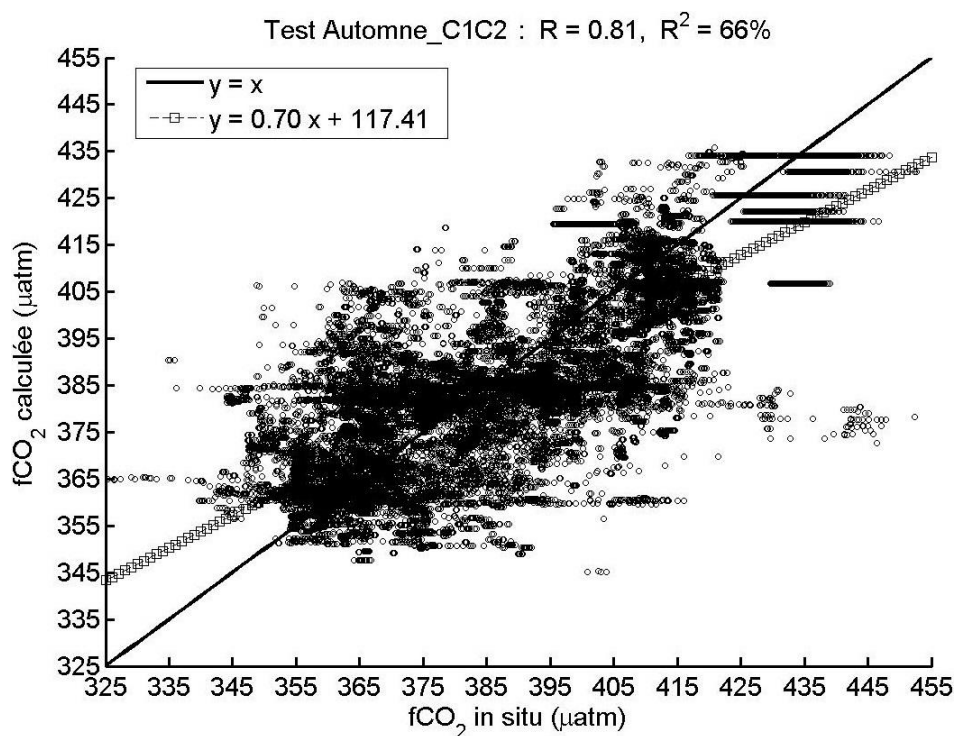


FIGURE C.14 – Test de la classe C1C2 de la saison de l'automne : fCO₂ calculée par le RN à partir des données satellitaire, en fonction de fCO₂ *in situ*.

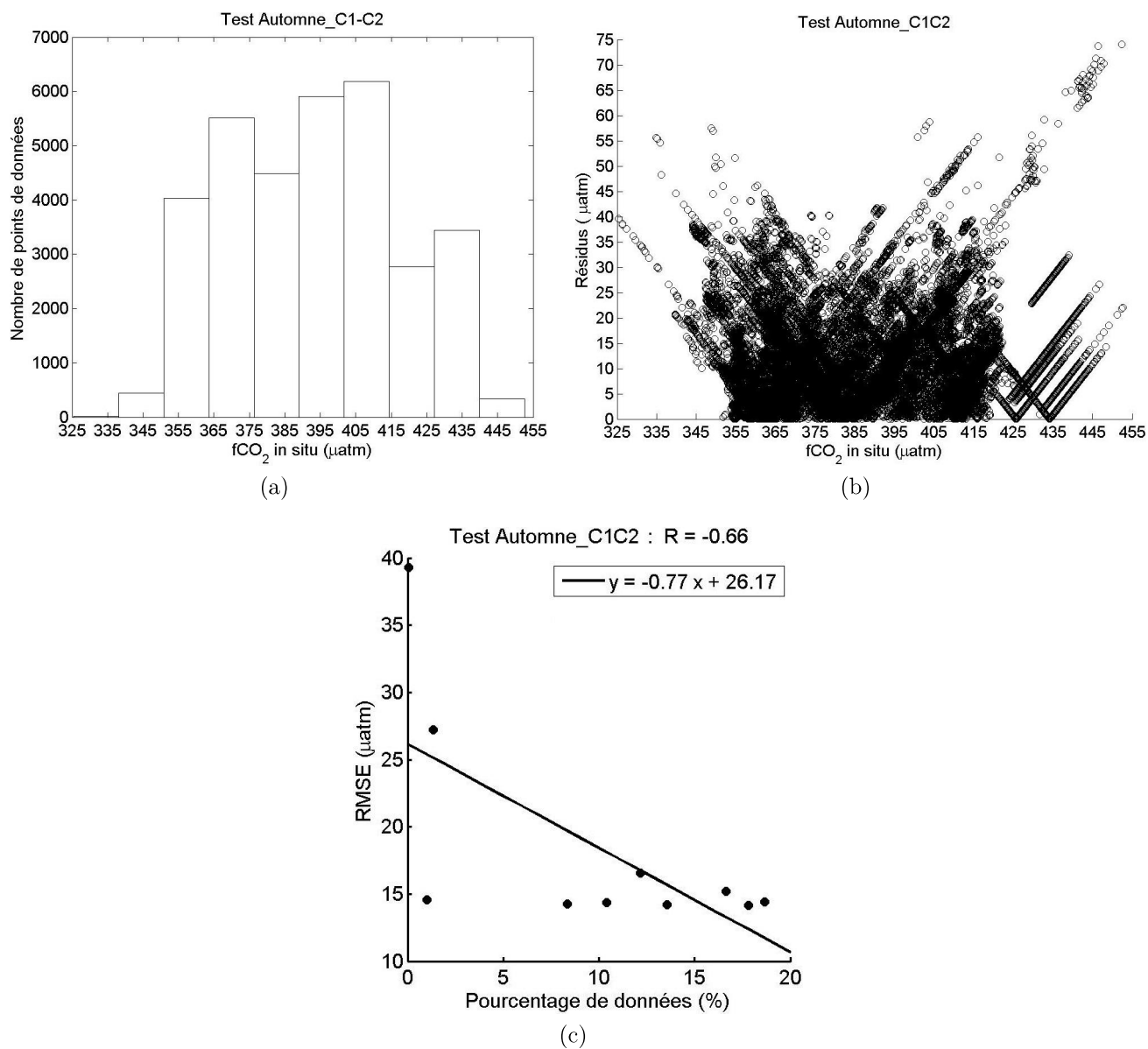


FIGURE C.15 – Test de la classe C1C2 de la saison de l'automne : (a) histogramme de fCO_2 , (b) relation entre les résidus et les valeurs de fCO_2 , (c) relation entre le pourcentage de données (par rapport à la BD d'apprentissage) de chaque intervalle de l'histogramme, et les RMSE correspondants.

N°	Borne inf. (μatm)	Borne sup. (μatm)	NP *	PD (%) **	RMSE (μatm)
1	325,71	337,71	20	0,06	39,32
2	338,44	350,44	446	1,35	27,21
3	351,18	363,18	4025	12,15	16,54
4	363,92	375,92	5516	16,65	15,18
5	376,65	388,65	4487	13,54	14,21
6	389,39	401,39	5907	17,83	14,15
7	402,12	414,12	6188	18,68	14,4
8	414,86	426,86	2766	8,35	14,26
9	427,6	439,6	3444	10,4	14,39
10	440,33	452,33	332	1	14,57

TABLEAU C.15 – Test de la classe C1C2 de la saison de l’automne : relation entre le nombre de points et le RMSE, pour chaque intervalle de l’histogramme de fCO₂. * nombre de points ** pourcentage de données par rapport au nombre total de données de la BD d’apprentissage

C.2.2 Classe 3**C.2.2.1 Apprentissage**

Année	RMSE (μatm)	Biais (μatm)	NP *	PD (%) **
2000	21,55	-16,46	794	0,98
2001	10,04	-4,23	1378	1,7
2002	16,54	-10,41	23919	29,46
2003	8,34	5,43	3452	4,25
2004	14,12	3,59	4421	5,44
2005	7,97	-0,98	4794	5,9
2006	12,51	4,54	3628	4,47
2007	13,61	4,89	2919	3,59
2008	15,41	-12,86	2235	2,75
2009	21,06	10,79	4118	5,07
2010	13,12	4,32	15508	19,1
2011	15,68	5,75	6209	7,65
2012	12,63	5,19	2797	3,44
2013	16,75	14,6	5028	6,19
Total	14,89	-0,01	81200	/

TABLEAU C.16 – Apprentissage de la classe C3 de la saison de l’automne : RMSE et biais correspondants à chaque année. * nombre de points, ** pourcentage de données par rapport à la BD d’apprentissage

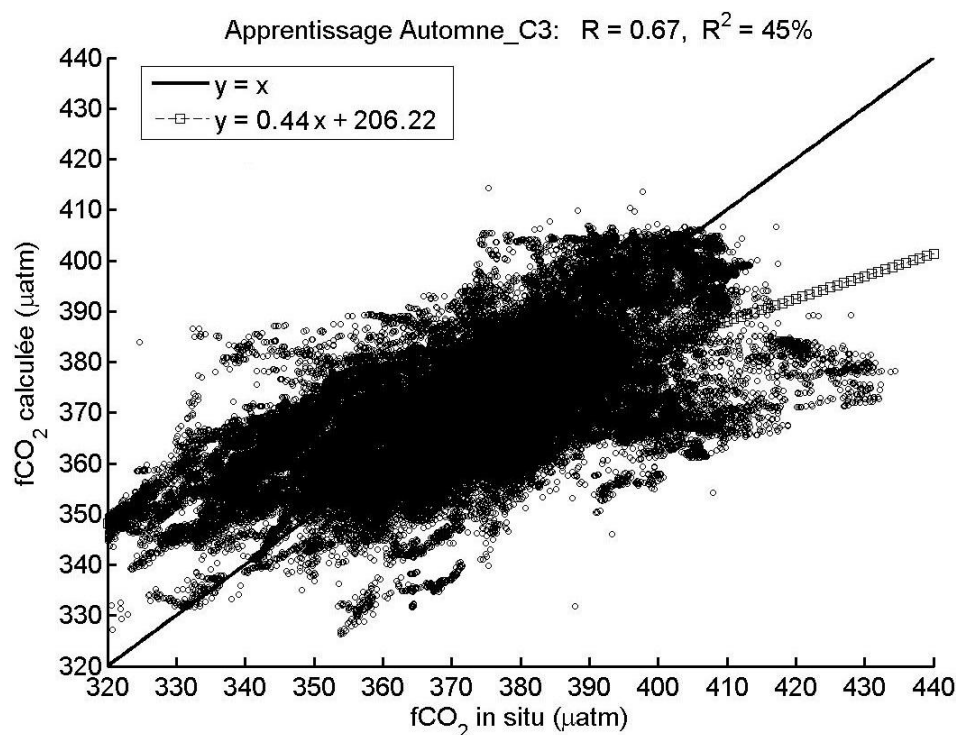


FIGURE C.16 – Apprentissage de la classe C3 de la saison de l’automne : fCO_2 estimée par le RN en fonction de fCO_2 *in situ*.

N°	Borne inf. (µatm)	Borne sup. (µatm)	NP *	PD (%) **	RMSE (µatm)
1	320,72	330,72	1800	2,22	24,81
2	332,16	342,16	5501	6,77	19,91
3	343,6	353,6	11655	14,35	16,47
4	355,04	365,04	13504	16,63	10,52
5	366,48	376,48	17788	21,91	9,47
6	377,92	387,92	15913	19,6	10,92
7	389,37	399,37	8802	10,84	17,68
8	400,81	410,81	5402	6,65	21,64
9	412,25	422,25	599	0,74	37,3
10	423,69	433,69	236	0,29	50,51

TABLEAU C.17 – Apprentissage de la classe C3 de la saison de l’automne : relation entre le nombre de points et le RMSE, pour chaque intervalle de l’histogramme de fCO_2 . * nombre de points ** pourcentage de données par rapport au nombre total de données de la BD d’apprentissage

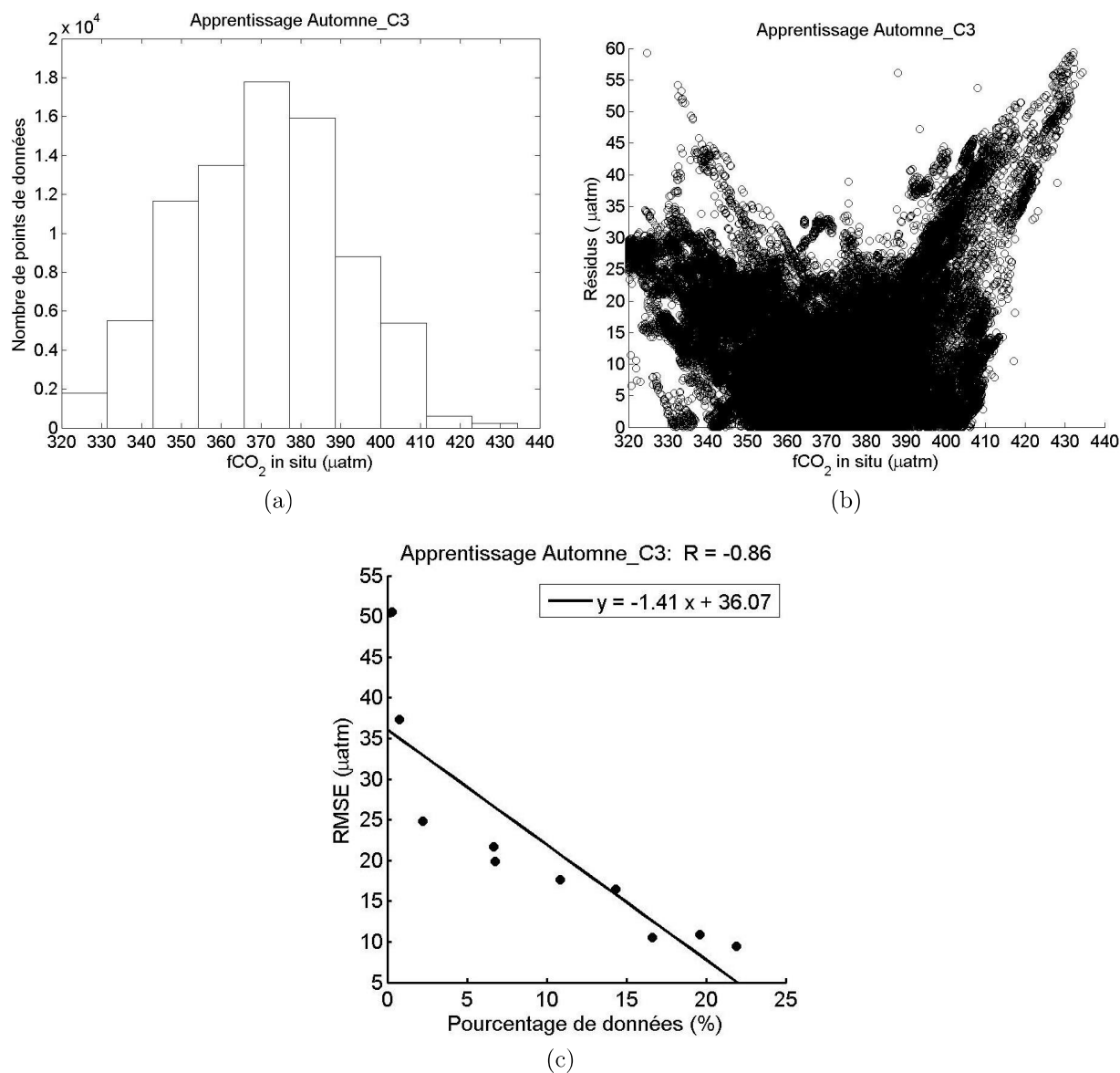


FIGURE C.17 – Apprentissage de la classe C3 de la saison de l'automne : (a) histogramme de fCO_2 , (b) relation entre les résidus et les valeurs de fCO_2 , (c) relation entre le pourcentage de données (par rapport à la BD d'apprentissage) de chaque intervalle de l'histogramme, et les RMSE correspondants.

C.2.2.2 Validation

Année	RMSE (μatm)	Biais (μatm)	NP *	PD (%) **
2000	21,82	-16,76	200	0,99
2001	9,98	-4,12	345	1,7
2002	16,51	-10,39	5982	29,49
2003	8,33	5,41	862	4,25
2004	14,06	3,68	1106	5,45
2005	8,09	-0,98	1189	5,86
2006	12,54	4,57	904	4,46
2007	13,78	5,06	732	3,61
2008	15,34	-12,94	557	2,75
2009	21	10,75	1030	5,08
2010	13,16	4,33	3883	19,14
2011	15,6	5,65	1548	7,63
2012	12,54	5,1	699	3,45
2013	16,77	14,61	1251	6,17
Total	14,88	-0,01	20288	/

TABLEAU C.18 – Validation de la classe C3 de la saison de l'automne : RMSE est biais correspondant à chaque année. * nombre de points, ** pourcentage de données par rapport à la BD de validation

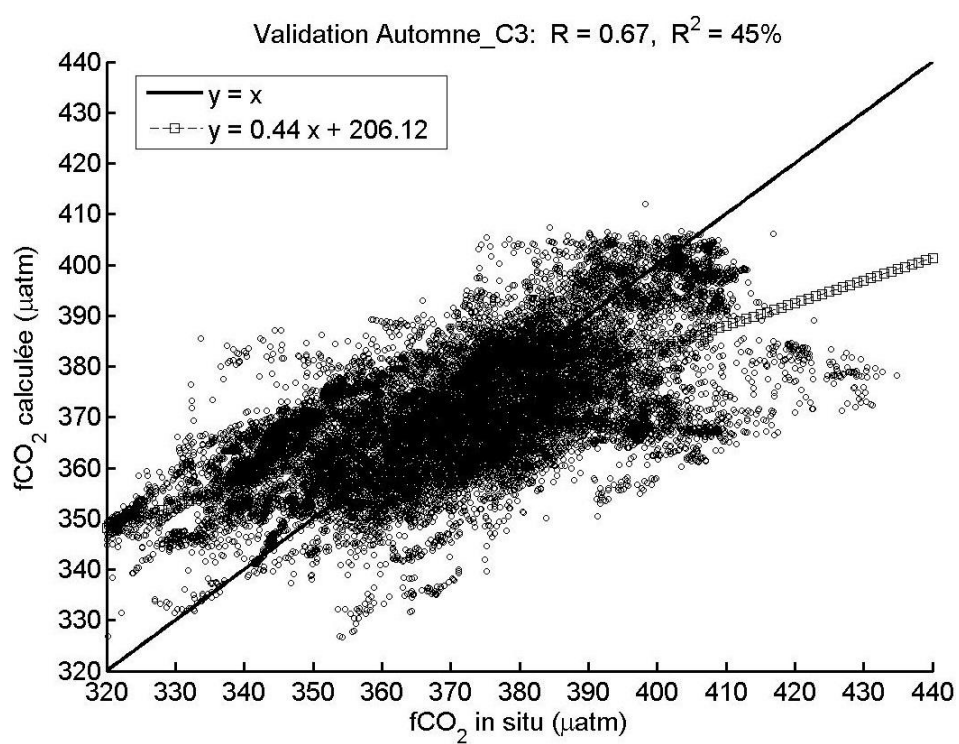


FIGURE C.18 – Validation de la classe C3 de la saison de l'automne : fCO_2 estimée par le RN en fonction de fCO_2 *in situ*.

C.2.2.3 Test

Année	fCO ₂ (µatm)		SST (°C)		SSS		NP *	PD (%) **
	RMSE	Biais	RMSE	Biais	RMSE	Biais		
2002	18.27	-7,9	0,43	0,02	0,837	0,238	28550	47,3
2003	9.5	0,21	0,4	0,08	0,87	-0,697	4274	7,08
2004	12.97	2,27	0,56	0,28	0,446	-0,369	5504	9,12
2005	9.17	0,78	0,51	0,2	0,353	-0,089	5983	9,91
2006	13.56	2,09	0,32	0,08	0,463	-0,296	4524	7,49
2007	16.77	4,05	0,44	-0,11	0,639	0,092	3612	5,98
2008	15.44	-7,86	0,61	0,48	0,545	-0,226	2770	4,59
2009	21.59	7,63	0,47	0,31	0,475	-0,258	5148	8,53
2010	16.25	4,92	0,42	0,19	0,662	0,225	17693	56,34
2011	26.1	0,34	0,4	-0,26	0,95	0,079	5426	17,28
2012	11.7	3,33	0,37	-0,11	0,548	0,06	2994	9,53
2013	18.11	15,3	0,35	0,07	0,374	0,135	5290	16,85
Total	17.08	0,15	0,44	0,09	0,691	0,039	91768	/

TABLEAU C.19 – Test de la classe C3 de la saison de l’automne : RMSE et biais (de fCO₂, SST, et SSS) correspondants à chaque année. * nombre de points, ** pourcentage de données par rapport à la BD de test

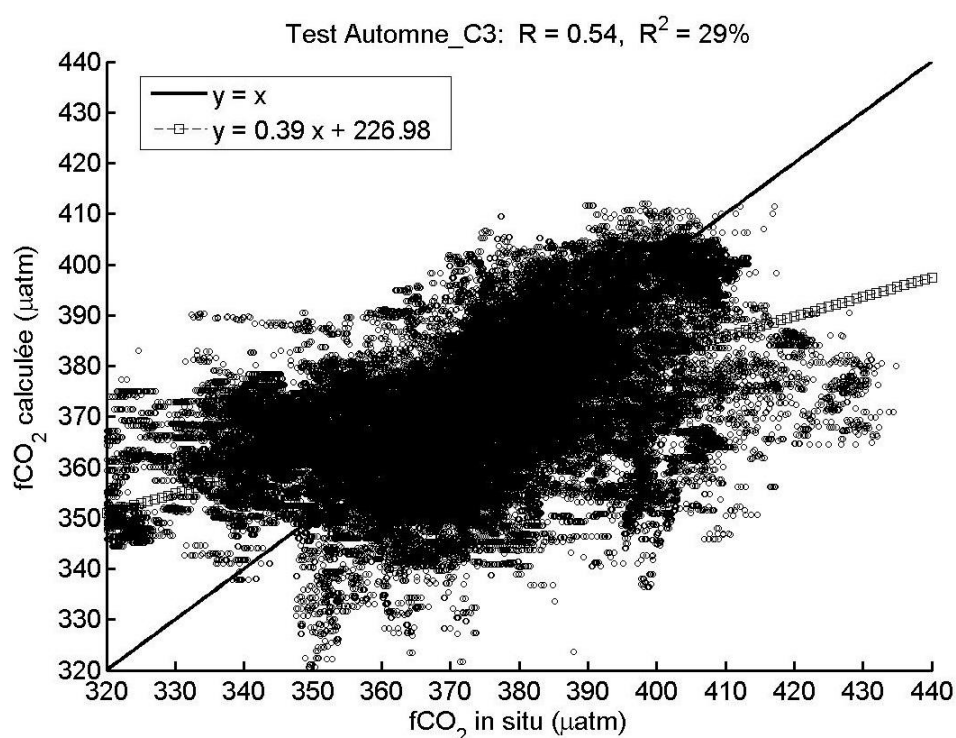


FIGURE C.19 – Test de la classe C3 de la saison de l’automne : fCO₂ calculée par le RN à partir des données satellitaire, en fonction de fCO₂ *in situ*.

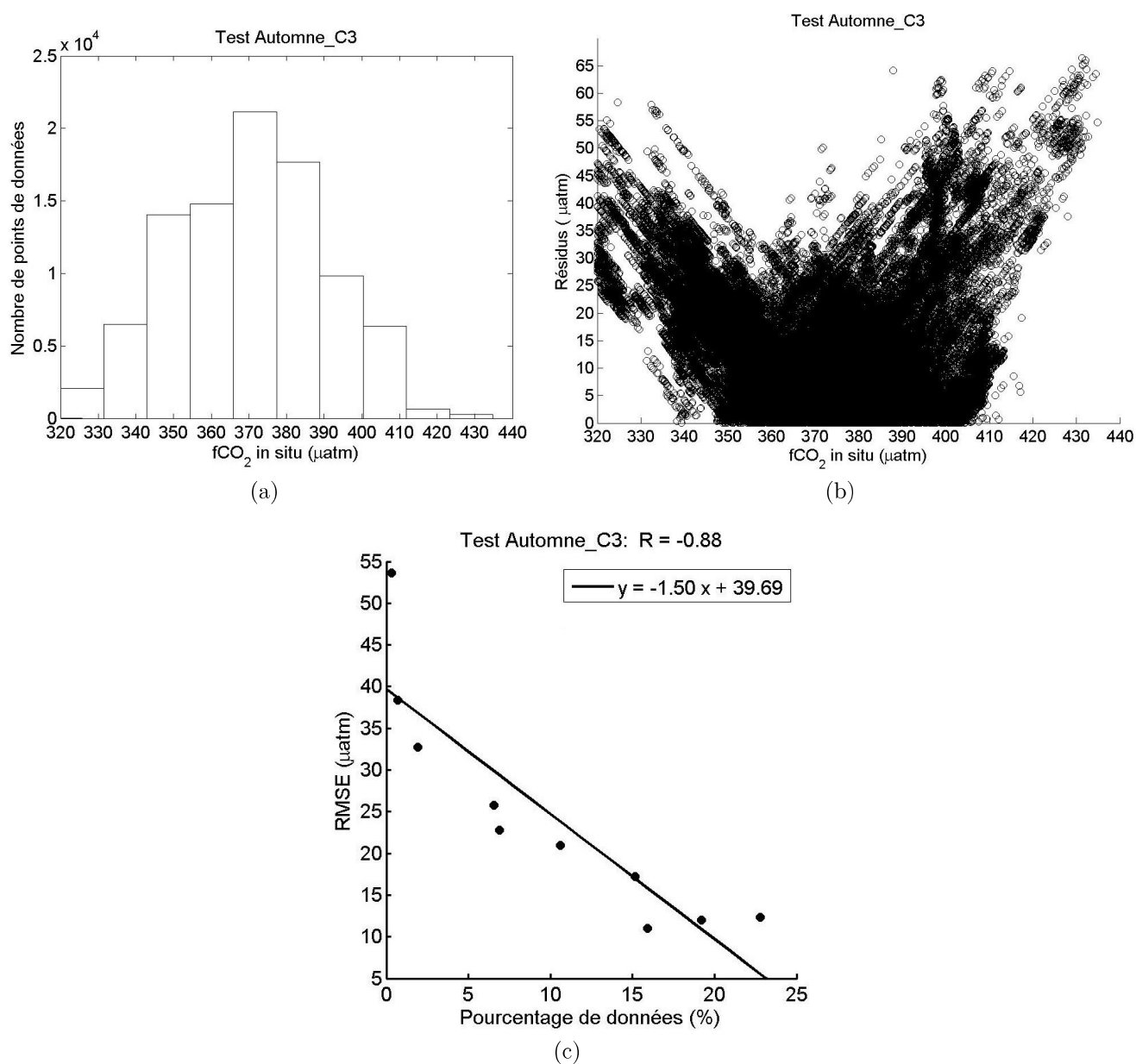


FIGURE C.20 – Test de la classe C3 de la saison de l'automne : (a) *histogramme de $f\text{CO}_2$* , (b) *relation entre les résidus et les valeurs de $f\text{CO}_2$* , (c) *relation entre le pourcentage de données (par rapport à la BD d'apprentissage) de chaque intervalle de l'histogramme, et les RMSE correspondants.*

N°	Borne inf. (μatm)	Borne sup. (μatm)	NP *	PD (%) **	RMSE (μatm)
1	320,74	330,74	1746	1,9	32,72
2	332,22	342,22	6029	6,57	25,79
3	343,71	353,71	13898	15,14	17,27
4	355,19	365,19	14615	15,93	11,03
5	366,67	376,67	20912	22,79	12,31
6	378,15	388,15	17629	19,21	12,03
7	389,63	399,63	9714	10,59	20,99
8	401,12	411,12	6310	6,88	22,8
9	412,6	422,6	644	0,7	38,34
10	424,08	434,08	271	0,3	53,64

TABLEAU C.20 – Test de la classe C3 de la saison de l’automne : relation entre le nombre de points et le RMSE, pour chaque intervalle de l’histogramme de fCO₂. * nombre de points ** pourcentage de données par rapport au nombre total de données de la BD d’apprentissage

Annexe D

TABLEAUX

D.1 Hiver

D.1.1 SST

Année	Minimum (°C)	Maximum (°C)	Moyenne (°C)	Écart-type (°C)
2002/2003	19,19	29,25	25,82	1,7
2003/2004	19,23	29,25	25,86	1,74
2004/2005	19,32	29,25	25,99	1,69
2005/2006	19,19	29,25	25,78	1,78
2006/2007	19,41	29,25	25,74	1,73
2007/2008	19,24	29,25	25,6	1,74
2008/2009	19,21	29,25	25,74	1,87
2009/2010	19,77	29,25	26,3	1,49
2010/2011	19,94	29,25	26,05	1,71
2011/2012	19,32	29,16	25,16	1,8
2012/2013	19,57	29,25	25,67	1,79
2013	19,63	29,25	25,46	1,85

TABLEAU D.1 – Paramètres statistiques de SST satellitaire pendant la saison de l’hiver entre 2002 et 2013 (de 23,5°S-70°O à 23,5°N-15°E).

D.1.2 SSS

Année	Minimum	Maximum	Moyenne	Écart-type
2002/2003	32,215	37,723	36,37	0,611
2003/2004	32,168	37,621	36,435	0,61
2004/2005	32,186	37,786	36,377	0,61
2005/2006	32,286	37,579	36,363	0,603
2006/2007	32,25	37,559	36,414	0,556
2007/2008	32,219	37,579	36,371	0,598
2008/2009	32,163	37,568	36,381	0,606
2009/2010	32,122	37,889	36,246	0,713
2010/2011	32,097	38,01	36,059	0,811
2011/2012	32,097	37,928	36,142	0,834
2012/2013	32,09	38,029	36,121	0,81
2013	32,082	37,974	36,041	0,842

TABLEAU D.2 – Paramètres statistiques de SSS satellitaire pendant la saison de l’hiver entre 2002 et 2013 (de 23,5°S-70°O à 23,5°N-15°E).

D.1.3 Chl-*a*

Année	Minimum (mg.m ⁻³)	Maximum (mg.m ⁻³)	Moyenne (mg.m ⁻³)	Écart-type (mg.m ⁻³)
2002/2003	0,002	5,35	0,11	0,19
2003/2004	0,002	5,35	0,11	0,2
2004/2005	0,002	5,35	0,11	0,21
2005/2006	0,004	5,35	0,12	0,21
2006/2007	0,002	5,35	0,11	0,2
2007/2008	0,002	5,34	0,11	0,18
2008/2009	0,002	5,35	0,11	0,21
2009/2010	0,002	5,34	0,1	0,15
2010/2011	0,002	5,34	0,1	0,13
2011/2012	0,002	5,34	0,1	0,15
2012/2013	0,002	5,35	0,1	0,13
2013	0,001	5,31	0,1	0,12

TABLEAU D.3 – Paramètres statistiques de Chl-*a* satellitaire pendant la saison de l’hiver entre 2002 et 2013 (de 23,5°S-70°O à 23,5°N-15°E).

D.1.4 fCO₂

Année	Minimum (μatm)	Maximum (μatm)	Moyenne (μatm)	Écart-type (μatm)
2002/2003	320,37	450,29	383,35	25,18
2003/2004	320,4	450,1	383,44	24,91
2004/2005	320,39	448,8	385,31	24,73
2005/2006	320,48	448,91	383,04	24,12
2006/2007	325,39	449,8	382,87	25,04
2007/2008	320,83	448,8	384,35	25,71
2008/2009	320,43	449,81	385,13	25,74
2009/2010	320,38	449,74	383,42	26,94
2010/2011	320,36	448,8	383,53	28,63
2011/2012	320,45	448,8	379,4	28,49
2012/2013	320,39	448,8	382,86	28,37
2013	320,4	448,8	378,29	27,5

TABLEAU D.4 – Paramètres statistiques de fCO₂ satellitaire pendant la saison de l’hiver entre 2002 et 2013 (de 23,5°S-70°O à 23,5°N-15°E).

D.2 Printemps

D.2.1 SST

Année	Minimum (°C)	Maximum (°C)	Moyenne (°C)	Écart-type (°C)
2003	18,41	30,74	26,21	2,18
2004	18,41	30,74	25,74	2,05
2005	18,41	30,75	26,34	2,01
2006	18,41	30,73	26,25	2,11
2007	18,41	30,71	26,05	2,13
2008	18,41	30,76	26,31	2,14
2009	18,41	30,75	26,15	2,35
2010	18,99	30,75	27,11	1,81
2011	18,49	30,76	26,42	1,96
2012	18,46	30,58	25,81	1,87
2013	18,49	30,75	26,46	1,9

TABLEAU D.5 – Paramètres statistiques de SST satellitaire pendant la saison du printemps entre 2003 et 2013 (de 23,5°S-70°O à 23,5°N-15°E).

D.2.2 SSS

Année	Minimum	Maximum	Moyenne	Écart-type
2003	32,092	37,63	36,332	0,642
2004	32,084	37,631	36,383	0,661
2005	32,036	37,625	36,367	0,646
2006	32,114	37,627	36,352	0,636
2007	32,061	37,573	36,374	0,644
2008	32,065	37,608	36,345	0,643
2009	32,305	37,607	36,292	0,664
2010	32,036	37,624	36,022	0,807
2011	32,026	37,626	36,111	0,87
2012	32,041	37,62	36,268	0,813
2013	32,028	37,626	36,115	0,894

TABLEAU D.6 – Paramètres statistiques de SSS satellitaire pendant la saison du printemps entre 2003 et 2013 (de 23,5°S-70°O à 23,5°N-15°E).

D.2.3 Chl-*a*

Année	Minimum (mg.m ⁻³)	Maximum (mg.m ⁻³)	Moyenne (mg.m ⁻³)	Écart-type (mg.m ⁻³)
2003	0,003	8,85	0,15	0,44
2004	0,003	8,85	0,17	0,5
2005	0,002	8,85	0,15	0,39
2006	0,002	8,85	0,15	0,41
2007	0,003	8,85	0,16	0,46
2008	0,004	8,85	0,14	0,4
2009	0,002	8,85	0,16	0,48
2010	0,002	8,81	0,09	0,14
2011	0,002	8,84	0,1	0,18
2012	0,003	8,83	0,1	0,23
2013	0,002	8,83	0,11	0,22

TABLEAU D.7 – Paramètres statistiques de Chl-*a* satellitaire pendant la saison du printemps entre 2003 et 2013 (de 23,5°S-70°O à 23,5°N-15°E).

D.2.4 fCO₂

Année	Minimum (μatm)	Maximum (μatm)	Moyenne (μatm)	Écart-type (μatm)
2003	329,25	444,18	384,47	19,19
2004	329,78	444,2	384,29	17,71
2005	331,15	444,28	386,22	18,54
2006	327,68	444,24	384,96	19,51
2007	327,09	444,26	383,74	18,68
2008	330,51	444,27	383,64	20,81
2009	328,78	444,27	383,23	20,55
2010	329,94	444,21	387,72	20,9
2011	328,3	444,19	383,72	19,67
2012	327,04	443,03	382,35	19,45
2013	328,38	444,26	383,59	18,77

TABLEAU D.8 – Paramètres statistiques de fCO₂ satellitaire pendant la saison du printemps entre 2003 et 2013 (de 23,5°S-70°O à 23,5°N-15°E).

D.3 Été

D.3.1 SST

Année	Minimum (°C)	Maximum (°C)	Moyenne (°C)	Écart-type (°C)
2002	17,36	30,63	25,1	2,48
2003	17,36	30,62	25,57	2,27
2004	17,36	30,57	25,34	2,4
2005	17,36	30,69	25,55	2,7
2006	17,36	30,68	25,47	2,4
2007	17,36	30,62	25,36	2,35
2008	17,36	30,68	25,65	2,39
2009	17,36	30,66	25,42	2,36
2010	17,36	30,69	26,03	2,58
2011	17,36	30,67	25,59	2,44
2012	17,36	30,64	25,31	2,36
2013	17,36	30,65	25,5	2,38

TABLEAU D.9 – Paramètres statistiques de SST satellitaire pendant la saison de l'été entre 2002 et 2013 (de 23,5°S-70°O à 23,5°N-15°E).

D.3.2 SSS

Année	Minimum	Maximum	Moyenne	Écart-type
2002	32,099	37,62	36,205	0,784
2003	32,062	37,689	36,252	0,727
2004	32,052	37,681	36,259	0,73
2005	32,018	37,64	36,177	0,767
2006	32,029	37,65	36,19	0,803
2007	32,044	37,638	36,226	0,738
2008	32,048	37,715	36,185	0,791
2009	32,022	37,702	36,155	0,737
2010	32,004	37,673	35,875	0,907
2011	32,002	37,683	35,918	0,944
2012	32,002	37,656	35,971	0,914
2013	32,005	37,69	35,997	0,872

TABLEAU D.10 – Paramètres statistiques de SSS satellitaire pendant la saison de l'été entre 2002 et 2013 (de 23,5°S-70°O à 23,5°N-15°E).

D.3.3 Chl-*a*

Année	Minimum (mg.m ⁻³)	Maximum (mg.m ⁻³)	Moyenne (mg.m ⁻³)	Écart-type (mg.m ⁻³)
2002	0,002	16,17	0,22	0,61
2003	0,003	16,18	0,24	0,71
2004	0,003	16,18	0,27	0,79
2005	0,004	16,17	0,25	0,65
2006	0,002	16,18	0,24	0,7
2007	0,003	16,17	0,24	0,69
2008	0,002	16,19	0,22	0,69
2009	0,003	16,19	0,27	0,77
2010	0,003	16,08	0,21	0,56
2011	0,003	15,73	0,16	0,34
2012	0,002	16,17	0,16	0,36
2013	0,002	16,11	0,17	0,37

TABLEAU D.11 – Paramètres statistiques de Chl-*a* satellitaire pendant la saison de l'été entre 2002 et 2013 (de 23,5°S-70°O à 23,5°N-15°E).

D.3.4 fCO₂

Année	Minimum (μatm)	Maximum (μatm)	Moyenne (μatm)	Écart-type (μatm)
2002	329,44	487,62	375,91	18,6
2003	327,86	484,09	379,35	18,59
2004	328,13	486,72	378,49	18,45
2005	330,52	491,68	379,56	18,81
2006	328,88	492,05	376,84	17,92
2007	327,31	486,45	377,25	18,3
2008	329,68	488,29	376,75	18,31
2009	329,93	481,62	378,36	19,17
2010	331,43	479,11	378,7	17,33
2011	329,28	474,4	373,89	17,32
2012	326,76	471,9	372,86	17,55
2013	328,72	473,86	375,09	17,71

TABLEAU D.12 – Paramètres statistiques de fCO₂ satellitaire pendant la saison de l'été entre 2002 et 2013 (de 23,5°S-70°O à 23,5°N-15°E).

D.4 Automne

D.4.1 SST

Année	Minimum (°C)	Maximum (°C)	Moyenne (°C)	Écart-type (°C)
2002	16,6	30,77	25,52	2,7
2003	16,6	31,03	26,07	2,73
2004	16,6	31,03	25,87	2,85
2005	16,6	30,97	25,94	2,86
2006	16,6	30,92	25,98	2,79
2007	16,6	30,84	25,7	2,73
2008	16,6	31,04	26,04	2,78
2009	16,6	30,83	25,96	2,57
2010	16,6	31,03	26,27	2,8
2011	16,6	30,91	25,74	2,66
2012	16,73	31,03	26,2	2,74
2013	16,6	30,65	25,8	2,74

TABLEAU D.13 – Paramètres statistiques de SST satellitaire pendant la saison de l'automne entre 2002 et 2013 (de 23,5°S-70°O à 23,5°N-15°E).

D.4.2 SSS

Année	Minimum	Maximum	Moyenne	Écart-type
2002	32,026	37,693	36,149	0,785
2003	32,111	37,612	36,279	0,706
2004	32,04	37,751	36,15	0,738
2005	32,041	37,588	36,119	0,78
2006	32,046	37,556	36,108	0,774
2007	32,031	37,694	36,12	0,795
2008	32,043	37,673	36,14	0,832
2009	32,046	37,723	36,098	0,752
2010	32,039	37,71	36,142	0,784
2011	32,015	37,69	35,869	0,982
2012	32,013	37,744	35,876	1,051
2013	32,013	37,728	35,927	0,959

TABLEAU D.14 – Paramètres statistiques de SSS satellitaire pendant la saison de l'automne entre 2002 et 2013 (de 23,5°S-70°O à 23,5°N-15°E).

D.4.3 Chl-*a*

Année	Minimum (mg.m ⁻³)	Maximum (mg.m ⁻³)	Moyenne (mg.m ⁻³)	Écart-type (mg.m ⁻³)
2002	0,002	5,84	0,18	0,34
2003	0,002	5,84	0,17	0,33
2004	0,003	5,84	0,17	0,31
2005	0,002	5,84	0,17	0,32
2006	0,002	5,84	0,17	0,33
2007	0,002	5,84	0,16	0,32
2008	0,002	5,84	0,16	0,34
2009	0,002	5,84	0,18	0,37
2010	0,002	5,84	0,16	0,31
2011	0,002	5,73	0,11	0,15
2012	0,001	5,39	0,11	0,15
2013	0,003	5,83	0,14	0,18

TABLEAU D.15 – Paramètres statistiques de Chl-*a* satellitaire pendant la saison de l'automne entre 2002 et 2013 (de 23,5°S-70°O à 23,5°N-15°E).

D.4.4 fCO₂

Année	Minimum (μatm)	Maximum (μatm)	Moyenne (μatm)	Écart-type (μatm)
2002	320,02	441,05	379,07	19,75
2003	320,21	440,6	379,84	15,3
2004	320,06	440,44	379,38	18,51
2005	320,28	440,69	379,57	18,23
2006	320,17	438,89	379,26	17,69
2007	320,01	442,27	380,16	19,99
2008	320,05	441,16	377,85	16,99
2009	320,65	440,24	380,43	18,86
2010	320,05	442,44	378,94	17,48
2011	320,01	446,02	376,29	20,79
2012	320	448,05	377,25	19,58
2013	320,02	441,21	377,74	20,14

TABLEAU D.16 – Paramètres statistiques de fCO₂ satellitaire pendant la saison de l'automne entre 2002 et 2013 (de 23,5°S-70°O à 23,5°N-15°E).

Annexe E

CONTRIBUTIONS SCIENTIFIQUES

- E.1 Publication 1 : A comparison of multiple non-linear regression and neural network techniques for sea surface salinity estimation in the tropical Atlantic ocean based on satellite data**

A Comparison Of Multiple Non-Linear Regression And Neural Network Techniques For Sea Surface Salinity Estimation In The Tropical Atlantic Ocean Based On Satellite Data

H. MOUSSA^{1,2} M.A. BENALLAL^{1,2} C. GOYET^{1,2} N. Lefèvre^{3,4} M.C. EL JAI^{1,2}
V. GUGLIELMI^{1,2} F. TOURATIER^{1,2}

ESAIM: PROCEEDINGS AND SURVEYS, February 2015, Vol. 49, p. 65-78
CMACS 2014 - Colloque de Modélisation, Analyse et Contrôle des Systèmes
(final draft post-refereeing)

DOI:<http://dx.doi.org/10.1051/proc/201549006>

¹IMAGES ESPACE-DEV, Université de Perpignan Via Domitia, 52 Avenue Paul Alduy, 66860 Perpignan Cedex 9, France.

²ESPACE-DEV, UG UA UR UM IRD, Maison de la télédétection, 500 Rue Jean-François Breton, 34093 Montpellier Cedex 5, France.

³IRD LOCEAN, Université Pierre et Marie Curie, UMR 7159, Paris, France.

⁴Lab. Oceanografia Física Estuarina e Costeira, Universidade Federal de Pernambuco, Av. Arquitetura, Brasil.

`hadjer.moussa@univ-perp.fr`

Abstract

Using measurements of Sea Surface Salinity and Sea Surface Temperature in the Western Tropical Atlantic Ocean, from 2003 to 2007 and 2009, we compare two approaches for estimating Sea Surface Salinity : Multiple Non-linear Regression and Multi Layer Perceptron. In the first experiment, we use 18,300 *in situ* data points to establish the two models, and 503 points for testing their *extrapolation*. In the second experiment, we use 15,668 *in situ* measurements for establishing the models, and 3,232 data points to test their *interpolation*. The results show that the Multiple Non-linear Regression is an admissible solution whether it be *interpolation* or *extrapolation*. Yet, the Multi Layer Perceptron can be used only for *interpolation*.

Résumé

En utilisant des mesures de Salinité et de Température à la surface de la mer, dans l'ouest de l'océan Atlantique tropical, de 2003 à 2007 puis 2009, on compare deux approches pour la prédiction de la Salinité dans l'eau de mer de surface : la Régression Non-linéaire Multiple et le Perceptron Multi Couches. Dans la première expérience, 18 300 mesures *in situ* sont utilisées dans la construction des deux modèles et 503 points pour tester leur *extrapolation*. Dans la deuxième expérience, 15 668 mesures *in situ* sont utilisées pour établir les deux modèles et 3 232 points pour tester leur *interpolation*. Les résultats montrent que la Régression Non-linéaire Multiple peut être appliquée à la fois pour l'*extrapolation* et l'*interpolation*. Cependant, le Perceptron Multi Couches ne peut être utilisé que pour l'*interpolation*.

Published online 13 March 2015 by EDP Sciences

1 Introduction

The tropical Atlantic Ocean (23°S - 23°N) is a highly dynamic region. It is generally characterized by high carbon dioxide fugacity (fCO_2) in surface sea-water. Consequently, this region is considered as a source of CO_2 for the atmosphere (Goyet et al., 1998). In order to quantify the variation of fCO_2 in this area, we will determine the variation of the CO_2 absorption capacity of the ocean (Goyet et al., 1998). Data used in such studies, are usually collected during oceanographic cruises which are expensive in terms of time and money. Hence, here we also use another data source like satellite data, to cover much wider area than oceanographic cruises.

In order to compute fCO_2 , we need several parameters, such as SSS (Sea Surface Salinity) and SST (Sea Surface Temperature). Unlike other satellite parameters such as SST which are available for several years, the SSS data provided from $SMOS$ (Soil Moisture and Ocean Salinity) mission are available only since 2010. Yet the available *in situ* data are only from 2003 to 2007 and 2009. Thus it is necessary to determine SSS from other available parameters. Figure 1 illustrates the temporal variations of SSS and SST . Figure 2.a illustrates that there is an *inverse* relationship between these two parameters (when SST increases, SSS decreases and vice versa), so we chose to use the widely available SST to estimate SSS . However, due to the

wide range of latitudes (Figure 1), we have to consider the dynamically complex system of currents flows in the western tropical Atlantic Ocean (Lefèvre et al., 2014). Each cruise covers a limited area of this region. Figure 1 and Figure 2 show the difference of temperature and salinity between these areas. In order to consider this criterion, we include *Latitude* in the models. Figure 2.b illustrates the relation between *SSS* and *Latitude*. In both techniques the function can be written as follows :

$$SSS = f(Latitude, SST) \quad (1)$$

In this paper, we compare two mathematical tech-

niques to estimate *SSS* from *SST* measured by satellite : Multiple Non-linear Regression and Multi-Layer Perceptron. Two experiments are implemented. In both experiments, we start by establishing the models from a dataset which contain these three parameters, namely *Latitude*, *SST* and *SSS*. Thus, in the first experiment, models are tested with another dataset that is outside of the original observation area, this is an "*Extrapolation*". In the second experiment, the dataset used in the test is also different from the original dataset, but it is inside the observation area, this is an "*Interpolation*". We used MATLAB™ software for the implementations of these programs.

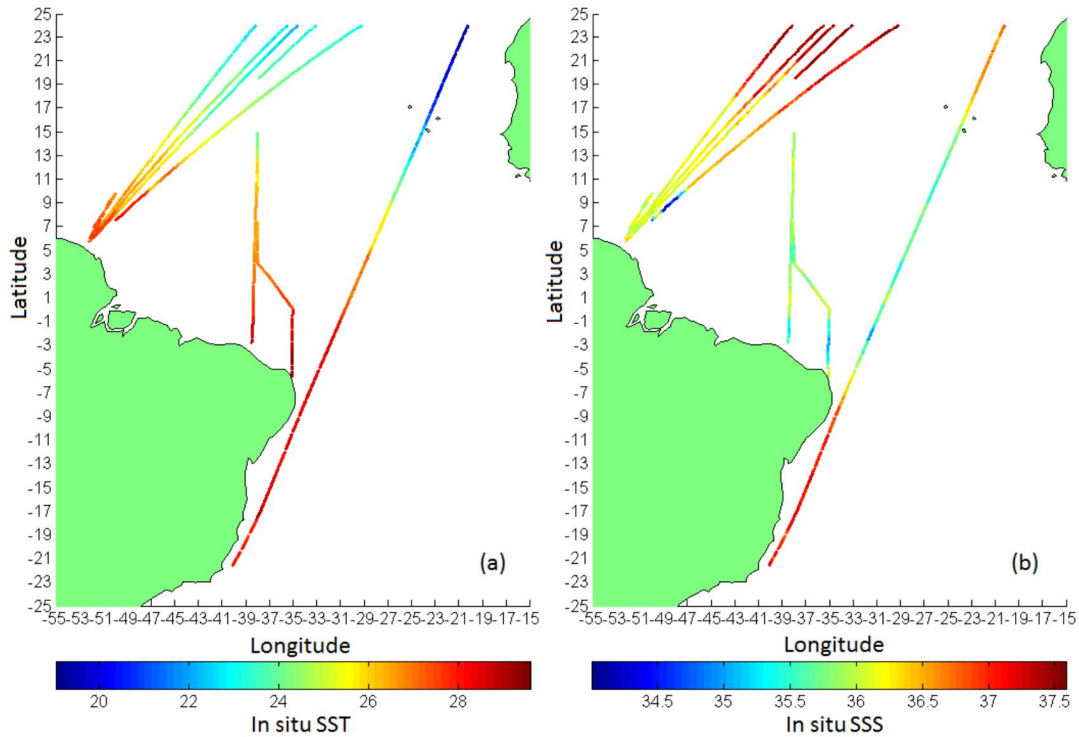


Figure 1: (a): *in situ* SST, (b): *in situ* SSS, collected during cruises : COLIBRI (western track, in 2006/2007/2009), PIRATA BRXI (central track, in 2009), Monte Olvia (eastern track, in 2009).

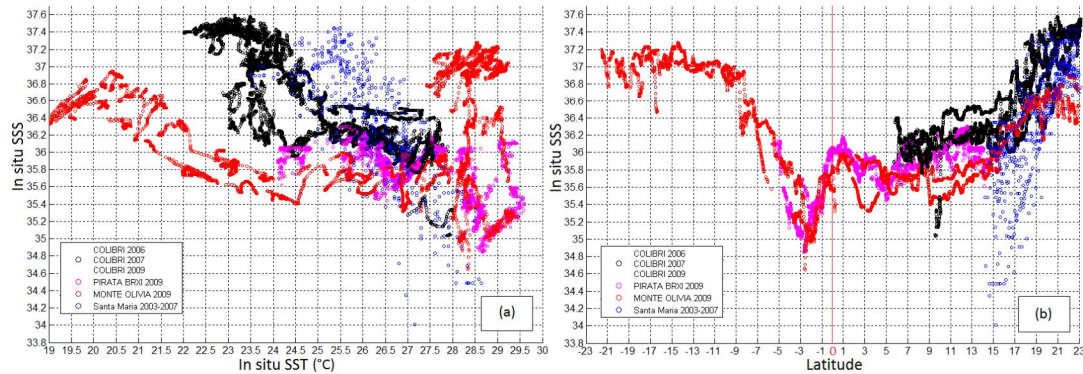


Figure 2: (a) : *in situ* SSS in terms of *in situ* SST. (b) : *in situ* SSS in terms of Latitude.

2 Datasets and methods

2.1 Datasets

This study concentrates on one period of the year from February to April. Five datasets are used. Four of them contain data collected during the following cruises:

- Santa Maria (in 2003/2004/2005/2006/2007),
- COLIBRI (in 2006/2007/2009),
- PIRATA BRXI (in 2009),
- Monte Olivia (in 2009),

which have been performed as part of the EC-funded project CARBOOCEAN IP program (<http://dataportal.carbo-ocean.org/>). These datasets

cover a large part of the Western Tropical Atlantic Ocean, with 18803 data points. The Underway fCO_2 , SST , and SSS were recorded on the research vessels. In this work, we use only SST , and SSS . The fifth dataset contains SST satellite data from MODISAqua (Moderate Resolution Imaging Spectroradiometer Aqua) mission (<http://oceandata.sci.gsfc.nasa.gov/>) at 4 Km x 4 Km resolution. This satellite was launched on May 4, 2002. It passes south to north over the equator in the afternoon, and it is viewing the entire earth's surface every 1 to 2 days. Data are produced in HDF (Hierarchical Data Format) format which is the standard data format for all NASA Earth Observing System (EOS) data products. Cruises dates, tracks, and vessel names are presented in Table 1. Figure 3 shows the cruises tracks.

Cruises	Dates	Latitude Min-Max of cruises tracks	Research vessels
Santa Maria	January - 2003	14°N - 23°N	MV Santa Maria
	Jan, Feb, Mar, April - 2004		
	Feb, Mar, April - 2005		
	Jan, Feb, April - 2006		
	Feb, Mar, April - 2007		
Colibri	March - 2006	5°N - 23°N	MN Colibri
	March, April - 2007		
	Feb, Mar, April - 2009		
Monte Olivia	March, April - 2009	23°S - 23°N	Monte Olivia
PIRATA	March, April - 2009	5°S - 15°N	N.Oc. Antares

Table 1: Datasets informations : cruises dates, latitudes, and research vessels.

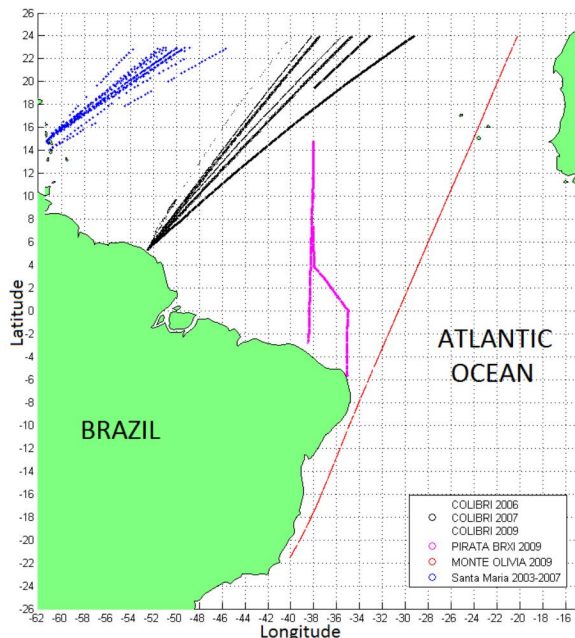


Figure 3: Cruises tracks : COLIBRI (2006/2007/2009), PIRATA BRXI (2009), Monte Olivia (2009), Santa Maria (2003/2004/2005/2006/2007).

In the first experiment, Colibri, PIRATA, and Monte Olivia data are used for building the models. MODISAqua data is used only for validation. Santa Maria data cruise is used for testing the *extrapolation*. In the second experiment, models are established using data from Santa Maria, Colibri, and Monte Olivia cruises. Validation is made using data from MODISAqua. Data from the PIRATA cruise is used for testing the *interpolation*. The origin and number of data used for each model is reported in Table 2.

2.2 MNR (Multiple Non-linear Regression) Model

A simple linear regression illustrates the relation between the dependent variable y and the independent variable x based on the regression equation (Brown, 2009):

$$y_i = \beta_0 + \beta_1 x_i, \quad i = 1, 2, \dots, n \quad (2)$$

The regression coefficients β_0 and β_1 , represent the *intercept* and the *slope*, respectively.

	Building models	Data points number	Test	Data points number
Experiment I	Colibri	18300	Santa Maria	503
	PIRATA Monte Olivia		MODISAqua	9139
Experiment II	Santa Maria	15668	PIRATA	3232
	Colibri Monte Olivia		MODISAqua	8247

Table 2: The origin and number of data used in the two experiments.

In our case, we have to find a relation between three variables SSS , SST , and $Latitude$. The dependent variable (SSS) is related to two independent variables (SST and $Latitude$). The linear model for k variables is:

$$\begin{aligned} y_i &= \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_k x_{ik}, \\ i &= 1, 2, \dots, n \end{aligned} \quad (3)$$

Figure 2 shows that there is a non-linear relationship between (SSS , $Latitude$) and (SSS , SST). Fitting a regression plane (4) will not be representative.

$$y_i = \beta_0 + \beta_1 * Lat + \beta_2 * SST \quad (4)$$

We consider a quadratic surface to fit the data sets. The regression equation is:

$$\begin{aligned} y_i &= \beta_0 + \beta_1 * Lat \\ &+ \beta_2 * SST + \beta_3 * Lat^2 \\ &+ \beta_4 * Lat * SST + \beta_5 * SST^2 \end{aligned} \quad (5)$$

The Curve Fitting Toolbox in MATLABTM provides the most widely used techniques for fitting curves and surfaces to data, using linear and non-linear regression, interpolation, and smoothing. To find the *surface* that best fits the data, we use the most popular method to estimate the regression coefficients in 5, namely the *Least Squares Method*.

2.2.1 Least Squares Method

This method is one of the oldest techniques of modern statistics. It was developed in the late 18th century and early 19th. It allows to compare experimental data, generally flawed, to a mathematical model supposed to describe this data. This method allows to minimize the experimental errors impact, by adding information in the measuring process Abdi (2010). In matrix form, the model is given by the formula:

$$y = f(X, \beta) + \epsilon \quad (6)$$

Where:

- y is an n -by-1 vector of responses.
- f is a function of β and X .
- β is a m -by-1 vector of coefficients.
- X is the n -by- m design matrix for the model.
- ϵ is an n -by-1 vector of errors.
- n is the number of data points included in fit.

The m regression coefficients are held in the vector β . To obtain this coefficient estimates, this method minimizes the summed square of residuals. The residual for the i^{th} data point r_i is defined as the deviation of the points from the regression line, in another term, it represent the difference between the observed response value y_i and the fitted response value \hat{y}_i . The MathWorks (2014a):

$$\begin{aligned} \text{Residual } r_i &= y_i - \hat{y}_i \\ y_i &\text{ observed value} \\ \hat{y}_i &\text{ fitted response} \\ S &= \sum_i^n r_i^2 = \sum_i^n (y_i - \hat{y}_i)^2 \end{aligned}$$

In Figure 4, we show some data points and a regression line, the plotted squares represent the squares of residuals. If we change the slope and/or the intercept of this line, the sizes of squares would be changed. The *Least Squares Method* find the line that minimizes the total area of these squares (Baker, 2010).

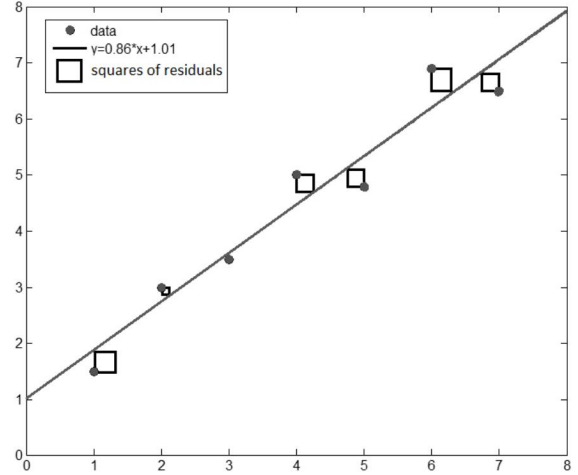


Figure 4: Representation of linear regression with the residuals squares.

We sought to compute a non-linear relationship expressing SSS as a function of $Latitude$ (Lat) and SST . We performed the multi-variable non-linear regression function which gives less errors :

$$\begin{aligned} SSS &= f(Lat, SST) = \beta_0 + \beta_1 * Lat \\ &+ \beta_2 * SST + \beta_3 * Lat^2 \\ &+ \beta_4 * Lat * SST + \beta_5 * SST^2 \end{aligned} \quad (7)$$

with

First experiment:

$$\begin{aligned}\beta_0 &= -5.855 \\ \beta_1 &= +0.5229 \\ \beta_2 &= +3.061 \\ \beta_3 &= +0.002243 \\ \beta_4 &= -0.01997 \\ \beta_5 &= -0.05623\end{aligned}$$

Second experiment:

$$\begin{aligned}\beta_0 &= -3.925 \\ \beta_1 &= +0.5998 \\ \beta_2 &= +2.828 \\ \beta_3 &= +0.001887 \\ \beta_4 &= -0.02251 \\ \beta_5 &= -0.05022\end{aligned}$$

2.3 MLP (Multi-Layer Perceptron) Model

The Artificial Neural Networks (ANNs) introduced in the 1960s, are based on the human nervous system functioning to design processing information machines (Seung, 2002). It is composed of two or more layers. Each one contains a set of neurons (e.g layers in Figure 5). The connections between the layers are associated with weights. There are many types of Neural Network (NN). They differ in structure, and the learning algorithm. In this work we proposed to apply the MLP (Multi-Layer Perceptron) that is a supervised method (Seung, 2002). It requires a desired output in order to learn using a back-propagation algorithm, for more details see Seung (2002). Model that maps the input to the output is then created. The goal is to produce an output when the desired output is unknown. Figure 5 shows the structure of MLP.

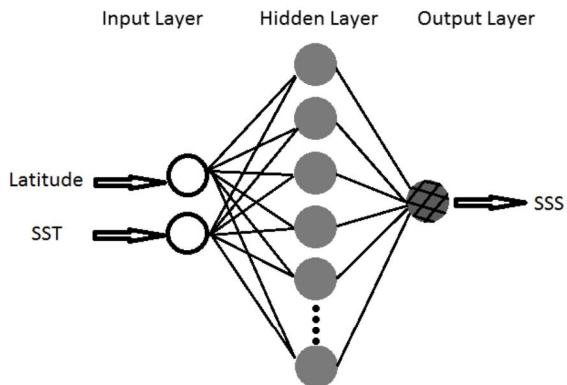


Figure 5: MLP structures obtained after data training: 2 input neurons correspond to *Latitude* and *SST*, 1 hidden layer (with 48 neurons in first experiment and 83 neurons in the second one), and 1 output neuron corresponds to estimated *SSS*.

The standard MLP network that is used for function fitting in the Neural Network Toolbox of MATLABTM, is a two-layer feedforward network, with a sigmoid transfer function (Equation (9)) in the hidden layer

and a linear transfer function (Equation (10)) in the output layer (The MathWorks, 2014b). Before learning starts, the default dataset division is made as follow:

- 70% for training,
- 15% to validate that the network is generalizing,
- 15% used as a completely independent test of network generalization.

$$net_j = \sum w_{ij} * I_i \quad (8)$$

$$f(net_j) = \frac{1}{1 - e^{-net_j}} \quad (9)$$

$$O_j = f(net_j) \quad (10)$$

I is the input vector which contains *Latitude* and *SST*. O is the output vector that represents *SSS*. w_{ij} is the weight from input i to neuron j .

In the algorithm, we propose, initially the hidden layer contained one neuron, at each new iteration a new NN is created, on which one hidden neuron is added. Finally there are as many networks as iterations. In order to find the NN that gives a best performance, we compute MSE (Equation(13)) for each one. In the first experiment, the NN with 48 hidden neurons was the best one, while it is the NN with 83 hidden neurons in the second experiment.

2.4 Models evaluation

Willmott and Matsuura (2005) have suggested that RMSE (Root Mean Square Error) is an inappropriate and misinterpreted measure of average error, and that MAE (Mean Absolute Error) is a more natural measure. (Chai and Draxler, 2014) have demonstrated that the RMSE is not ambiguous in its meaning, and is more appropriate to use than the MAE when model errors follow a normal distribution.

In model evaluation studies, there are three rules which are regularly employed to measure the average magnitude of the error.

- RMSE: As shown in Equation (11), using this quadratic scoring rule, since the errors are squared before they are averaged, the RMSE gives a relatively high weight to large errors. Thus, this rule is most useful when large errors are particularly undesirable Chai and Draxler (2014).

$$RMSE = \sqrt{\frac{\sum_i (x_i - y_i)^2}{n}} \quad (11)$$

- MAE: This is a linear score which means that all individual differences are weighted equally in the average [5].

$$MAE = \frac{\sum_i |x_i - y_i|}{n} \quad (12)$$

- MSE: We look that MSE (Mean Square Error) is analogue to RMSE Chai and Draxler (2014).

$$MSE = \frac{\sum_i (x_i - y_i)^2}{n} \quad (13)$$

- MBE: (Mean Bias Error) Here, the signs of the errors are not removed.

$$MBE = \frac{\sum_i (x_i - y_i)}{n} \quad (14)$$

Where n is the number of data, y is the output of the model and x is the target.

3 Results and discussion

To evaluate the goodness of fit test, in both experiments, for each technique, SS was calculated at the *in*

situ SST and compared with the *in situ SSS* data. In the first experiment, Figure 6.a and Figure 6.b show fitted SSS by the MNR and the MLP, respectively, as function of *in situ SSS*. Figure 7.a and Figure 7.b show results of the second experiment. We note that in all figures which represent estimated SSS as function of *in situ SSS*, two lines are represented, the theoretical line " $y = x$ ", and the regression line " $y = a.x + b$ " that represents the relationship between this two parameters. The MBE, the RMSE, and the MAE of models are reported in Table 3. In building models, for both experiments, MLP gives smaller errors than MNR.

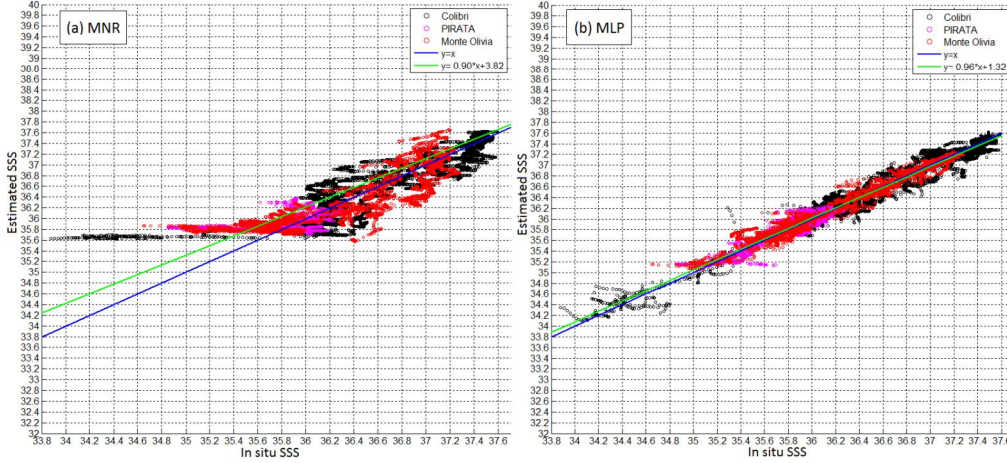


Figure 6: Experiment I (for *extrapolation*): evaluation of MNR (a) and MLP (b) using the *in situ* data; SSS estimated as function of *in situ SST*.

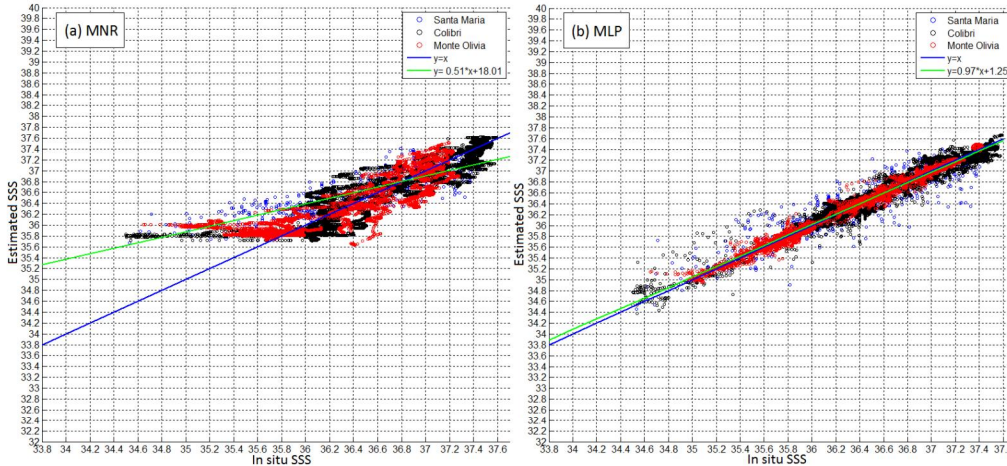


Figure 7: Experiment II (for *interpolation*): evaluation of MNR (a) and MLP (b) using the *in situ* data; SSS estimated as function of *in situ SST*.

Experiments	Models	MBE	RMSE	MAE	
I	a	MNR	4.2598×10^{-15}	0.3051	0.2306
	b	MLP	-9.3441×10^{-4}	0.1206	0.0900
II	a	MNR	-4.9171×10^{-13}	0.2844	0.2144
	b	MLP	-0.0290	0.1894	0.1177

Table 3: Goodness of fit for each model using *in situ* data : (I) first experiment, and (II) second experiment.

Experiment	MBE	RMSE	MAE
I	0.4464	0.7477	0.5271
II	0.4289	0.7533	0.5150

Table 4: Satellite SST residuals (compared to *in situ* SST) in both experiments.

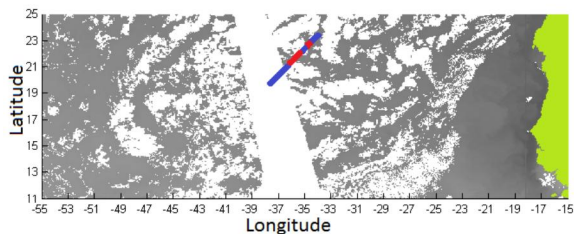


Figure 8: One day averaged Satellite SST data (gray scale); *in situ* SST data (blue); *in situ* SST data and satellite SST data (red).

In order to assess which approach would best estimate *SSS*, we used the remaining data subsets (Table 2). To validate the two models, we used daily averaged satellite data (from MODISAqua mission) at the dates of the *in situ* data used to build the models. Figure 8 shows an example of one day of satellite *SST* as well as the measured *in situ* data of *SST*. To extract data we need, we sought satellite points corresponding to *in situ* points and containing *SST* information. Then, we use points that include both *in situ SST* and satellite *SST* (Figure 8 in red). Several parameters, such as cloud cover (Figure 8 in white) prevent satellite observations, hence decrease of data point number that we can use from 18300 (*in situ* data) to 9139 (with satellite data) in the first experiment, and from 15668 to 8247 in the second one (Figure 8 in blue represents points with only *in situ SST* data). We used both polynomial functions and neural networks to estimate *SSS* from these satellite data. Estimated *SSS* is then compared with *in situ SST* used in building the models.

After this first step of validation, the models are tested with data points which are situated in a different geographical area. In the first experiment, we use data from the Santa Maria cruise. In the second experiment, we use data from the PIRATA cruise. Because of the use of a dataset which is in another geographical area,

errors have increased compared to the validation step. RMSE of satellite *SST* (Table 4), show that in datasets validation, there are some satellite points with a large error in *SST*, which leads to large errors with the MLP method.

3.1 Estimation of SSS using MNR

Results of the first step of validation are reported in Table 5.I.1.a for the first experiment, and Table 5.II.3.a for the second one. Comparing these results with those of Table 3, we note that in the two experiments, errors have improved. This is because the validation dataset is in the same geographical area than the original data, and the number of data points is smaller than that of construction model.

Table 5.I.2.a contains results of the *extrapolation*. In Figure 9.a, we see that the MNR agree relatively well on estimating salinity. Table 5.II.4.a shows that *interpolation* results are better than those of *extrapolation*. However, Figure 10.a shows a large margin between the theoretical and the regression lines. If we compare this fit with the MLP one (Figure 10.b), we see that the MLP fit is better, which is contradictory with the values of errors (Table 5.II.4.a and Table 5.II.4.b). For this reason, we computed the histogram of errors. As shown in Figure 11, in *interpolation* with MNR we have 67.98% of errors between $[0 - 0.2]$, and 32.02% between $]0.2 - 1.2]$. Unlike *interpolation* with MLP, we have 35.18% in $[0 - 0.2]$, and 64.82% in $]0.2 - 1.2]$.

We note that in all MNR regression (construction of models and tests), the model tends towards to overestimate *SSS* at the lowest *in situ SST*.

3.2 Estimation of SSS using MLP

Table 5.I.1.b and Table 5.II.3.b show results of the first validation step. Unlike MNR, MLP is sensitive to large errors. For both experiments, we note that due to the large errors in satellite *SST*, validation errors are large compared to errors in building model.

The large errors in *extrapolation* results (Table 5.I.2.b, Figure 9.b), show that in spite of small errors in building model (Table 3.II.b), and in spite of an excellent regression (Figure 6.b), it is not adapted at all for *extrapolation*. However, for an *interpolation*, it can provide reasonable results (see Table 5.II.4.b and Figure 10.b).

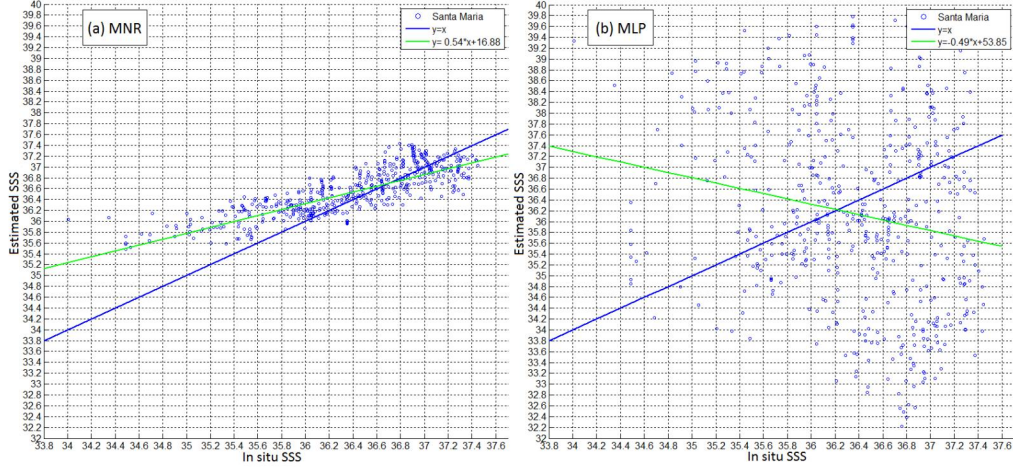


Figure 9: *Extrapolation* results with Santa Maria data cruises in experiment I; (a): MNR, (b): MLP.

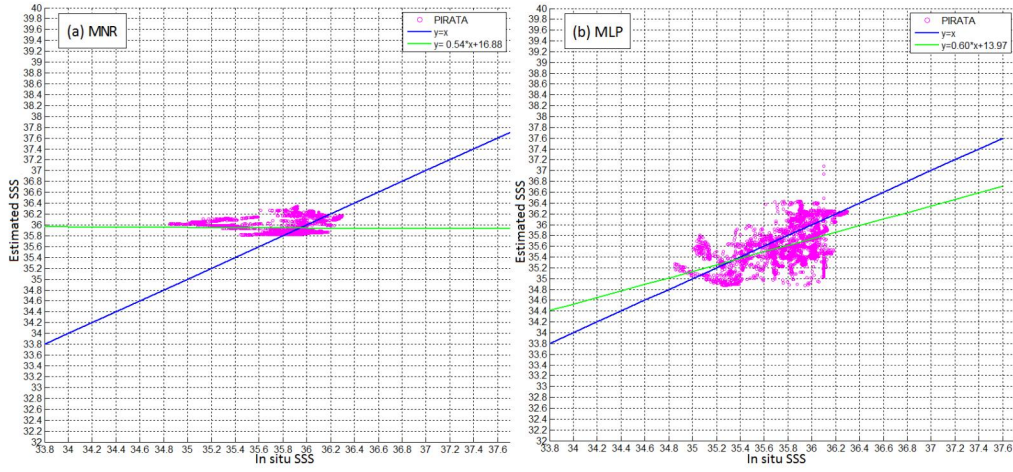


Figure 10: *Interpolation* results with PIRATA data cruise in experiment II; (a): MNR, (b): MLP.

Experiments		Datasets	Models	MBE	RMSE	MAE
I	1	Satellite data (MODISAqua)	a MNR	0.0901	0.2522	0.1965
			b MLP	-0.1785	0.5810	0.2700
	2	<i>in situ</i> data (Santa Maria cruises)	a MNR	-0.1871	0.4336	0.3316
			b MLP	0.2247	1.9608	1.5621
II	3	Satellite data (MODISAqua)	a MNR	0.0972	0.2674	0.2032
			b MLP	0.3389	2.3112	0.6521
	4	<i>in situ</i> data (PIRATA cruise)	a MNR	-0.1588	0.3614	0.2473
			b MLP	-0.1781	0.3775	0.3081

Table 5: Estimated residuals *SSS* with the four test datasets (1,2,3,4), and the two techniques (a: MNR, b: MLP).

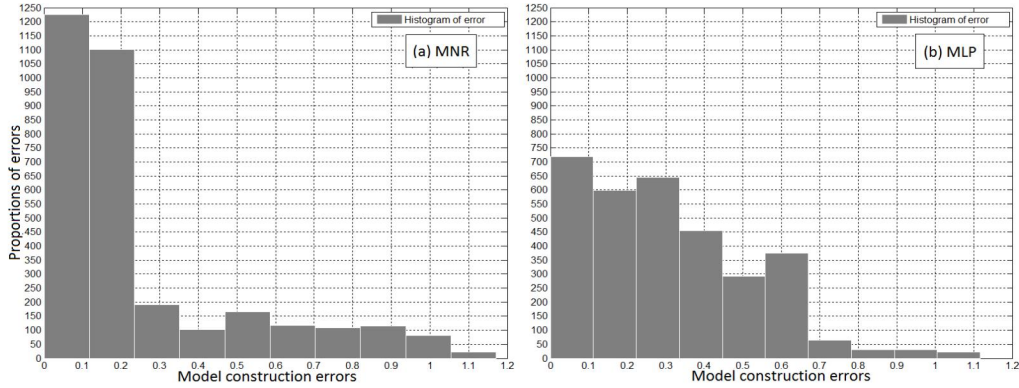


Figure 11: Error histograms in *interpolation*, (a): MNR, (b): MLP.

This work shows that the MLP method needs to be significantly improved to be used to interpolate and extrapolate *SSS* from *SST*. A neural network is more complicated than a polynomial function. Thus for obtaining a robust neural network, many parameters can be changed, especially the structure, for example change the number of hidden layers, the number of hidden neurones, connections between neurones... The training algorithm could also be changed. This study further indicates that the popular Non-linear Regression with Least Squares method provides better results than MLP in this area and for this parameters (*SST*, *SSS*). The MNR equation:

$$\begin{aligned}
 SSS &= f(Lat, SST) = \beta_0 + \beta_1 * Lat \\
 &+ \beta_2 * SST + \beta_3 * Lat^2 \\
 &+ \beta_4 * Lat * SST + \beta_5 * SST^2
 \end{aligned} \quad (15)$$

is an admissible solution whether it is for *extrapolation* with $RMSE = 0.3$, within 23°S - 61°W to 23°N - 20°W :

$$\begin{aligned}
 \beta_0 &= -5.855 \\
 \beta_1 &= +0.5229 \\
 \beta_2 &= +3.061 \\
 \beta_3 &= +0.002243 \\
 \beta_4 &= -0.01997 \\
 \beta_5 &= -0.05623
 \end{aligned}$$

Or for *interpolation* with $RMSE = 0.2$, within 23°S - 55°W to 23°N - 20°W :

$$\begin{aligned}
 \beta_0 &= -3.925 \\
 \beta_1 &= +0.5998 \\
 \beta_2 &= +2.828 \\
 \beta_3 &= +0.001887 \\
 \beta_4 &= -0.02251 \\
 \beta_5 &= -0.05022
 \end{aligned}$$

At present, additional datasets are required to better assess the model. We propose to perform a model for each month, or for each season of the year (Fall: September-October-November, Winter: December-January-February, Spring: March-April-May, Summer: June-July-August) depending upon the availability of *in situ* data. A threshold could be used

to exclude the wrong satellite *SST*. Overall, this study opens the route of mapping *SSS* at the same temporal and spatial resolution as *SST* from satellites.

Acknowledgements

We are grateful to the MESRS (Ministère Algérien de l'Enseignement Supérieur et de la Recherche Scientifique) for funding H.MOUSSA's PhD.

References

- Abdi, H. (2010). The Method of Least Squares. In *Encyclopedia of Research Design*, pages 705–708.
- Baker, S. (2010). Simple Regression Theory I.
- Brown, S. (2009). Multiple Linear Regression Analysis: A Matrix Approach with MATLAB. *ALABAMA JOURNAL OF MATHEMATICS*.
- Chai, T. and Draxler, R. R. (2014). Root mean square error (RMSE) or mean absolute error (MAE)? – Arguments against avoiding RMSE in the literature. *Geosci. Model Dev.*, 7(3):1247–1250.
- Goyet, C., Adams, R., and Eiseid, G. (1998). Observations of the CO₂ system properties in the tropical Atlantic Ocean. *Marine Chemistry*, 60(1–2):49–61.
- Lefèvre, N., Urbano, D. F., Gallois, F., and Diverrès, D. (2014). Impact of physical processes on the seasonal distribution of the fugacity of CO₂ in the western tropical Atlantic: FCO₂ IN THE WESTERN TROPICAL ATLANTIC. *Journal of Geophysical Research: Oceans*, 119(2):646–663.
- Seung, S. (2002). Multilayer perceptrons and backpropagation learning.
- The MathWorks (2014a). *Curve Fitting Toolbox User's Guide*.
- The MathWorks (2014b). *Neural Network Toolbox User's Guide*.
- Willmott, C. J. and Matsuura, K. (2005). Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Climate research*, 30(1):79.

E.2 Publication 2 : Satellite-derived CO₂ fugacity in surface seawater of the tropical Atlantic ocean using feedforward neural network

Satellite-derived CO₂ fugacity in surface seawater of the tropical Atlantic ocean using a feedforward neural network

H. MOUSSA^{1,2} M.A. BENALLAL^{1,2} C. GOYET^{1,2} N. Lefèvre^{3,4}

International Journal of Remote Sensing (final draft post-refereeing)
DOI:<http://dx.doi.org/10.1080/01431161.2015.1131872>

¹IMAGES ESPACE-DEV, Université de Perpignan Via Domitia, 52 Avenue Paul Alduy, 66860 Perpignan Cedex 9, France.

²ESPACE-DEV, UG UA UR UM IRD, Maison de la télédétection, 500 Rue Jean-François Breton, 34093 Montpellier Cedex 5, France.

³IRD LOCEAN, Université Pierre et Marie Curie, UMR 7159, Paris, France.

⁴Lab. Oceanografia Física Estuarina e Costeira, Universidade Federal de Pernambuco, Av. Arquitetura, Brasil.

`hadjer.moussa@univ-perp.fr`

Abstract

A feedforward neural network is used to quantify the fugacity of CO₂ in surface seawater ($f(\text{CO}_2^{\text{sw}})$) of the tropical Atlantic ocean, exclusively from satellite data: sea-surface temperature (SST), sea-surface salinity (SSS), and chlorophyll-*a* (chl-*a*), at a 4 km × 4 km spatial resolution, for the period of spring (March and April). The model was constructed using 7188 *in situ* data provided by the "Surface Ocean CO₂ Atlas" (SOCAT) products, and the "EC-funded project CARBOOCEAN IP program" products, available for the years: 2001, 2002, 2004, 2006, 2007, and 2009. The model was tested using remote sensing data of the Moderate Resolution Imaging Spectroradiometer (MODIS) Aqua. This approach was validated over the area extending from 8°N-61°W to 23°N-20°W. A comparison with multiple linear regression model was established. The neuron network has provided better results (root mean square error (RMSE) of 8.7 μatm (0.8815275 Pa)) than linear regression (RMSE of 9.6 μatm (0.97272 Pa)) for $f(\text{CO}_2^{\text{sw}})$ interpolation using remote sensing data. Since the required input data are available, this approach could be applied to the whole tropical Atlantic ocean and for the remaining seasons (summer, fall, and winter).

Received 18 July 2015, accepted 07 December 2015

Key words: fugacity; temperature; salinity; chlorophyll-*a*; satellite; neural network; ocean; tropical Atlantic; MODIS; SeaWiFS.

1 Introduction

Anthropogenic activities are responsible for much of the increase of the second major greenhouse gas, namely, carbon dioxide. Because of emissions of anthropogenic CO₂, the atmospheric CO₂ fugacity ($f(\text{CO}_2^{\text{atm}})$) has increased from 277 μatm (according to Bureau International Des Poids Et Mesures (1955), 1 atm = 101325 Pa) in 1750 (prior to the industrial revolution) to 395 μatm in 2013 (Dlugokencky and Tans, 2015; Tans and Keeling, 2015; Le Quéré et al., 2015). The monthly average of $f(\text{CO}_2^{\text{atm}})$ at the Mauna Loa station, reached 400 μatm in February 2015 (Tans and Keeling, 2015).

In contrast to the Pacific ocean (North of 50°S) which occupies 49% of the ocean area, and takes up only 18% of the global CO₂ uptake (Goyet and Peltzer, 1994, 1997), the Atlantic ocean (North of 50°S) represents 24% of the ocean area, and takes up over 40% of the global CO₂ (Chen, 1982; Takahashi et al., 1997, 2002, 2009). In particular, the tropical Atlantic region is characterized by strong dynamical processes, due to

the seasonal variations of surface currents. Thus, it is important to understand the total inorganic carbon distribution and its variability in this ocean area.

Although the tropical Atlantic ocean, especially the equatorial region, appears to be a source of CO₂ throughout the year (Takahashi, 1961; Goyet et al., 1998; Lefèvre et al., 1998; Landschützer et al., 2013), regional sinks of CO₂ have been found (Park and Wanninkhof, 2012; Schuster et al., 2013). Park and Wanninkhof (2012) have further shown that from 2002 to 2009, in the region 19°N-20°N, 65°W-68°W, seawater CO₂ fugacity ($f(\text{CO}_2^{\text{sw}})$), is lower than $f(\text{CO}_2^{\text{atm}})$, which lead to a large ocean CO₂ sink of $0.06 \pm 0.18 \text{ mol m}^{-2} \text{ year}^{-1}$. According to the study of Lefèvre et al. (2014), in March/April 2009, the region between 6°S-15°N, 55°W-25°W, would be a CO₂ sink of $0.40 \text{ mmol m}^{-2} \text{ day}^{-1}$ ($0.15 \text{ mol m}^{-2} \text{ year}^{-1}$).

Although the number of $f(\text{CO}_2^{\text{sw}})$ measurements has increased greatly during the last decades (since the 1990s (Sabine et al., 2010)), it is still very difficult to evaluate the air-sea CO₂ fluxes over an ocean basin and thus over the global ocean. Finding correlation

among *in situ* measurements (sea-surface temperature (SST), sea-surface salinity (SSS), chlorophyll-*a* (chl-*a*) ...) allows us to make maps of $f(\text{CO}_2^{\text{sw}})$ where *in situ* $f(\text{CO}_2^{\text{sw}})$ data are not available, and then to estimate regional air-sea CO_2 fluxes.

The objective of this study is to elaborate a model for the tropical Atlantic ocean that best represents this correlation during spring (March and April), and then use it to map $f(\text{CO}_2^{\text{sw}})$ using satellite data (SST, SSS, chl-*a*) which have a much larger spatio-temporal coverage, and with finer resolution than *in situ* data.

In order to develop an accurate model, we need to perform several performance tests, as well as, to determine why and where residuals are increasing. That is why we first choose to validate this approach over a relatively small ocean area, prior to proceed over the whole tropical Atlantic ocean. Thus, here we describe an approach applied for the limited sub-zone: from 8°N - 61°W to 23°N - 20°W .

2 Data and methods

Lefèvre et al. (2005) reported that neural network techniques predict $f(\text{CO}_2^{\text{sw}})$ in the Atlantic subpolar gyre with a better accuracy than multiple regression. In many ocean areas, the two techniques: self-organizing map (SOM) and feedforward neural network (FNN), have been widely used to interpolate $f(\text{CO}_2^{\text{sw}})$ (Lefèvre et al., 2005; Friedrich and Oschlies, 2009; Telszewski et al., 2009; Landschützer et al., 2013; Nakaoka et al., 2013; Zeng et al., 2014). SOM is commonly viewed as a flexible tool for data analysis, such as visualization and clustering of high dimensional data spaces (Flexer, 2001; Xu and Wunsch, 2005). It is also used for nonlinear regression and time series prediction tasks (Barreto and Araújo, 2002; Barreto, 2007). However, compared to the number of models developed based on FNN, this field is still in its infancy. Here, we chose to compare FNN algorithm with multiple linear regression model, to interpolate the $f(\text{CO}_2^{\text{sw}})$ spatio-temporal variations.

Sea-surface $f(\text{CO}_2^{\text{sw}})$ depends upon several parameters. Changes in solubility, which is mainly driven by temperature, leads to an important spatial and temporal variation of $f(\text{CO}_2^{\text{sw}})$. In earlier works, $f(\text{CO}_2^{\text{sw}})$ -SST relationships have been studied (Gordon and Jones, 1973; Copin-Montegut, 1988; Copin-montegut, 1989; Goyet and Poisson, 1989; Goyet et al., 1993), and used to map $f(\text{CO}_2^{\text{sw}})$ (Stephens et al., 1995; Lefèvre and Taylor, 2002; Inoue et al., 2003; Olsen et al., 2003; Lefèvre et al., 2005). The $f(\text{CO}_2^{\text{sw}})$ is also affected by marine biology that modifies the total dissolved inorganic carbon content of the water.

Jamet et al. (2007) further investigated the parameters needed to generate seasonal $f(\text{CO}_2^{\text{sw}})$ maps in the North Atlantic. They reported that interpolating $f(\text{CO}_2^{\text{sw}})$ as a function of SST, chl-*a* and mixed layer depth (MLD) provides results more accurate than those of previous works based only upon SST, or upon the three parameters: SST, latitude, and longitude. In this work, we chose to interpolate among parameters that have the highest influence on $f(\text{CO}_2^{\text{sw}})$, namely;

SST, SSS, and chl-*a*.

2.1 Data

2.1.1 *in situ* data

The spatial and temporal domain of this study extends from 8°N - 61°W to 23°N - 20°W (See figure 1), for a period from March to April. The database is a combination of: 1) Surface Ocean CO_2 Atlas (SOCAT) product, version 1.4 (Pfeil et al., 2013), and 2) The product of the EC-funded project CARBOOCEAN IP program (<http://dataportal.carboocean.org/>). This database is used to construct the models. It contains 7188 vectors with following parameters: date, time, latitude, longitude, $f(\text{CO}_2^{\text{sw}})$, SST, and SSS. The ocean depth of all data is not more than 5 m. Since *in situ* chl-*a* was not available, we merged *in situ* database with satellite chl-*a* provided from 2 sensors: MODIS Aqua (Moderate Resolution Imaging Spectroradiometer Aqua, <http://oceandata.sci.gsfc.nasa.gov/MODISA/>) and SeaWiFS (Sea-viewing Wide Field-of-view Sensor, <http://oceandata.sci.gsfc.nasa.gov/SeaWiFS/>) (Section 2.1.2). These *in situ* data are available for the years: 2001, 2002, 2004, 2006, 2007, and 2009. We note that unlike $f(\text{CO}_2^{\text{atm}})$, the surface $f(\text{CO}_2^{\text{sw}})$ is highly variable, yet, because of the sparsity of its spatial and temporal coverage, the interannual variability of $f(\text{CO}_2^{\text{sw}})$ remains poorly understood. That is why we combine data of 6 years to develop the model.

2.1.2 Satellite data

This database is used to create maps of $f(\text{CO}_2^{\text{sw}})$. It contains SST and chl-*a* data, which are generated from the MODIS Aqua sensor. Data are available from 2003 at $4\text{ km} \times 4\text{ km}$ resolution. In addition to these data, as mentioned above (Section 2.1.1), we used satellite chl-*a* generated from SeaWiFS, at $9\text{ km} \times 9\text{ km}$ resolution, for the years 2001 and 2002.

Knowing that, for daily averaged satellite data, there are a lot of data points missing (primarily due to cloud cover), in order to get a better spatio-temporal coverage, all satellite data used in this study are monthly averaged.

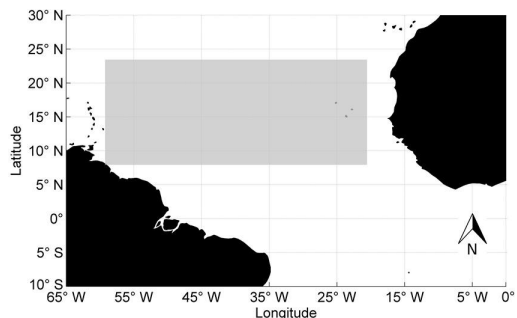


Figure 1: The study area, which is represented by the gray rectangle.

2.2 Methods

2.2.1 Feedforward neural network (FNN)

An artificial neural network (ANN) is an abstraction of biological neurons, used to solve different problems in several fields. Many researchers have worked to improve ANN computing capability (McCulloch and Pitts, 1943; Rosenblatt, 1958; Rumelhart and McClelland, 1986).

ANN consists in a set of interconnected neurons (computing units) in the form of layers, which aim to compute an output from one or more input(s) through learning. It can be used for several problems, such as linear and non-linear regressions, pattern recognition, anomaly detection... The network architecture and learning algorithm can be adapted depending on the scientific question of interest.

The number of input neurons depends on the number of estimation parameters. The precise number of hidden layers and hidden neurons needs to be set by the user, depending on the complexity of the problem. Several researchers proved that three-layer ANNs are preferred for non-linear function approximation (de Villiers and Barnard, 1993; Tamura and Tateishi, 1997). In order to define the optimal number of hidden neurons, we could repeat the ANN learning, each time increasing the number, until the best result is obtained. For a regression problem, generalization is the principal object, since too many hidden neurons may lead the ANN to overlearning resulting in increasing errors (Panchal et al., 2011), a large number of hidden neurons should be avoided. The FNN is the simplest type of ANN, in which the information circulates in one direction, from the input to the output, and there are no connections from output to input.

The learning technique we used here to train the FNN is a variant of the backpropagation algorithm (Rumelhart and McClelland, 1986) which is the most popular in FNN learning field. In order to minimize the mean squared error between the desired output and the actual output of the network, this algorithm aims to adjust connection weights, proceeding from the last to the first layer. This variant is called "levenberg-marquard backpropagation" (Levenberg, 1944; Marquardt, 1963), which is one of the fast training algorithms presenting a stable convergence.

2.2.2 Multiple linear regression (MLR)

The MLR method is used to determine the coefficients ($\beta_0, \beta_1, \beta_2, \dots, \beta_n$) that illustrate the relationship between a dependent variable (Y) and two or more explanatory variables (X_i). This leads to an equation of the form: $Y = \beta_0 + \beta_1 \times X_1 + \beta_2 \times X_2 + \dots + \beta_n \times X_n$, n represents the number of observations.

To do this, we used the most popular method, namely, least squares method, which aims to find the coefficients regression that best fit the observations, by minimizing the summed square of residuals. We can

define the residuals as the deviation of the points from the regression line.

3 Results and discussion

In the aim of comparing FNN with MLR, for each method we pass through 3 steps: 1) learning (model construction): using 4792 *in situ* data points, 2) validation: used to assess model's ability to estimate $f(\text{CO}_2^{\text{sw}})$ from another *in situ* dataset. The observations are located in the same area as the learning set (from 8°N-61°W to 23°N-20°W), and in the same time period of the year (spring). This set contains 2396 points, and 3) test: using satellite data, this step represents the most robust test of model's capacity to estimate basin wide $f(\text{CO}_2^{\text{sw}})$. It consists in computing $f(\text{CO}_2^{\text{sw}})$ exclusively from satellite data, after that, comparing results with *in situ* data.

3.1 Models construction

In order to avoid overfitting, *in situ* data was separated into 2 subsets, one for model construction, and the other one for validation. The separation was done in such a way that the learning set, as well as the validation set cover the whole $f(\text{CO}_2^{\text{sw}})$ interval.

The FNN which is used for the comparison, was chosen among 9 networks, with a hidden neuron number ranging from 2 to 10 (since a neuron network with one hidden neuron is similar to linear regression). The FNN with 6 hidden neurons gave the best results, with a root mean square error (RMSE) of 8.7 μatm in the test step (Table 1). Furthermore, the multiple linear equation which was developed for this comparison is:

$$\begin{aligned} f(\text{SST, SSS, chl-}a) &= \beta_0 + \beta_1 \times (\text{SST}) \\ &+ \beta_2 \times (\text{SSS}) \\ &+ \beta_3 \times (\text{chl-}a) \end{aligned} \quad (1)$$

With: $\beta_0 = +180.47$, $\beta_1 = -1.89$, $\beta_2 = +6.61$, $\beta_3 = +26.50$

Errors of learning, validation and test are presented in Table 1. We note that FNN with 6 hidden neurons provided good RMSE values compared to linear regression. To better see the difference between these two models, we present in figures 2, 3, and 4, the results of the three steps cited above. It is clear in figures 2 and 3, that in learning and validation, the slope and the intercept are good using FNN. Furthermore, the coefficient of determination (R^2) is high using FNN (it reaches 70% , versus 53% using MLR). We note also that in the three figures, data presented with triangles, are data for the day 16 April 2009 which contains most of the high errors (reasons are explained in the end of this section) in the three steps, as we can see, for this day, there are more data misfitted in (b) than in (a). Thus, we chose to use the neural network with 6 hidden neurons to interpolate $f(\text{CO}_2^{\text{sw}})$.

Hidden neuron number	Learning		Validation		Test	
	RMSE (μatm)	Bias (μatm)	RMSE (μatm)	Bias (μatm)	RMSE (μatm)	Bias (μatm)
2	9.09	0.25	9.08	0.23	8.91	1.24
3	8.64	0.01	8.66	-0.05	10.07	1.07
4	8.38	0.01	8.41	-0.05	9.75	1.70
5	7.88	-0.05	7.91	-0.10	10.74	1.23
6	7.75	0.03	7.77	0.04	8.72	0.51
7	7.94	-0.01	7.94	-0.04	9.97	2.08
8	6.79	0.15	6.83	0.10	11.12	2.04
9	6.55	0.01	6.56	-0.03	10.02	2.53
10	5.74	0.14	5.79	0.12	11.12	1.73
Linear regression	9.42	approx. 0.00	9.41	-0.03	9.63	0.82

Table 1: Results of learning, validation and test of the 9 neural networks, as well as linear regression. The best RMSE (8.72 μatm) corresponds to the FNN with 6 hidden neurons (from 8°N-61°W to 23°N-20°W).

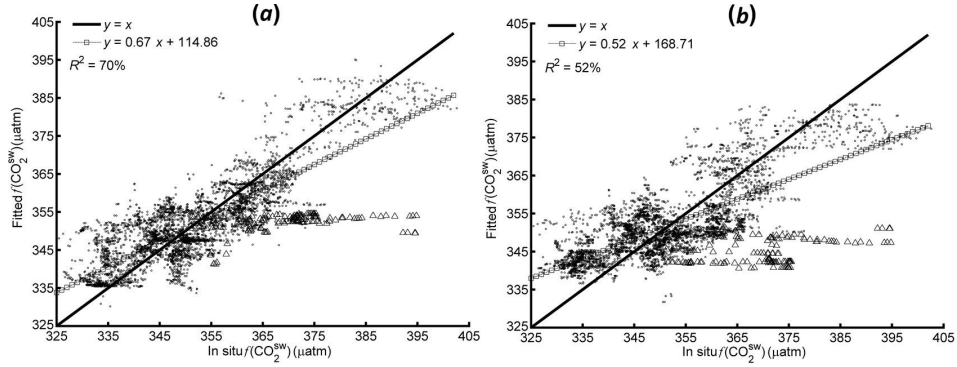


Figure 2: Result of models construction. Fitted $f(\text{CO}_2^{\text{sw}})$ as function of *in situ* $f(\text{CO}_2^{\text{sw}})$ (from 8°N-61°W to 23°N-20°W). (a): FNN with 6 hidden neurons, and (b): linear regression. Triangle points represent data for the day 16 April 2009.

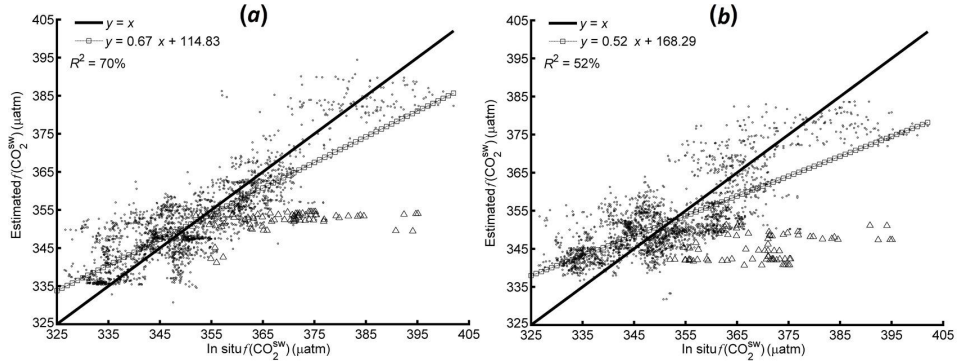


Figure 3: Result of models validation. Estimated $f(\text{CO}_2^{\text{sw}})$ as function of *in situ* $f(\text{CO}_2^{\text{sw}})$ (from 8°N-61°W to 23°N-20°W). (a): FNN with 6 hidden neurons, and (b): linear regression. Triangle points represent data for the day 16 April 2009.

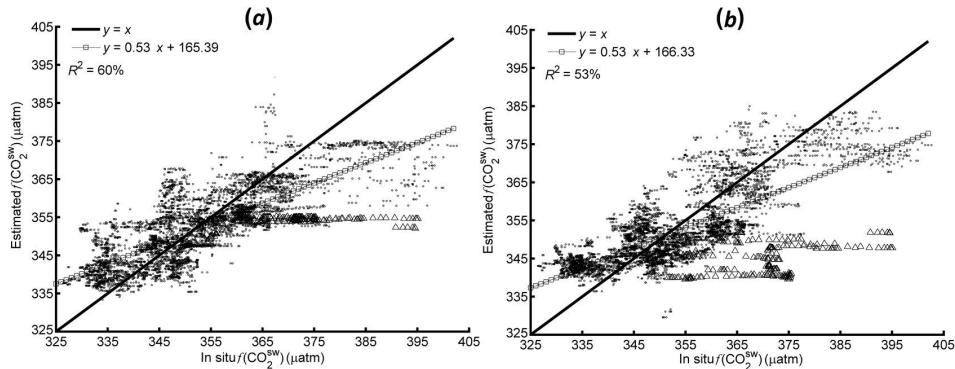


Figure 4: Result of models test. Estimated $f(\text{CO}_2^{\text{sw}})$ as function of *in situ* $f(\text{CO}_2^{\text{sw}})$ (from 8°N - 61°W to 23°N - 20°W). (a): FNN with 6 hidden neurons, and (b): linear regression. Triangle points represent data for the day 16 April 2009.

The FNN configuration is as follow: Three input neurons, corresponding to the three parameters: SST, SSS, and chl-*a*, 6 hidden neurons, and one output neuron for $f(\text{CO}_2^{\text{sw}})$. The model is valid for the following intervals: SST: [19°C – 25°C], SSS: [34.0 – 37.6], chl-*a*: [0.01 mg m^{-3} – 1.12 mg m^{-3}], and $f(\text{CO}_2^{\text{sw}})$: [325

μatm – $403\text{ }\mu\text{atm}$]. chl-*a* measurements have a particularly skewed distribution, in order to make it less skewed, we apply the \log_{10} transformation. The RMSE between *in situ* and fitted $f(\text{CO}_2^{\text{sw}})$ in learning step is of $7.8\text{ }\mu\text{atm}$ with a tiny bias ($0.03\text{ }\mu\text{atm}$) (Figure 2, Table 2).

Years	Months	RMSE (μatm)	Bias (μatm)	Number of data points
2001	April	9.95	-7.13	184
2002	March	13.23	-12.45	62
2004	March	6.61	3.10	364
2006	March	5.28	-3.54	591
2007	March	6.54	-3.63	358
	April	5.52	3.41	847
2009	March	7.38	-1.81	1746
	April	11.90	7.46	640
Total		7.75	0.03	4792

Table 2: Comparison between *in situ* $f(\text{CO}_2^{\text{sw}})$ and fitted $f(\text{CO}_2^{\text{sw}})$ in the FNN learning step (from 8°N - 61°W to 23°N - 20°W).

The largest residuals ($> 20\text{ }\mu\text{atm}$) are located between the longitudes 29°W and 20°W . In this band, the correlation coefficient between $f(\text{CO}_2^{\text{sw}})$ and estimation parameters (SST, SSS, and chl-*a*) are as follow: $f(\text{CO}_2^{\text{sw}})$ -SST: -0.6 , $f(\text{CO}_2^{\text{sw}})$ -SSS: 0.4 and $f(\text{CO}_2^{\text{sw}})$ -chl-*a*: 0.6 . Since the $f(\text{CO}_2^{\text{sw}})$ -SSS correlation is low, we present in figure 5 the associated residuals of *in situ* $f(\text{CO}_2^{\text{sw}})$ and: SST in (a) and chl-*a* in (b). We note that the highest residuals (red squares) correspond to points in which the relationship between the 2 parameters ($f(\text{CO}_2^{\text{sw}})$ and SST/chl-*a*) is different from the general trend. As presented in Table 3, for $\text{SST} \geq 22.5^\circ\text{C}$ and $\text{chl-}a \leq 0.15\text{ mg m}^{-3}$, most of the $f(\text{CO}_2^{\text{sw}})$ values (97%) are $\leq 371\text{ }\mu\text{atm}$. This is not the case of data for the red squares in figure 5; this may be one of the reasons why the residual increases.

We see in Table 2 that there are 2 months in which RMSE is large: March 2002 and April 2009. In March

2002, residual values are between $15\text{ }\mu\text{atm}$ and $26\text{ }\mu\text{atm}$. In this month, half of the set contains data with low density (according to $f(\text{CO}_2^{\text{sw}})$ values of all training data set); 40% of the points have $f(\text{CO}_2^{\text{sw}}) \leq 333\text{ }\mu\text{atm}$. As shown in figure 6, there are 238 data points with $f(\text{CO}_2^{\text{sw}}) \leq 333\text{ }\mu\text{atm}$, which represent only 5% of all training data set.

	$f(\text{CO}_2^{\text{sw}}) \leq 371$	$f(\text{CO}_2^{\text{sw}}) \geq 371$
Proportion of data (%)*	97	3
Minimum residual (μatm)	0.0	19.7
Maximum residual (μatm)	26.4	47.3

Table 3: Minimum and maximum residuals according to SST, chl-*a*, and $f(\text{CO}_2^{\text{sw}})$ values (from 8°N - 61°W to 23°N - 20°W). *Percentage relative to data with $\text{SST} \geq 22.5^\circ\text{C}$ and $\text{chl-}a \leq 0.15\text{ (mg m}^{-3})$.

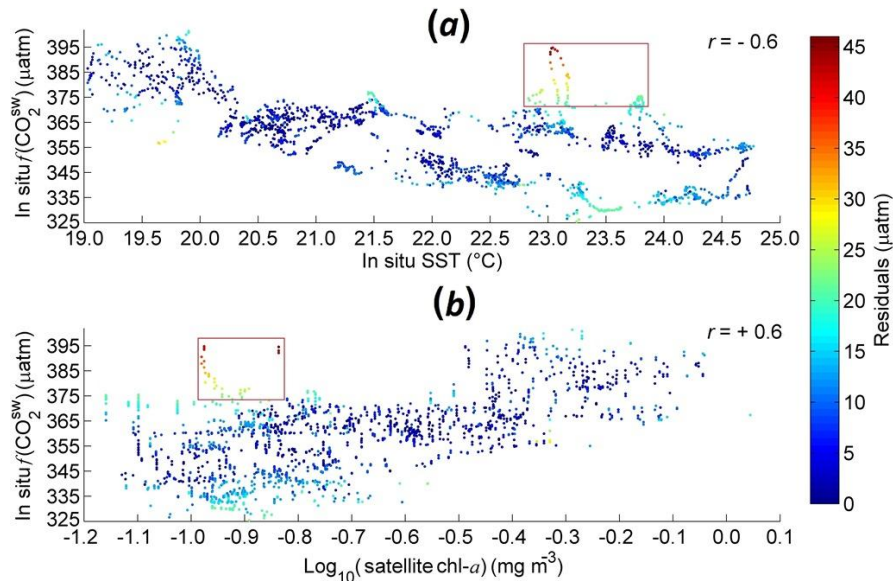


Figure 5: Residuals of FNN learning step as function of: (a): $f(\text{CO}_2^{\text{sw}})$ -SST relationship, and (b): $f(\text{CO}_2^{\text{sw}})$ -chl- a relationship (from 8°N - 29°W to 23°N - 20°W).

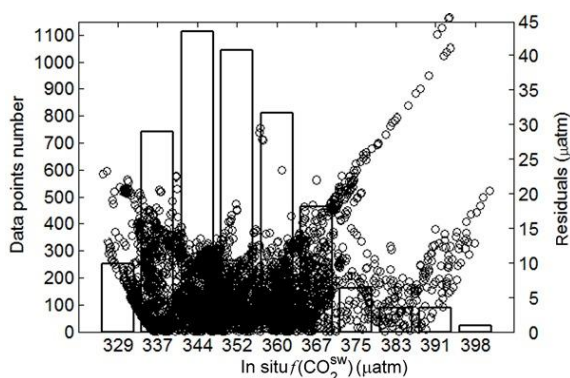


Figure 6: Relationship between $f(\text{CO}_2^{\text{sw}})$ density (in learning) data set) and residual values (of learning step)(from 8°N - 61°W to 23°N - 20°W).

Unlike March 2002, in April 2009, data points with $f(\text{CO}_2^{\text{sw}}) \geq 371 \mu\text{atm}$ represent 19% of all data in this month, but for most of these points, the residual values are very large, reaching $47 \mu\text{atm}$. Figure 6 shows that data points with $f(\text{CO}_2^{\text{sw}}) \geq 371 \mu\text{atm}$ represent 8% (392 points) of all training data set. But this low density is not the only cause of these large residuals. Figure 7 shows all the data for April 2009. We can see that there is one day in which residuals are large: 16 April 2009. At the bottom of this figure (figure 7 (b)), we see that when $f(\text{CO}_2^{\text{sw}})$ variability is high, residuals are high (day (i): 16 April 2009), and decrease when $f(\text{CO}_2^{\text{sw}})$ variability is low (day (ii): 17 April 2009, and day (iii): 18 April 2009).

There are 160 data points for 16 April 2009, in a duration of 10 hours, with a time difference of between 2 to 10 minutes between 2 successive points. Yet, there is a difference of $43 \mu\text{atm}$ between maximum and minimum of $f(\text{CO}_2^{\text{sw}})$ for this day (Table 4). Other parameters (SST, SSS, and chl- a) have low vari-

ation comparing to $f(\text{CO}_2^{\text{sw}})$, we see in Table 4 that SST (that most affect $f(\text{CO}_2^{\text{sw}})$) represents a difference of 1.8°C between maximum and minimum value. We conclude that there is another process causing this high variability. There are several current systems in the tropical Atlantic ocean that affect the $f(\text{CO}_2^{\text{sw}})$ distribution. In boreal spring, South of 20°N , the wind field off the North African coast is Southward, leading to the African Coastal Current, and local upwelling (Stramma and Schott, 1999; Mittelstaedt, 1983). We note that data points of 16 April 2009 (figure 7) are close to an upwelling zone (Stramma and Schott, 1999; Mittelstaedt, 1983). That could be the reason of high $f(\text{CO}_2^{\text{sw}})$ variability.

Parameter	Minimum	Maximum	Range (Max-Min)
$f(\text{CO}_2^{\text{sw}})$ (μatm)	351	394	43
SST ($^\circ\text{C}$)	23.0	24.8	1.8
SSS	35.67	35.89	0.23
chl- a (mg m^{-3})	0.07	0.15	0.08
Latitude	9	13	4
Longitude	-27	-25	2

Table 4: Maximum and minimum of each parameter for 16 April 2009.

3.2 The FNN validation step

The comparison between *in situ* $f(\text{CO}_2^{\text{sw}})$ observations and estimated $f(\text{CO}_2^{\text{sw}})$ is presented in figure 3 (a). The validation set is approximately the same as the learning one, and we note that in figure 3 (a) the trend is close to that in learning step (figure 2 (a), Section 3.1) with almost the same slope (S) and intercept (I) (Learning: $S=0.67$, $I=114.86$; Validation: $S=0.67$, $I=114.83$), thus we can say that the FNN has achieved the validation step, and it is capable of pro-

ducing $f(\text{CO}_2^{\text{sw}})$ from a different *in situ* dataset. Table 5 shows that residual values of each year are close to residuals of the learning step (Table 2) and RMSE increases in March 2002 and April 2009 for the same reasons (the low density of $f(\text{CO}_2^{\text{sw}})$ values, and the high $f(\text{CO}_2^{\text{sw}})$ variability caused by a local upwelling).

In both learning and validation steps, we note in

Table 2 and Table 5 that bias tends to be larger in early years. This could be caused by the $f(\text{CO}_2^{\text{sw}})$ trend, which increased of about $30 \mu\text{atm}$ between the years 2001 and 2009, and by the FNN trained to fit densely sampled year 2009 (see "Number of data points" in Table 2 and Table 5).

Years	Months	RMSE (μatm)	Bias (μatm)	Number of data points
2001	April	10.04	-7.12	91
2002	March	13.12	-12.30	32
2004	March	6.79	3.32	182
2006	March	5.26	-3.49	295
2007	March	6.46	-3.72	179
	April	5.53	3.42	423
2009	March	7.39	-1.83	873
	April	11.90	7.40	321
Total		7.77	0.04	2396

Table 5: Comparison between in situ $f(\text{CO}_2^{\text{sw}})$ and estimated $f(\text{CO}_2^{\text{sw}})$ in the FNN validation step (from 8°N - 61°W to 23°N - 20°W).

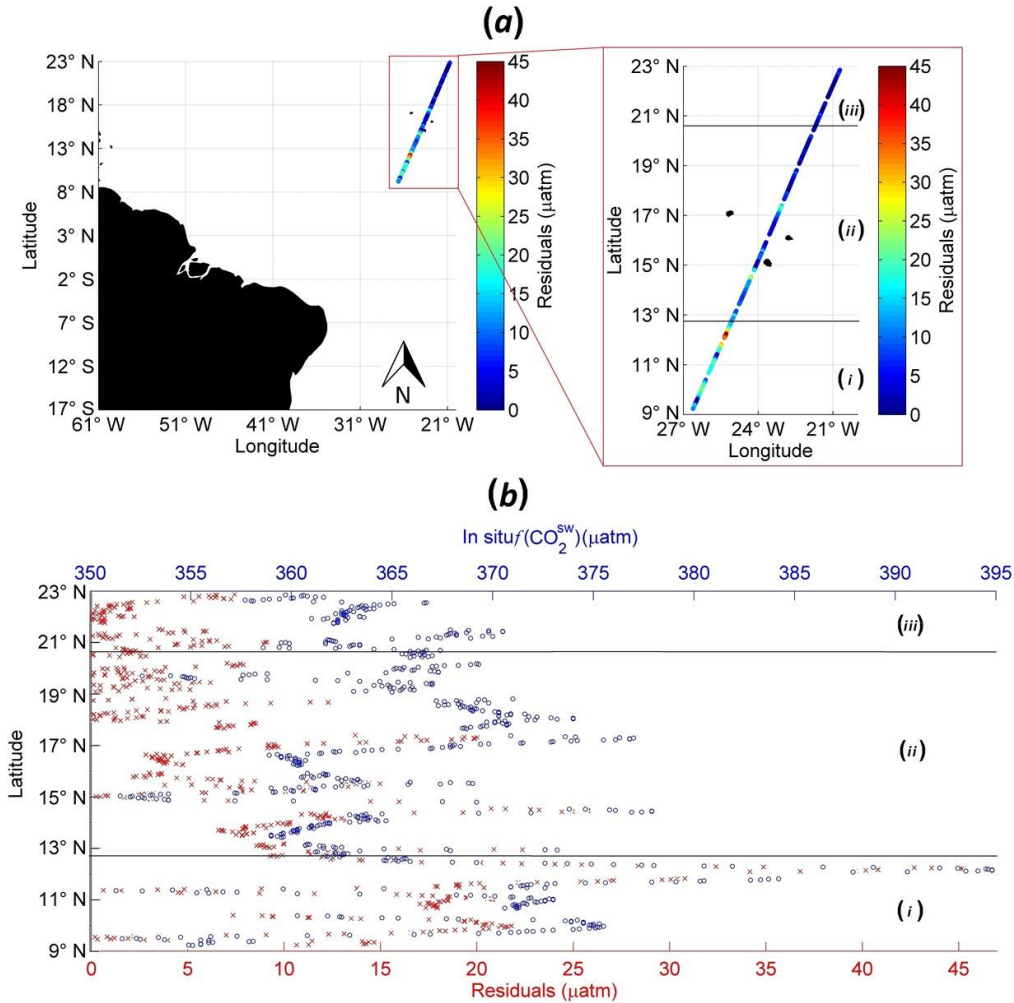


Figure 7: April 2009; (a): Residuals of learning step as function of days and geographical position. (b): Residuals of learning step as function of days, $f(\text{CO}_2^{\text{sw}})$ variations and latitude. (i): 16 April 2009, 13h02 to 23h57. (ii): 17 April 2009, 00h02 to 23h57. (iii): 18 April 2009, 00h00 to 10h40.

3.3 Test of the FNN with satellite data

In order to obtain monthly $f(\text{CO}_2^{\text{sw}})$ map with 4 km \times 4 km resolution, we used data provided from MODIS Aqua (<http://oceandata.sci.gsfc.nasa.gov/MODISA/Mapped/Monthly/4km/>) which are available since July 2002. Thus, to make a comparison with *in situ* data and then to compute residuals, we used satellite data for the years: 2004, 2006, 2007 and 2009. FNN should estimate $f(\text{CO}_2^{\text{sw}})$ from satellite SST, satellite SSS, and satellite chl-*a* with good accuracy. Since satellite SSS is available only since 2010 (provided by SMOS (Soil Moisture and Ocean Salinity) and Aquarius/SAC-D) at \sim 100-300 km spatial resolution with uncertainty around 0.3 in the Tropics (Lagerloef et al., 2008; Reul et al., 2012), we follow the approach used in Moussa et al. (2015) and Benallal et al. (2015) to get interpolated SSS data from satellite SST and latitude (Lat). We performed a model with all *in situ* data we dispose (7188 points). We determined the following equation:

$$\begin{aligned}
 f(\text{SST}, \text{Lat}) &= \beta_0 + \beta_1 \times (\text{SST}) \\
 &+ \beta_2 \times (\text{Lat}) + \beta_3 \times (\text{SST}) \\
 &\times (\text{Lat}) + \beta_4 \times (\text{SST})^2 + \beta_5 \times (\text{Lat})^2
 \end{aligned}
 \tag{2}$$

With 4 km \times 4 km spatial resolution, RMSE = 0.2 and bias = 0.02, within 8°N-61°W to 23°N-20°W:

$$\beta_0 = +61.91, \beta_1 = -1.92, \beta_2 = -0.7262, \beta_3 = +0.0305, \beta_4 = +0.03312, \beta_5 = +0.005104.$$

Then, we computed interpolated $f(\text{CO}_2^{\text{sw}})$. In order to allow comparison between *in situ* and interpolated $f(\text{CO}_2^{\text{sw}})$, we took all points in which we have both *in situ* and satellite data (6819 points). Results are presented in figure 4 (a) and Table 6. We note that, except RMSE of March 2007 which increased by 4 μatm , and RMSE of March 2009 by 2 μatm , RMSE of the remaining months are close to those of learning step (Table 2; Section 3.1) and those of validation step (Table 5; Section 3.2). This shows that model is capable of providing interpolated $f(\text{CO}_2^{\text{sw}})$ exclusively from satellite data with a low RMSE (8.7 μatm).

Years	Months	RMSE (μatm)	Bias (μatm)	Number of data points
2004	March	5.91	-0.21	546
2006	March	5.97	-3.65	886
2007	March	10.68	-8.78	537
	April	5.96	4.15	1270
2009	March	9.57	-0.20	2619
	April	11.29	7.09	961
Total		8.72	0.51	6819

Table 6: Comparison between *in situ* $f(\text{CO}_2^{\text{sw}})$ and estimated $f(\text{CO}_2^{\text{sw}})$ in the FNN test step (from 8°N-61°W to 23°N-20°W).

Although the biases of satellite SST and interpolated SSS are insignificant (-0.02 °C and 0.02, respectively), we tried to correct them, however, the improvement of the RMSE value of $f(\text{CO}_2^{\text{sw}})$ interpolation was slight (an improvement of 0.02 μatm). Moreover, since the main objective of the paper is to map $f(\text{CO}_2^{\text{sw}})$ using satellite data, without having *in situ* data to correct them, it would be better not to correct these biases.

We represent in figure 8 an example of $f(\text{CO}_2^{\text{sw}})$ mapping for March, and April 2009, using satellite SST, interpolated SSS, and satellite chl-*a*. The resolution is of 4 km \times 4 km. Since the model was developed for data points with SST within the interval [19 °C—25 °C], points not belonging to this interval are presented in white in figure 8.

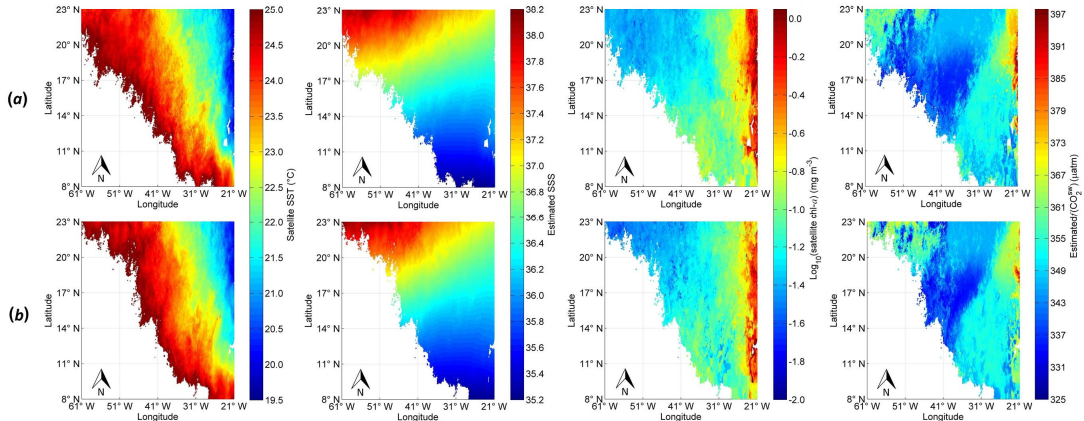


Figure 8: Mapped $f(\text{CO}_2^{\text{sw}})$ of spring 2009, with resolution of 4 km \times 4 km, using satellite SST, interpolated SSS, and satellite chl-*a*. (a): March 2009, and (b): April 2009 (from 8°N-61°W to 23°N-20°W).

We see in this figure, that $f(\text{CO}_2^{\text{sw}})$ increases Eastward, from $325 \mu\text{atm}$ to $397 \mu\text{atm}$. For the 2 months (March and April), West of 35°W , we note an increase of $f(\text{CO}_2^{\text{sw}})$ Northward. These variations are associated with those of SST, SSS, and chl-*a*. We note that SST increases Westward from 19.7°C to 24.8°C , SSS increases Northward from 35.2 to 38.0, and chl-*a* increases Eastward from 0.003 mg m^{-3} to 1.1 mg m^{-3} . Since the spatial variation of $f(\text{CO}_2^{\text{sw}})$ is non-linear, for each parameter (SST, SSS, and chl-*a*) it is difficult to determine where the correlation with $f(\text{CO}_2^{\text{sw}})$ is high. Yet, for both months, East of 35°W , $f(\text{CO}_2^{\text{sw}})$ variations are mainly influenced by those of SST and chl-*a*, for both parameters, correlation with $f(\text{CO}_2^{\text{sw}})$ is high; -0.6 and 0.8 respectively (in contrast to that between $f(\text{CO}_2^{\text{sw}})$ and SSS, which is of 0.1 East of 35°W). However, West of 35°W , the variations of $f(\text{CO}_2^{\text{sw}})$ are mostly affected by SSS. The correlation between these two parameters ($f(\text{CO}_2^{\text{sw}})$ and SSS) is high; 0.6 (unlike $f(\text{CO}_2^{\text{sw}})$ -SST and $f(\text{CO}_2^{\text{sw}})$ -chl-*a* correlations, which are of 0.2 and -0.2 respectively). Therefore, each of the three parameters (SST, SSS, and chl-*a*) affect significantly $f(\text{CO}_2^{\text{sw}})$.

In order to present an example of comparison with the climatology of Takahashi et al. (2014) developed using *in situ* measurements, we produced $f(\text{CO}_2^{\text{sw}})$ map for March 2005, based upon satellite data. In figure 9, we present in (a) the climatology of Takahashi et al. (2014) for March 2005, and in (b) the $f(\text{CO}_2^{\text{sw}})$ map produced here using the FNN, presented in the same $f(\text{CO}_2^{\text{sw}})$ color scale as in (a). Furthermore, in order to see the $f(\text{CO}_2^{\text{sw}})$ distribution more clearly, we present in (c) the same map as shown in (b) with a different $f(\text{CO}_2^{\text{sw}})$ color scale. We note that for the 2 products, the $f(\text{CO}_2^{\text{sw}})$ increase Eastward from $339 \mu\text{atm}$ to $377 \mu\text{atm}$ in (a), and from $323 \mu\text{atm}$ to $376 \mu\text{atm}$ in (b). There is almost the same $f(\text{CO}_2^{\text{sw}})$ range in these 2 products with a difference of resolution; the Takahashi et al. (2014) map has a spatial resolution of $4^\circ \times 5^\circ$, yet, the FNN produces a $f(\text{CO}_2^{\text{sw}})$ map with a higher resolution: $0.04^\circ \times 0.04^\circ$.

In order to best represent the differences between these two products, we show in figure 10 (a) the climatological $f(\text{CO}_2^{\text{sw}})$ of Takahashi et al. (2014) on the color scale of $f(\text{CO}_2^{\text{sw}})$ produced by FNN. Moreover, we present in figure 10 (b) the FNN product with a reduced resolution, performed by computing the $f(\text{CO}_2^{\text{sw}})$ mean of each bloc of 4° latitude \times 5° longitude, therefore, the scale of $f(\text{CO}_2^{\text{sw}})$ produced by FNN was of $[323 \mu\text{atm} - 357 \mu\text{atm}]$. Figure 10 (c) presents the difference between the two products, the maximum difference value is of $38 \mu\text{atm}$. We note that large differences are located in West of 36°W , where climatological $f(\text{CO}_2^{\text{sw}})$ are overestimated compared with FNN $f(\text{CO}_2^{\text{sw}})$.

Overall, these results suggest that in this ocean area, the climatological estimates (Takahashi et al., 2014) would tend to overestimate $f(\text{CO}_2^{\text{sw}})$ by a mean of $12 \mu\text{atm}$ over this ocean area. Since figure 9 (b) presents a higher resolution than figure 9 (a), it is now possible to detect gradual changes in $f(\text{CO}_2^{\text{sw}})$.

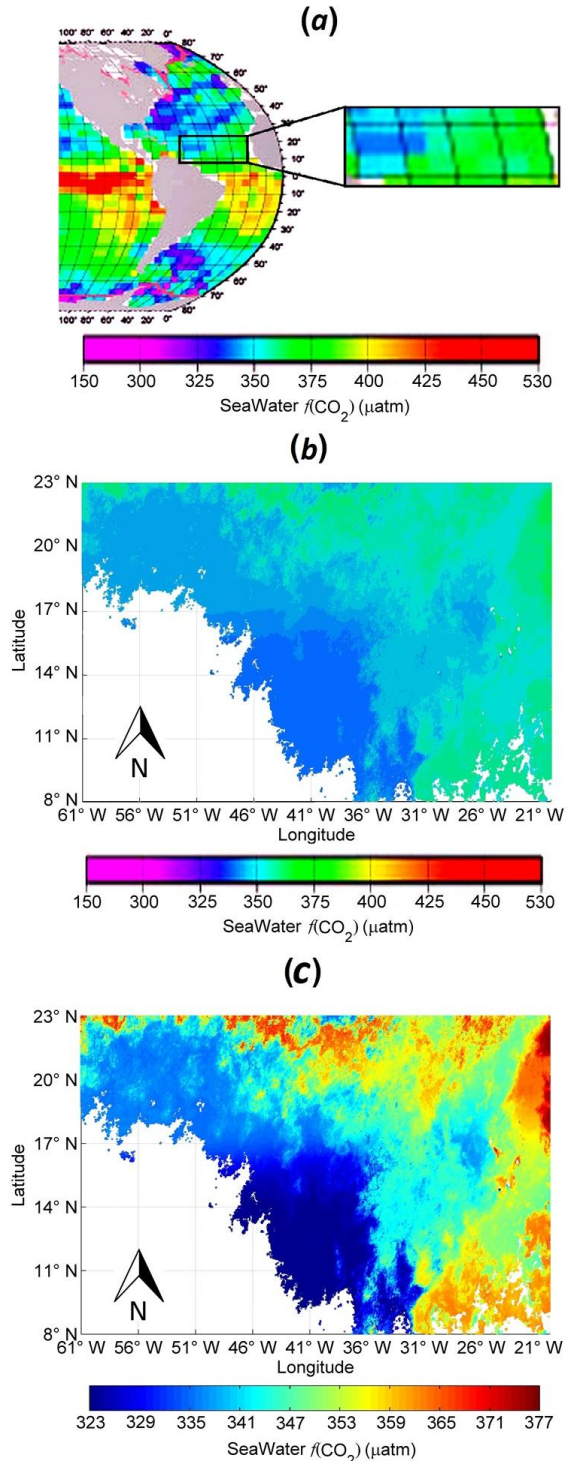


Figure 9: Distributions of surface seawater $f(\text{CO}_2^{\text{sw}})$ in March 2005: (a) Climatological map of Takahashi et al. (2014) ($4^\circ \times 5^\circ$ spatial resolution). (b) Map produced by the developed FNN, presented on the same $f(\text{CO}_2^{\text{sw}})$ color scale as in (a) ($0.04^\circ \times 0.04^\circ$ spatial resolution). (c) The same map as in (b) with a color scale between minimum and maximum of $f(\text{CO}_2^{\text{sw}})$ in this region.

Therefore, the fine resolution that offer remote sensing data, allows us to get a smoothness representation of detailed $f(\text{CO}_2^{\text{sw}})$ concentrations, and signifi-

cantly improve our knowledge from the relatively rough climatological resolution (only $4^\circ \times 5^\circ$). The use of fine resolution ($0.04^\circ \times 0.04^\circ$) satellite products to determine $f(\text{CO}_2^{\text{sw}})$ over large ocean areas, will undoubtedly improve the accuracy of the CO_2 air-sea fluxes estimates over the global ocean.

4 Conclusion

We have presented a methodological study that shows: 1) the feedforward neural network represents a good alternative for $f(\text{CO}_2^{\text{sw}})$ interpolation, 2) the feedforward neural network provides better results than multiple linear regression, and 3) it becomes possible to estimate surface $f(\text{CO}_2^{\text{sw}})$ exclusively from satellite products, with a fine resolution ($4 \text{ km} \times 4 \text{ km}$) over a large ocean area (from 8°N - 61°W to 23°N - 20°W) with an uncertainty of $\pm 8.7 \mu\text{atm}$ (which is in the order of a diurnal variation magnitude). We have investigated the parameters needed to estimate $f(\text{CO}_2^{\text{sw}})$, the results show that SST, SSS, and chl-*a* are sufficient as input of the model to provide a basin-wide maps of $f(\text{CO}_2^{\text{sw}})$ with a good accuracy.

This study further shows that the most crucial limiting factor for the accuracy of the interpolation is the high variability of $f(\text{CO}_2^{\text{sw}})$. The second reason is the lack of data in some areas. Additional *in situ* observations are still needed in addition to satellite data to validate interpolation.

The comparison with the climatological $f(\text{CO}_2^{\text{sw}})$ map of Takahashi et al. (2014), shows that the $f(\text{CO}_2^{\text{sw}})$ map produced by the FNN provides new insights. In particular, the FNN method improves the accuracy of the $f(\text{CO}_2^{\text{sw}})$ gradient, which provides a substantial benefits regarding to the CO_2 fluxes estimates.

Overall, this study is a first step toward, and opens the route to the quantification of $f(\text{CO}_2^{\text{sw}})$ over the whole ocean based upon fine spatial and temporal resolution of satellite SST, SSS, and chl-*a* fields.

4.1 Acknowledgements

We greatly acknowledge Rawa Hassan EL Fallah and Elissar Gemayel for very helpful discussions and comments. This study has been performed as part of H. MOUSSA's Ph.D., which is funded by the MESRS (Ministère Algérien de l'Enseignement Supérieur et de la Recherche Scientifique).

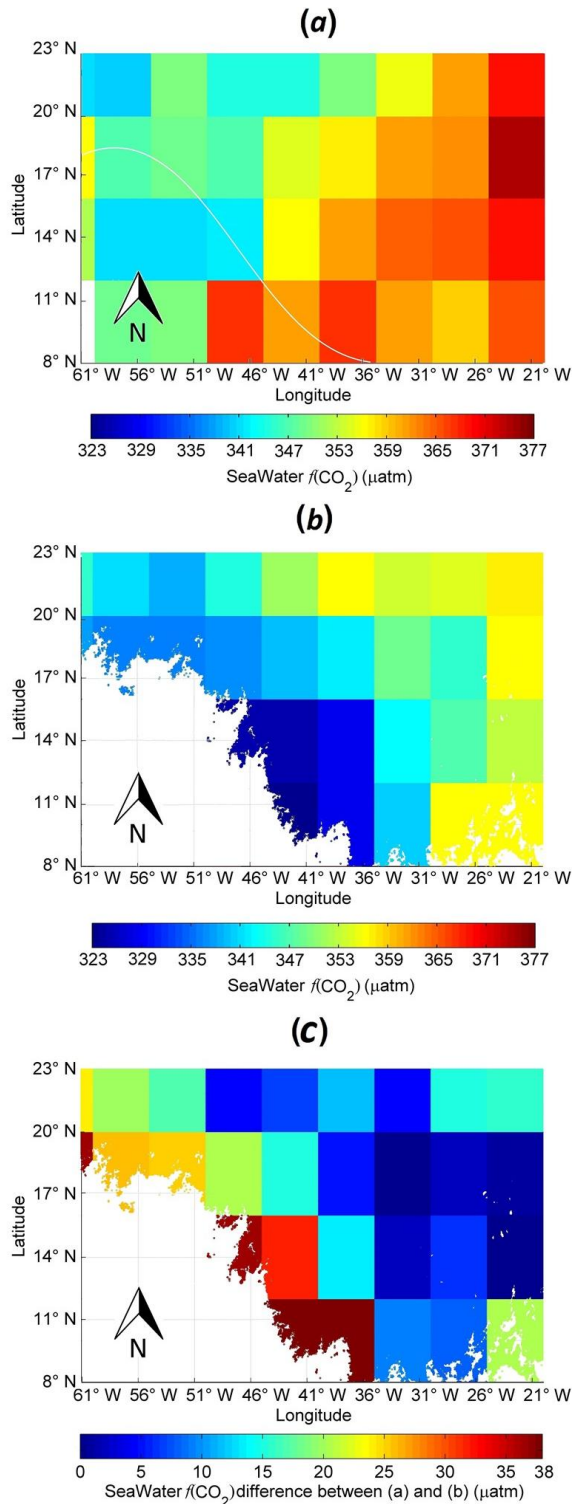


Figure 10: Distributions of surface seawater $f(\text{CO}_2^{\text{sw}})$ in March 2005: (a) Climatological map of Takahashi et al. (2014) ($4^\circ \times 5^\circ$ spatial resolution), the white line represents the limit below which there are no data in (b) and (c). (b) Map produced by the developed FNN ($4^\circ \times 5^\circ$ spatial resolution). (c) The difference between $f(\text{CO}_2^{\text{sw}})$ maps in (a) and (b).

References

- Barreto, G. A. (2007). Time Series Prediction with the Self-Organizing Map: A Review. In Hammer, D. B. and Hitzler, D. P., editors, *Perspectives of Neural-Symbolic Integration*, volume 77, pages 135–158. SPRINGER BERLIN HEIDELBERG.
- Barreto, G. A. and Araújo, A. F. R. (2002). Non-linear Modeling of Dynamic Systems with the Self-Organizing Map. In Dorronsoro, J. R., editor, *Artificial Neural Networks — ICANN 2002*, volume 2415, pages 975–980. SPRINGER BERLIN HEIDELBERG.
- Benallal, M. A., Moussa, H., Touratier, F., Goyet, C., and Poisson, A. (2015). Ocean salinity from satellite-derived temperature in the Antarctic Ocean. *Antarctic Science*, FirstView:1–8.
- Bureau International Des Poids Et Mesures (1955). Dixième conférence générale des poids et mesures. Technical report, Bureau International Des Poids Et Mesures, Quai des Grands-Augustins, 55, Paris.
- Chen, C.-T. A. (1982). On the distribution of anthropogenic CO₂ in the Atlantic and Southern oceans. *DEEP-SEA RESEARCH PART I-OCEANOGRAPHIC RESEARCH PAPERS*, 29:563–580.
- Copin-Montegut, C. (1988). A new formula for the effect of temperature on the partial pressure of CO₂ in seawater. *MARINE CHEMISTRY*, 25:29–37.
- Copin-montegut, C. (1989). A new formula for the effect of temperature on the partial pressure of carbon dioxide in seawater, corrigendum. *MARINE CHEMISTRY*, 27:143–144.
- de Villiers, J. and Barnard, E. (1993). Backpropagation neural nets with one and two hidden layers. *IEEE TRANSACTIONS ON NEURAL NETWORKS*, 4:136–141.
- Dlugokencky, E. and Tans, P. (2015). Trends in atmospheric carbon dioxide.
- Flexer, A. (2001). On the Use of Self-Organizing Maps for Clustering and Visualization. *INTELLIGENT DATA ANALYSIS*, 5:373–384.
- Friedrich, T. and Oschlies, A. (2009). Neural network-based estimates of North Atlantic surface pCO₂ from satellite data: A methodological study. *JOURNAL OF GEOPHYSICAL RESEARCH-OCEANS*, 114:C03020.
- Gordon, L. I. and Jones, L. B. (1973). The effect of temperature on carbon dioxide partial pressures in seawater. *MARINE CHEMISTRY*, 1:317–322.
- Goyet, C., Adams, R., and Eiseid, G. (1998). Observations of the CO₂ system properties in the tropical Atlantic Ocean. *MARINE CHEMISTRY*, 60:49–61.
- Goyet, C., Millero, F. J., Poisson, A., and Shafer, D. K. (1993). Temperature dependence of CO₂ fugacity in seawater. *MARINE CHEMISTRY*, 44:205–219.
- Goyet, C. and Peltzer, E. T. (1994). Comparison of the August–September 1991 and 1979 surface partial pressure of CO₂ distribution in the Equatorial Pacific Ocean near 150°W. *MARINE CHEMISTRY*, 45:257–266.
- Goyet, C. and Peltzer, E. T. (1997). Variation of CO₂ partial pressure in surface seawater in the equatorial Pacific Ocean. *DEEP-SEA RESEARCH PART I-OCEANOGRAPHIC RESEARCH PAPERS*, 44:1611–1625.
- Goyet, C. and Poisson, A. (1989). New determination of carbonic acid dissociation constants in seawater as a function of temperature and salinity. *DEEP-SEA RESEARCH PART I-OCEANOGRAPHIC RESEARCH PAPERS*, 36:1635–1654.
- Inoue, H. Y., Ishii, M., Matsueda, H., Kawano, T., Murata, A., and Takasugi, Y. (2003). Distribution of the partial pressure of CO₂ in surface water (pcO_2^w) between Japan and the Hawaiian Islands: pcO_2^w —SST relationship in the winter and summer. *TELLUS SERIES B-CHEMICAL AND PHYSICAL METEOROLOGY*, 55:456–465.
- Jamet, C., Moulin, C., and Lefèvre, N. (2007). Estimation of the oceanic pCO₂ in the North Atlantic from VOS lines in-situ measurements: parameters needed to generate seasonally mean maps. *ANNALES DE GEOPHYSIQUE*, 25:2247–2257.
- Lagerloef, G., Colomb, F. R., Le Vine, D., Wentz, F., Yueh, S., Ruf, C., Lilly, J., Gunn, J., Chao, Y., deCharon, A., Feldman, G., and Swift, C. (2008). The Aquarius/SAC-D Mission: Designed to Meet the Salinity Remote-Sensing Challenge. *OCEANOGRAPHY*, 21:68–81.
- Landschützer, P., Gruber, N., Bakker, D. C. E., Schuster, U., Nakaoka, S., Payne, M. R., Sasse, T. P., and Zeng, J. (2013). A neural network-based estimate of the seasonal to inter-annual variability of the Atlantic Ocean carbon sink. *BIOGEOSCIENCES*, 10:7793–7815.
- Le Quéré, C., Moriarty, R., Andrew, R. M., Peters, G. P., Ciais, P., Friedlingstein, P., Jones, S. D., Sitch, S., Tans, P., Arneeth, A., Boden, T. A., Bopp, L., Bozec, Y., Canadell, J. G., Chini, L. P., Chevallier, F., Cosca, C. E., Harris, I., Hoppema, M., Houghton, R. A., House, J. I., Jain, A. K., Johannessen, T., Kato, E., Keeling, R. F., Kitidis, V., Klein Goldewijk, K., Koven, C., Landa, C. S., Landschützer, P., Lenton, A., Lima, I. D., Marland, G., Mathis, J. T., Metzl, N., Nojiri, Y., Olsen, A., Ono, T., Peng, S., Peters, W., Pfeil, B., Poulter, B., Raupach, M. R., Regnier, P., Rödenbeck, C., Saito, S., Salisbury, J. E., Schuster, U., Schwinger, J., Sferian, R., Segsneider, J., Steinhoff, T., Stocker, B. D., Sutton, A. J., Takahashi, T., Tilbrook, B., van der

- Werf, G. R., Viovy, N., Wang, Y.-P., Wanninkhof, R., Wiltshire, A., and Zeng, N. (2015). Global carbon budget 2014. *EARTH SYSTEM SCIENCE DATA*, 7:47–85.
- Lefèvre, N., J. Watson, A. J., and R. Watson, A. R. (2005). A comparison of multiple regression and neural network techniques for mapping in situ pCO₂ data. *TELLUS SERIES B-CHEMICAL AND PHYSICAL METEOROLOGY*, 57:375–384.
- Lefèvre, N., Moore, G., Aiken, J., Watson, A., Cooper, D., and Ling, R. (1998). Variability of pCO₂ in the tropical Atlantic in 1995. *JOURNAL OF GEOPHYSICAL RESEARCH-OCEANS*, 103:5623–5634.
- Lefèvre, N. and Taylor, A. (2002). Estimating pCO₂ from sea surface temperatures in the Atlantic gyres. *DEEP-SEA RESEARCH PART I-OCEANOGRAPHIC RESEARCH PAPERS*, 49:539–554.
- Lefèvre, N., Urbano, D. F., Gallois, F., and Diverrès, D. (2014). Impact of physical processes on the seasonal distribution of the fugacity of CO₂ in the western tropical Atlantic. *JOURNAL OF GEOPHYSICAL RESEARCH-OCEANS*, 119:646–663.
- Levenberg, K. (1944). A method for the solution of certain non-linear problems in least squares. *QUARTERLY JOURNAL OF APPLIED MATHEMATICS*, 2:164–168.
- Marquardt, D. (1963). An Algorithm for Least-Squares Estimation of Nonlinear Parameters. *JOURNAL OF THE SOCIETY FOR INDUSTRIAL AND APPLIED MATHEMATICS*, 11:431–441.
- McCulloch, W. S. and Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *BULLETIN OF MATHEMATICAL BIOLOGY*, 5:115–133.
- Mittelstaedt, E. (1983). The upwelling area off Northwest Africa—A description of phenomena related to coastal upwelling. *PROGRESS IN OCEANOGRAPHY*, 12:307–331.
- Moussa, H., Benallal, M. A., Goyet, C., Lefèvre, N., EL Jai, M. C., Guglielmi, V., and Touratier, F. (2015). A comparison of Multiple Non-linear regression and neural network techniques for sea surface salinity estimation in the tropical Atlantic ocean based on satellite data. In El Yacoubi, S., Afifi, L., Zerrik, E. H., and Tridane, A., editors, *CMACS 2014 - Colloque de Modélisation, Analyse et Contrôle des Systèmes*, volume 49, pages 65–78, Al-Akhawayn University, Morocco. ESAIM: PROCEEDINGS AND SURVEYS.
- Nakaoka, S., Telszewski, M., Nojiri, Y., Yasunaka, S., Miyazaki, C., Mukai, H., and Usui, N. (2013). Estimating temporal and spatial variation of ocean surface pCO₂ in the North Pacific using a self-organizing map neural network technique. *BIOGEOSCIENCES*, 10:6093–6106.
- Olsen, A., Bellerby, R. G. J., Johannessen, T., Omar, A. M., and Skjelvan, I. (2003). Interannual variability in the wintertime air–sea flux of carbon dioxide in the northern North Atlantic, 1981–2001. *DEEP-SEA RESEARCH PART I-OCEANOGRAPHIC RESEARCH PAPERS*, 50:1323–1338.
- Panchal, G., Ganatra, A., Shah, P., and Panchal, D. (2011). Determination of over-learning and over-fitting problem in back propagation neural network. *INTERNATIONAL JOURNAL ON SOFT COMPUTING*, 2:40–51.
- Park, G.-H. and Wanninkhof, R. (2012). A large increase of the CO₂ sink in the western tropical North Atlantic from 2002 to 2009. *JOURNAL OF GEOPHYSICAL RESEARCH-OCEANS*, 117:C08029.
- Pfeil, B., Olsen, A., Bakker, D. C. E., Hankin, S., Koyuk, H., Kozyr, A., Malczyk, J., Manke, A., Metzl, N., Sabine, C. L., Akl, J., Alin, S. R., Bates, N., Bellerby, R. G. J., Borges, A., Boutin, J., Brown, P. J., Cai, W.-J., Chavez, F. P., Chen, A., Cosca, C., Fassbender, A. J., Feely, R. A., González-Dávila, M., Goyet, C., Hales, B., Hardman-Mountford, N., Heinze, C., Hood, M., Hoppema, M., Hunt, C. W., Hydes, D., Ishii, M., Johannessen, T., Jones, S. D., Key, R. M., Körtzinger, A., Landschützer, P., Lauvset, S. K., Lefèvre, N., Lenton, A., Lourantou, A., Merlivat, L., Midorikawa, T., Mintrop, L., Miyazaki, C., Murata, A., Nakadate, A., Nakano, Y., Nakaoka, S., Nojiri, Y., Omar, A. M., Padin, X. A., Park, G.-H., Paterson, K., Perez, F. F., Pierrot, D., Poisson, A., Ríos, A. F., Santana-Casiano, J. M., Salisbury, J., Sarma, V. V. S. S., Schlitzer, R., Schneider, B., Schuster, U., Sieger, R., Skjelvan, I., Steinhoff, T., Suzuki, T., Takahashi, T., Tedesco, K., Telszewski, M., Thomas, H., Tilbrook, B., Tjiputra, J., Vandemark, D., Veness, T., Wanninkhof, R., Watson, A. J., Weiss, R., Wong, C. S., and Yoshikawa-Inoue, H. (2013). A uniform, quality controlled Surface Ocean CO₂ Atlas (SOCAT). *EARTH SYSTEM SCIENCE DATA*, 5:125–143.
- Reul, N., Tenerelli, J., Boutin, J., Chapron, B., Paul, F., Brion, E., Gaillard, F., and Archer, O. (2012). Overview of the First SMOS Sea Surface Salinity Products. Part I: Quality Assessment for the Second Half of 2010. *IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING*, 50:1636–1647.
- Rosenblatt, F. (1958). The perceptron: A probabilistic model for information storage and organization in the brain. *PSYCHOLOGICAL REVIEW*, 65:386–408.
- Rumelhart, D. E. and McClelland, J. L., editors (1986). *Parallel Distributed Processing: Explorations in the Microstructure of Cognition, Vol. 1: Foundations*. MIT PRESS, Cambridge, MA, USA.
- Sabine, C., Ducklow, H., and Hood, M. (2010). International Carbon Coordination: Roger Revelle’s Legacy

- in the Intergovernmental Oceanographic Commission. *OCEANOGRAPHY*, 23(3):48–61.
- Schuster, U., McKinley, G. A., Bates, N., Chevallier, F., Doney, S. C., Fay, A. R., González-Dávila, M., Gruber, N., Jones, S., Krijnen, J., Landschützer, P., Lefevre, N., Manizza, M., Mathis, J., Metzl, N., Olsen, A., Rios, A. F., Rödenbeck, C., Santana-Casiano, J. M., Takahashi, T., Wanninkhof, R., and Watson, A. J. (2013). An assessment of the Atlantic and Arctic sea—air CO₂ fluxes, 1990–2009. *BIOGEOSCIENCES*, 10:607–627.
- Stephens, M. P., Samuels, G., Olson, D. B., Fine, R. A., and Takahashi, T. (1995). Sea-air flux of CO₂ in the North Pacific using shipboard and satellite data. *JOURNAL OF GEOPHYSICAL RESEARCH-OCEANS*, 100:13571–13583.
- Stramma, L. and Schott, F. (1999). The mean flow field of the tropical Atlantic Ocean. *DEEP-SEA RESEARCH PART II-TOPICAL STUDIES IN OCEANOGRAPHY*, 46:279–303.
- Takahashi, T. (1961). Carbon dioxide in the atmosphere and in Atlantic Ocean water. *JOURNAL OF GEOPHYSICAL RESEARCH-OCEANS*, 66:477–494.
- Takahashi, T., Feely, R. A., Weiss, R. F., Wanninkhof, R. H., Chipman, D. W., Sutherland, S. C., and Takahashi, T. T. (1997). Global air-sea flux of CO₂: An estimate based on measurements of sea—air pCO₂ difference. *PROCEEDING OF THE NATIONAL ACADEMY OF SCIENCE OF THE UNITED STATES OF AMERICA*, 94:8292–8299.
- Takahashi, T., Sutherland, S. C., Chipman, D. W., Goddard, J. G., Ho, C., Newberger, T., Sweeney, C., and Munro, D. R. (2014). Climatological distributions of pH, pCO₂, total CO₂, alkalinity, and CaCO₃ saturation in the global surface ocean, and temporal changes at selected locations. *MARINE CHEMISTRY*, 164:95–125.
- Takahashi, T., Sutherland, S. C., Sweeney, C., Poisson, A., Metzl, N., Tilbrook, B., Bates, N., Wanninkhof, R., Feely, R. A., Sabine, C., Olafsson, J., and Nojiri, Y. (2002). Global sea—air CO₂ flux based on climatological surface ocean pCO₂, and seasonal biological and temperature effects. *DEEP-SEA RESEARCH PART II-TOPICAL STUDIES IN OCEANOGRAPHY*, 49:1601–1622.
- Takahashi, T., Sutherland, S. C., Wanninkhof, R., Sweeney, C., Feely, R. A., Chipman, D. W., Hales, B., Friederich, G., Chavez, F., Sabine, C., Watson, A., Bakker, D. C. E., Schuster, U., Metzl, N., Yoshikawa-Inoue, H., Ishii, M., Midorikawa, T., Nojiri, Y., Körtzinger, A., Steinhoff, T., Hoppema, M., Olafsson, J., Arnarson, T. S., Tilbrook, B., Johannessen, T., Olsen, A., Bellerby, R., Wong, C. S., Delille, B., Bates, N. R., and de Baar, H. J. W. (2009). Climatological mean and decadal change in surface ocean pCO₂, and net sea—air CO₂ flux over the global oceans. *DEEP-SEA RESEARCH PART II-TOPICAL STUDIES IN OCEANOGRAPHY*, 56:554–577.
- Tamura, S. and Tateishi, M. (1997). Capabilities of a four-layered feedforward neural network: four layers versus three. *IEEE TRANSACTIONS ON NEURAL NETWORKS*, 8:251–255.
- Tans, P. and Keeling, R. (2015). Trends in Atmospheric Carbon Dioxide.
- Telszewski, M., Chazottes, A., Schuster, U., Watson, A. J., Moulin, C., Bakker, D. C. E., González-Dávila, M., Johannessen, T., Körtzinger, A., Lüger, H., Olsen, A., Omar, A., Padin, X. A., Ríos, A. F., Steinhoff, T., Santana-Casiano, M., Wallace, D. W. R., and Wanninkhof, R. (2009). Estimating the monthly pCO₂ distribution in the North Atlantic using a self-organizing neural network. *BIOGEOSCIENCES*, 6:1405–1421.
- Xu, R. and Wunsch, D. (2005). Survey of clustering algorithms. *IEEE TRANSACTIONS ON NEURAL NETWORKS*, 16:645–678.
- Zeng, J., Nojiri, Y., Landschützer, P., Telszewski, M., and Nakaoka, S. (2014). A Global Surface Ocean fCO₂ Climatology Based on a Feed-Forward Neural Network. *JOURNAL OF ATMOSPHERIC AND OCEANIC TECHNOLOGY*, 31:1838–1849.

E.3 Publication 3 : Ocean salinity from satellite-derived temperature in the Antarctic Ocean

Ocean salinity from satellite-derived temperature in the Antarctic Ocean

M.A. BENALLAL^{1,2} H. MOUSSA^{1,2} F. TOURATIER^{1,2} C. GOYET^{1,2}
A. POISSON^{1,3}

Antarctic Science (final draft post-refereeing)
DOI:<http://dx.doi.org/10.1017/S0954102015000516>

¹IMAGES ESPACE-DEV, Université de Perpignan Via Domitia, 52 Avenue Paul Alduy, 66860 Perpignan Cedex 9, France.

²ESPACE-DEV, UG UA UR UM IRD, Maison de la télédétection, 500 Rue Jean-François Breton, 34093 Montpellier Cedex 5, France.

³Laboratoire d’Océanographie et du Climat: Expérimentation et Approches Numériques (LOCEAN/IPSL), Université Pierre et Marie Curie, Paris, France.

Mohamed.benallal@univ-perp.fr

Abstract

The aim of the MINERVE project (Mesures à l’INterface Eau-aiR de la Variabilité des Échanges de CO₂) is to observe and understand the seasonal and inter-annual variability of the partial pressure of CO₂ ($p\text{CO}_2$) in surface waters using hydrological and biogeochemical data in the Southern Ocean south of Australia. Logistics routes of the R/V Astrolabe provide access to scarcely studied areas, thus allowing us to understand the different processes acting in this area of the Antarctic Ocean. The surface area covered by these cruises, however, is tiny compared with the total surface area of the Antarctic Ocean. We apply correlations between in situ surface temperature and salinity data to satellite images of sea surface temperature to map ocean surface salinity over a much wider area than under the cruise tracks. Comparing with salinity given by satellites which provide 100 km resolution and 0.1 accuracy, we are able to map salinity at 4 km resolution and almost the same accuracy of ± 0.1 .

Received 21 October 2014, accepted 23 September 2015

Key words: mapping, Southern Ocean, surface salinity, temperature fields.

1 Introduction

The Southern Ocean plays a significant role in the exchange of CO₂ between the global atmosphere and ocean (Lenton et al., 2013). The MINERVE project (Mesures à l’INterface Eau-aiR de la Variabilité des Échanges de CO₂) was designed to observe and understand the seasonal and interannual variability of the partial pressure of CO₂ ($p\text{CO}_2$) in surface waters in the Southern Ocean south of Australia. The logistics routes of the RV Astrolabe between Hobart, Tasmania and the French Antarctic station at Dumont d’Urville provide an opportunity to collect hydrological and biogeochemical data in this remote area. However, the surface area sampled during these cruises is tiny compared with the surface area of the entire region of interest. Therefore, we seek ways to extrapolate the in situ data to a broader area using satellite images.

The exchange of CO₂ between the ocean and atmosphere is affected by both sea surface temperature (SST) and the stability of the surface mixed layer. The SST is routinely monitored by satellites. However, in cold high-latitude oceans, surface mixed layer stability is often controlled by salinity, which has only recently been mapped by satellite even though sea surface salin-

ity (SSS) is one of the fundamental variables for which sustained global observations are required to improve our knowledge and prediction of the ocean circulation, global water cycle and climate (Lagerloef et al., 2010). Two satellite missions now provide estimates of SSS.

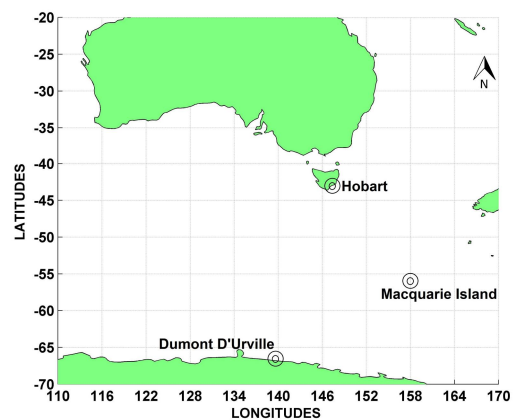


Figure 1: Regional map showing Hobart, Dumont D’Urville and Macquarie Island. Source: Google Earth.

The Aquarius satellite, a collaboration between NASA and the Space Agency of Argentina, was launched in June 2011 and operated until June 2015. Aquarius provided global maps of salinity with accuracy of ± 0.2 and a spatial resolution of 150 km (Le Vine et al., 2007). Data are available from <http://oceandata.sci.gsfc.nasa.gov/Aquarius/>. The Soil Moisture and Ocean Salinity (SMOS) satellite was launched in November 2009 and continues to operate, providing maps of salinity with an accuracy of ± 0.1 and a spatial resolution of 50–100 km (Michel, 2006). Data

are available from <http://cp34-bec.cmima.csic.es/data/data-access/>.

In this paper, we report *in situ* measurements of the interannual variability of SSS in the southern Indian Ocean during the early spring of each year between 2002 and 2009. We then describe a technique for mapping SSS in the same period of the year, but from 2010–12 using satellite-based maps of SST. The technique is then assessed by comparisons between the *in situ* data and mapped values, and with direct SSS estimates from the two salinity satellites.

Part	Year	From Hobart to Dumont D'Urville				From Dumont D'Urville to Hobart				D.N.
		Start date	End date	Lowest L.	Highest L.	Start date	End date	Highest L.	Lowest L.	
1	2002	16/10	21/10	-43.96	-63.62	30/10	03/11	-61.81	-43.11	294
	2003	21/10	26/10	-45.14	-62.82	06/11	11/11	-62.15	-43.66	388
	2004	19/10	23/10	-43.47	-61.96	04/11	09/11	-62.14	-43.29	421
	2005	18/10	22/10	-44.92	-61.88	03/11	08/11	-62.82	-43.26	470
	2006	20/10	29/10	-43.64	-66.56	03/11	12/11	-66.66	-43.45	201
	2007	28/10	01/11	-43.20	-63.44	16/11	20/11	-63.42	-44.28	348
	2008	21/10	26/10	-42.84	-61.17	02/11	11/11	-65.84	-43.36	4410
	2009	22/10	25/10	-48.58	-62.05	11/11	16/11	-62.47	-43.47	7671
	2	2010	21/10	27/10	-42.88	-63.53	31/10	07/11	-63.28	-43.51
2011		25/10	26/10	-54.59	-57.97	06/12	10/12	-63.46	-43.31	4181
2012		23/10	28/10	-42.84	-62.83	18/11	22/11	-61.70	-42.94	6169

Table 1: Details of the MINERVE database. Part 1 was used to establish the model. Part 2 was used to validate the model. L. = Latitude, D.N. = Data numbers

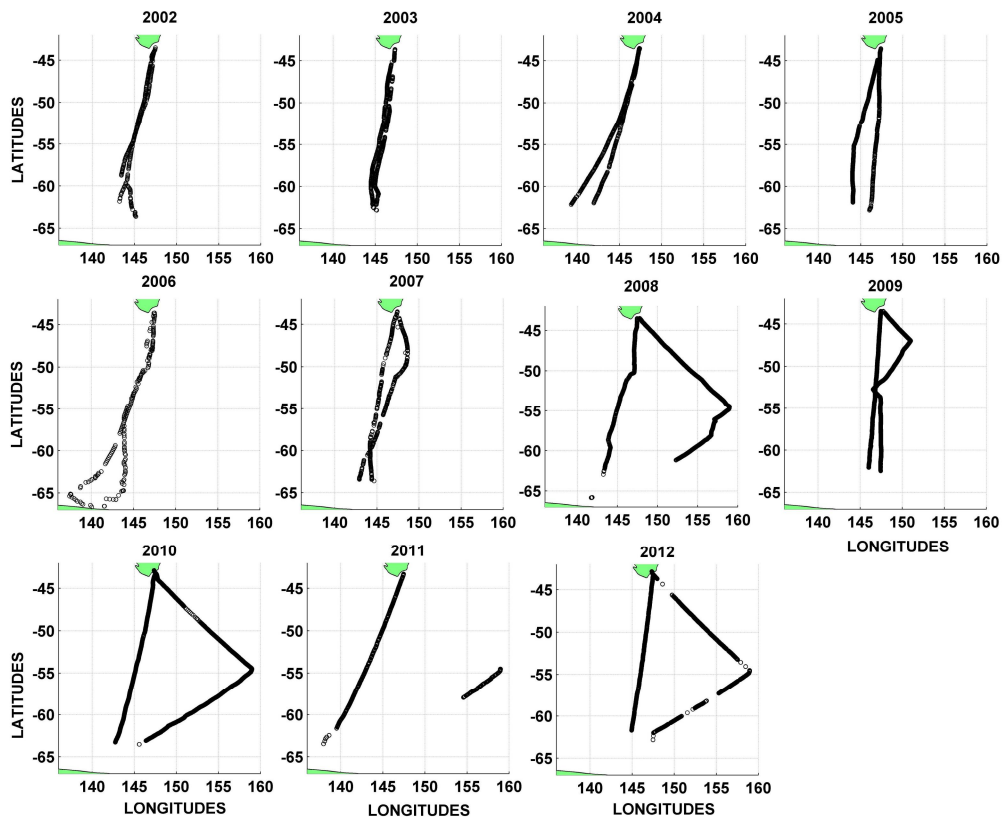


Figure 2: Cruise tracks of the MINERVE project where data are during the first cruise every year.

2 methods

Cruise details

The database from the MINERVE project was utilized for this study. The database was divided into two parts: one to build a model (2002–09) and the second (2010–12) to provide a validation test. There are 14 203 *in situ* data points available to establish the model and 21 395 *in situ* data points for the validation process. The MINERVE project acquired data during three cruises each summer from October to March, between Hobart (Tasmania, Australia, $\sim 43^\circ\text{S}$) and Dumont D’Urville (Adélie Land, Antarctica, $\sim 67^\circ\text{S}$) (Fig. 1). Only data from early summer is considered here, mainly in October and November; see Table 1 for dates and latitudes of each cruise used to establish (part 1) and validate (part 2) the model. The cruise tracks (Fig. 2) every year cross all latitudes between

Tasmania and Antarctica, although the presence of ice in the early summer prevents data collection south of $\sim 63^\circ\text{S}$. Changes of longitude are primarily due to weather conditions; however, in some years (2008, 2010, 2011 and 2012), the ship had to go to Macquarie Island (55°S , 158°E) (Fig. 1).

Data and model

The MINERVE project provides data on surface water parameters from southern Australia to the sea ice edge north of Antarctica. The latitudinal distribution of SST and SSS (Fig. 3) shows cooling and freshening from roughly $\text{SST} = 2^\circ\text{C}$ and $\text{SSS} = 35.1$ near Hobart to $\sim -1.9^\circ\text{C}$ (the freezing point) and 33.9 near 63°S . There is a substantial range in both SST and SSS at each latitude, associated with both the differences in cruise tracks (Fig. 2) and interannual variability.

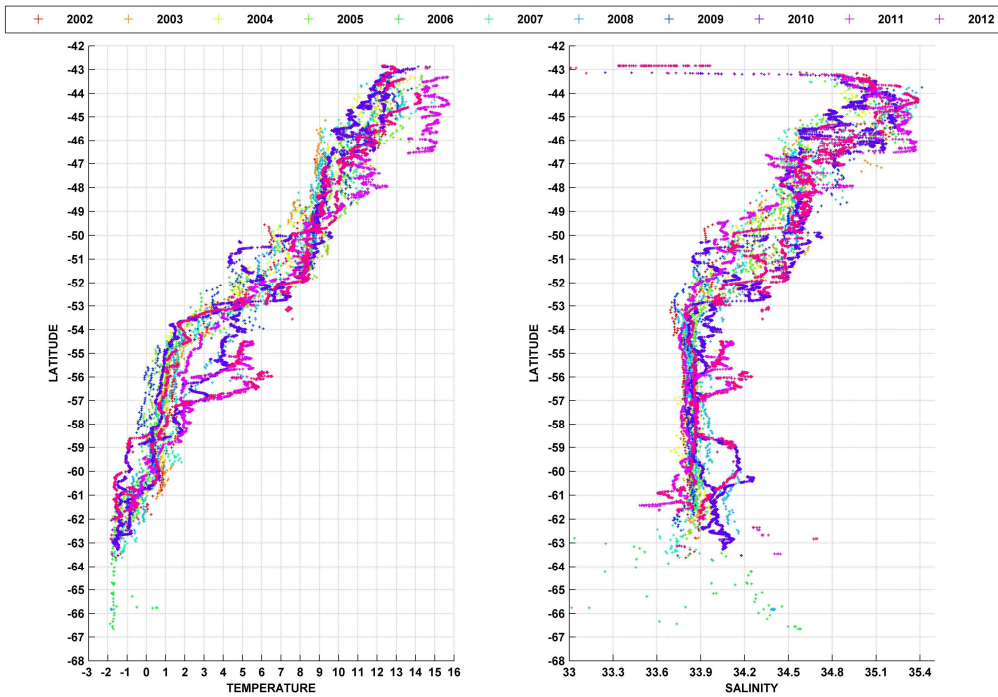


Figure 3: Latitudinal distribution of temperature and salinity in the Southern Ocean south of Australia during the 2002–12 MINERVE project.

The surface fitting toolbox of MATLABM was used to develop a model based on the MINERVE database of SST, SSS and latitude, φ . This toolbox performs multiple linear regressions between these three variables. This method provides good results for our purpose (Moussa et al., 2015). We carried out the analysis for Part 1 of the database (2002–09), setting the degree of φ at 1 and the degree of SST at 2, resulting in the following function:

$$\begin{aligned} \text{SSS}_{\text{mod}} = & a + b \times \varphi + c \times \text{SST} \\ & + d \times \text{Lat} \times \text{SST} + e \times \text{SST}^2 \end{aligned} \quad (1)$$

where the optimized values of coefficients were $a = 33.10000$, $b = -0.01257$, $c = 0.103600$, $d = 0.001929$ and $e = 0.008504$. The fitting surface and the residual plot are shown in Fig. 4.

Three statistical measures were calculated to evaluate the model. Model error at each point in parameter space is defined as $e_i = P_i - O_i$ where e_i is the error, P_i ($i = 1..n$) is the model-predicted value and O_i ($i = 1..n$) is the observed value. The root-mean-square error (RMSE; Willmott and Matsuura (2005)) is defined as:

$$RMSE = \left[\frac{1}{n} \sum_{i=1}^n e_i^2 \right]^{\frac{1}{2}} \quad (2)$$

The mean absolute error (MAE) is based on the absolute value of each error estimate (Willmott and Matsuura, 2005), i.e.

$$MAE = \left[\frac{1}{n} \sum_{i=1}^n |e_i| \right] \quad (3)$$

The mean bias error (MBE) is the simple average of individual error estimates:

$$MBE = \left[\frac{1}{n} \sum_{i=1}^n e_i \right] \quad (4)$$

The MBE indicates whether the average estimate is less than or greater than the observed values. The following measures were obtained for the modelled error in salinity: $MBE = -3.01 \times 10^{-14}$, $MAE = 0.0706$, and $RMSE = 0.0958$. The values for MBE and RMSE indicate that the model, in general, has a salinity accuracy that is comparable with satellite-derived SSS accuracy. However, some large values of e_i are found in the far southern portion of the domain (Fig. 4b).

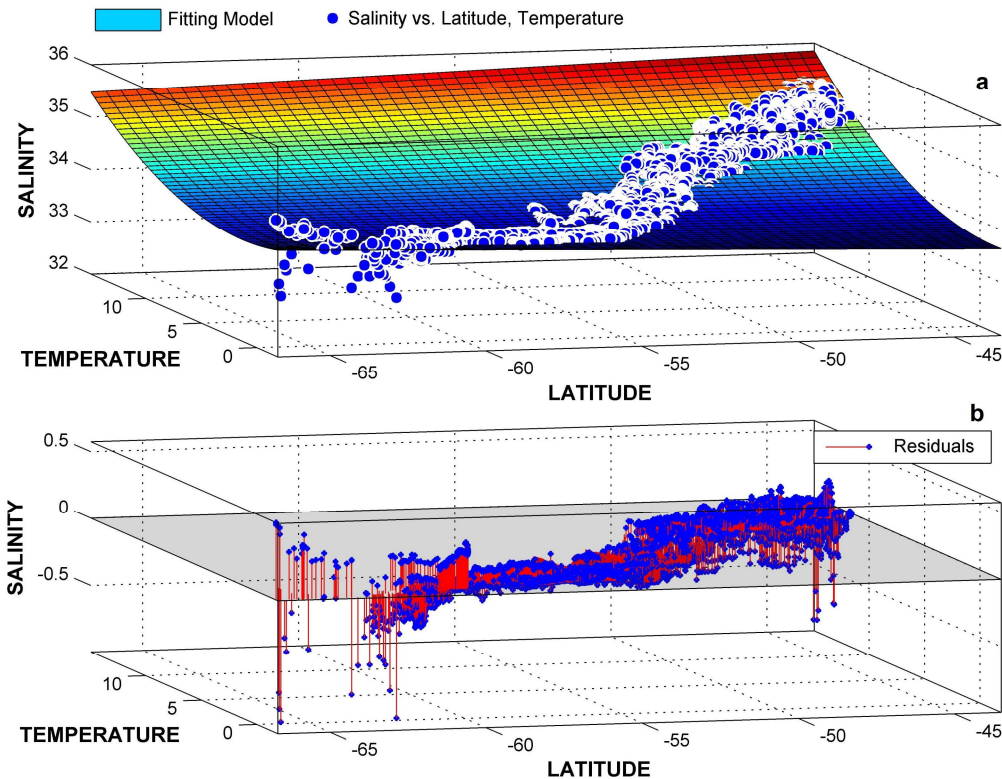


Figure 4: Model suggested by the surface fitting toolbox. a. Dataset points and fitting surface. b. Residuals plot.

3 Results

Satellite-derived maps of SST were obtained from Moderate Resolution Imaging Spectroradiometer (MODIS) on the NASA Aqua satellite. Data are provided as daily averages with a spatial resolution of 4×4 km (<http://oceancolor.gsfc.nasa.gov>). For comparison with the MINERVE database, MODIS SST points corresponding as closely as possible in space and time to our *in situ* datapoints were selected. Our statistical fit model (Eq. 1) was then applied to the satellite-derived SST and latitude φ to estimate SSS, and compare these values with *in situ* SSS data for part 2 data (2010–12) that were not included in the development of the model fit. The loss of satellite-derived SST data due to weather substantially reduces the number

of points for comparison, from 21 395 to 4672 points. There is a mismatch between satellite and *in situ* SST around 48°S and 51°S (Fig. 5a). These errors are probably due to the calibration of the satellite instruments, which is rather complex. At the moment, there are four main calibration methodologies. The aim of these methodologies is to have a good accuracy over the globe; however, calibration can cause significant errors when satellite data are compared with specific local *in situ* data. Note that these errors appear in two fronts (explained in the following paragraph) which increase the difficulty for modelling for the satellite algorithms.

The errors as functions of latitude (Fig. 5) are usually small but with some regions of large outliers. The study area was divided into three regions (Laika et al.,

2009) representing five different hydrographic zones of the Southern Ocean (Rintoul et al., 1997):

1. The Sub-Antarctic Region (SAR) which it is composed of three zones:
 - The Sub-Tropical Zone (STZ) from -43.5°S to -47.5°S.
 - The Sub-Antarctic Zone (SAZ) from -47.5°S to -51.5°S.
 - The Polar Frontal Zone (PFZ) from -51.5°S to -54.5°S.
2. The Permanent Open Ocean Zone (POOZ) from -54.5°S to -60°S.
3. The Seasonal Ice Zone (SIZ) from -60.5°S to -65°S.

Table 2 shows all residuals by zone and for the whole study area. Model errors are larger in the SAZ and STZ than in the other zones. However, the error in the estimation of salinity is correlated with the difference between the *in situ* and the satellite temperature (Fig. 5a). That is, the large errors in SSS for some latitude ranges are caused by errors in satellite-derived temperature rather than in our regression model (Eq. 1).

There is large mean bias between satellite and *in situ* temperature (Fig 5a), with MBE = -0.6952 (Table 2). When outliers are removed, keeping only points where residuals are between +1 and -1.5, the new MBE is -0.2974. This bias is applied to satellite temperature data so that the MBE becomes zero. After testing the model, new residuals for salinity and temperature are smaller; compare Fig. 6 with Fig. 5, and Table 3 with Table 2.

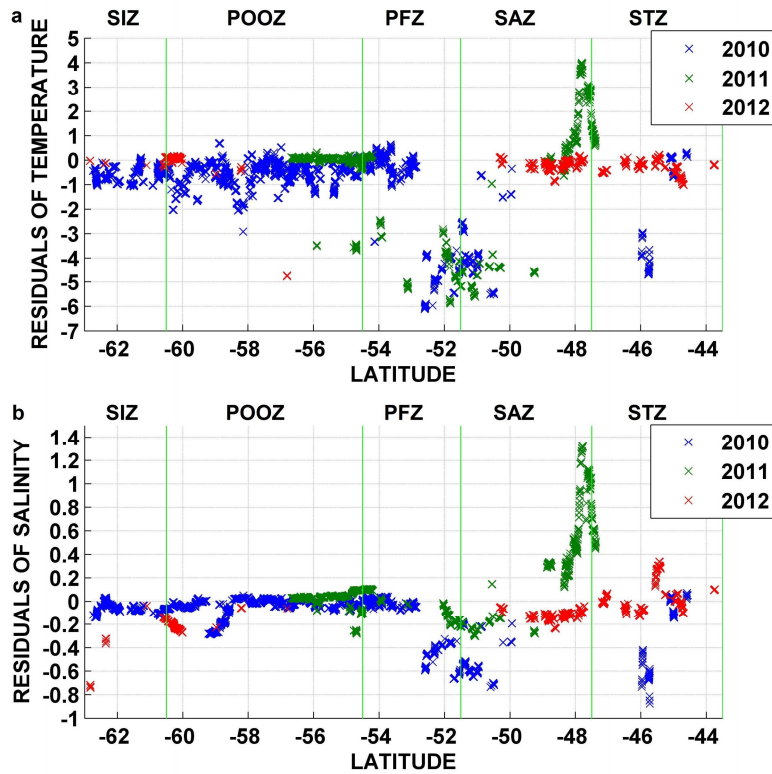


Figure 5: Salinity and temperature residuals (satellite to *in situ* data comparison). Residuals between a. satellite and observed temperature, and b. estimated and observed salinity.

Area	Numbers	Estimated salinity vs <i>in situ</i> salinity			Satellite temperature vs <i>in situ</i> temperature		
		MBE	MAE	RMSE	MBE	MAE	RMSE
STZ	354	0.0001	0.1763	0.2776	-0.4812	0.6958	1.3225
SAZ	678	-0.0223	0.3183	0.4296	-1.0521	1.7228	2.5537
PFZ	614	-0.1092	0.1387	0.2270	-1.7291	1.8288	2.7897
POOZ	2505	-0.0383	0.0577	0.0881	-0.4094	0.4611	0.6962
SIZ	521	-0.0758	0.0781	0.1102	-0.5322	0.5390	0.6064
Whole area	4672	-0.0466	0.1174	0.2119	-0.6952	0.8504	1.5500

Table 2: Model testing results (2010–12). MAE = mean absolute error, MBE = mean bias error, PFZ = Polar Frontal Zone, POOZ = Permanent Open Ocean Zone, RMSE = root-mean-square error, SAR = Sub-Antarctic Region, SAZ = Sub-Antarctic Zone, SIZ = Seasonal Ice Zone, STZ = Sub-Tropical Zone.

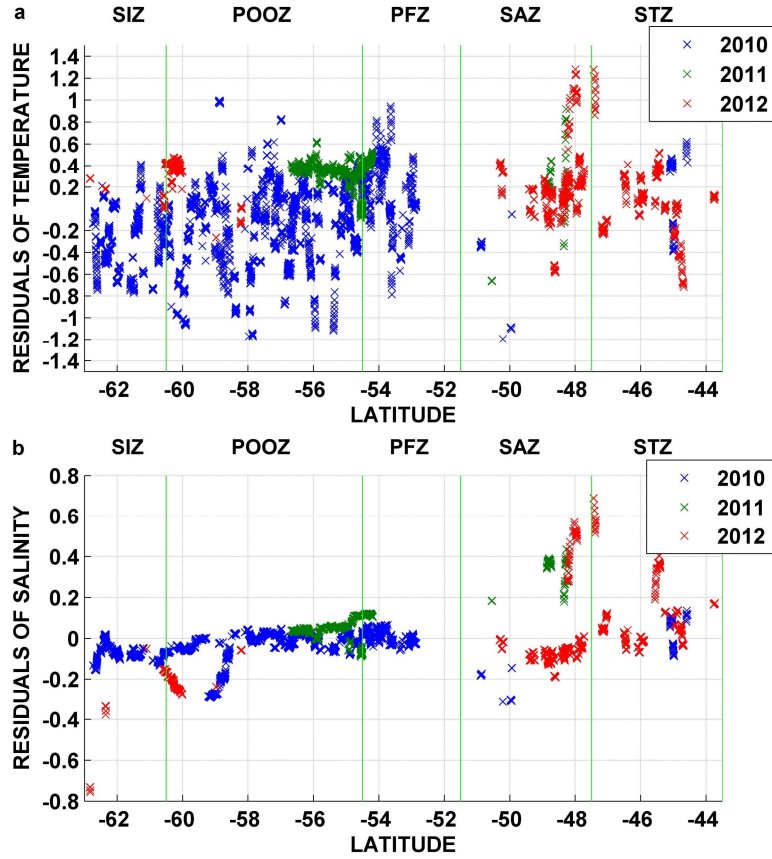


Figure 6: Salinity and temperature residuals after selecting process. Residuals between a. satellite and observed temperature, and b. estimated and observed salinity.

Area	Numbers	Estimated salinity vs <i>in situ</i> salinity			Satellite temperature vs <i>in situ</i> temperature		
		MBE	MAE	RMSE	MBE	MAE	RMSE
STZ	305	0.0969	0.1193	0.1805	0.1089	0.2923	0.3701
SAZ	401	0.0200	0.1480	0.2023	0.1457	0.2901	0.4001
PFZ	381	0.0268	0.0435	0.0567	0.2168	0.3197	0.3745
POOZ	2376	-0.0314	0.0578	0.0890	0.0218	0.3077	0.3837
SIZ	521	-0.0863	0.0872	0.1187	0.2348	0.3054	0.3735
Whole area	3984	-0.0180	0.0741	0.1160	0	0.3058	0.3822

Table 3: Model testing results after the selection process. L = Latitude. MAE = mean absolute error, MBE = mean bias error, PFZ = Polar Frontal Zone, POOZ = Permanent Open Ocean Zone, RMSE = root-mean-square error, SAR = Sub-Antarctic Region, SAZ = Sub-Antarctic Zone, SIZ = Seasonal Ice Zone, STZ = Sub-Tropical Zone.

4 Discussion

Satellite SST maps are available every day (with more or less points provided depending on the climatic conditions), at 4 km spatial resolution. The model we have established (Eq. 1) allows us to map SSS at the same resolution in the early summer in the Southern Ocean south of Australia.

This spatial resolution is much finer than we can obtain from existing satellite salinity sensors on the Aquarius and SMOS satellites (~ 150 km and 50–100 km, respectively). The daily sampling of SST by satellites is much more frequent than either salinity satellite provides. The accuracy of our model, in salinity units (RMSE = 0.116 for the entire region from 43°S to the ice edge near 63°S), is comparable to the cited accuracy of the satellite-derived values; therefore, our model may provide valuable information on SSS for temporal and

spatial scales that cannot be resolved by the satellite-based salinity records. An additional source of error, when modelling SSS based on satellite-derived SST, is that the RMSE for satellite vs *in situ* SST is $\sim 0.38^\circ\text{C}$ (Table 3); based on Eq. (1), this converts to ~ 0.1 in SSS.

As an example of the potential value of our model, Fig. 7 compares a one day modelled SSS (Eq. (1) applied to satellite-derived SST data) with coarser estimates from SMOS and Aquarius. The differences between resolutions are clear. Note that SMOS data are provided as an average of at a minimum of three days. Table 4 shows a pairwise comparison of modelled SSS with SMOS data and with Aquarius data using *in situ* data from 2011 to 2012 (Table I, part 2). The RMSE of our modelled SSS is much smaller than both SMOS and Aquarius data.

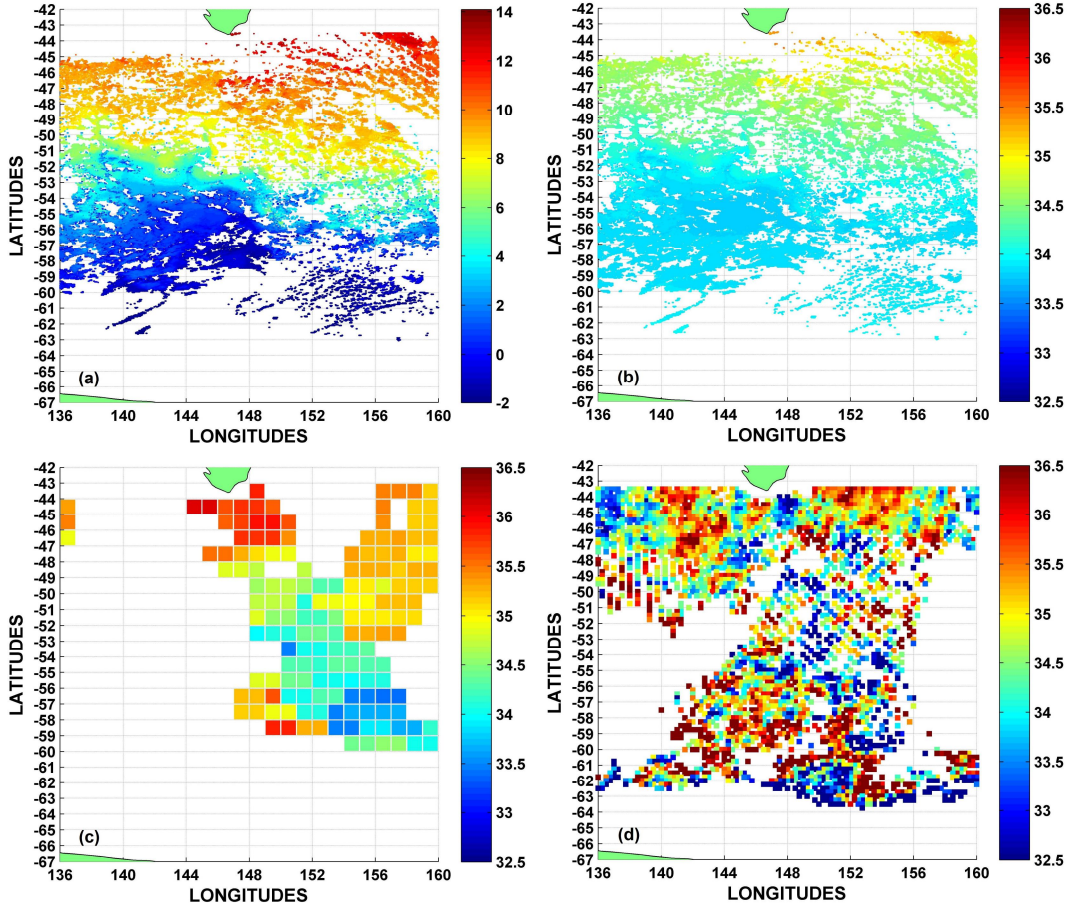


Figure 7: Comparison of modelled sea surface salinity (SSS) using sea surface temperature (SST) with SMOS and Aquarius satellite data. a. Satellite SST in the Antarctic Ocean south of Australia on 25 October 2011. b. Modelled estimated salinity from a. c. Aquarius satellite image from 25 October 2011. d. SMOS three-day average satellite image from 25 October 2011 to 27 October 2011.

	Modelled SSS vs SMOS data	Modelled SSS vs Aquarius data
Numbers	1083	696
RMSE of modelled SSS	0.1841	0.0755
RMSE of corresponding satellite data	1.9781	0.4989

Table 4: Comparison between modelled sea surface salinity (SSS) and two satellite products (SMOS and Aquarius). RMSE = root-mean-square error.

5 Conclusions

We have demonstrated the value of developing algorithms to relate SSS to satellite-derived SST, by providing the ability to map SSS at much finer spatial and temporal scales that can be resolved by existing salinity satellites, Aquarius (stopped on 17 June 2015) and SMOS. Furthermore, as these satellites have only been operating for a few years, the algorithm allows the extension of maps of SSS into time ranges for which no satellite maps of SSS are available. This approach can be further developed to determine SSS throughout the year, and to contribute to a broader range of applications including determination of surface density, air-sea exchanges, and estimation and analysis of air-sea CO₂ fluxes.

Collecting more in situ data is the best way to improve our models; therefore, it is important that the MINERVE project remains active.

6 Acknowledgements

I would like to express my very great appreciation to all the persons who built and continue to provide accurate databases for the MINERVE project, and the IPEV (Institut polaire français Paul Émile Victor) which provides the logistics framework, human resources, and technical and financial resources to the development of French research in the Polar Regions. I would like to thank NASA for providing satellite temperature images. I am particularly grateful for the MESRS (Ministère de l'Enseignement Supérieur et de la Recherche Scientifique) from Algeria which funded my PhD and gave me the opportunity to undertake research at the IMAGES laboratory. We are also grateful to the reviewers for their comments on the manuscript.

7 Authors contributions

Mrs H. Moussa was involved in the conception and writing of the programs for the study. Dr F. Touratier supervised this PhD research and assisted with writing and revising this paper. Professor C. Goyet assisted with analysis of the data and provided ideas for the selection process which enabled the achievement of better results, and also participated in revising the article. Professor A. Poisson participated on the MINERVE project where we found our data, and also assisted with revising the article.

References

- Lagerloef, G., Boutin, J., Chao, Y., Delcroix, T., Font, J., Niiler, P., Reul, N., Riser, S. J., Schmitt, R., Stammer, D., and others (2010). Resolving the global surface salinity field and variations by blending satellite and in situ observations. In *OceanObs 09*, pages 587–597. European Space Agency.
- Laïka, H., Goyet, C., Vouve, F., Poisson, A., and Touratier, F. (2009). Interannual properties of the CO₂ system in the Southern Ocean south of Australia. *Antarctic Science*, 21(06):663.
- Le Vinc, D., Lagerloef, G., Colomb, F., Yueh, S., and Pellerano, F. (2007). Aquarius: An Instrument to Monitor Sea Surface Salinity From Space. *IEEE Transactions on Geoscience and Remote Sensing*, 45(7):2040–2050.
- Lenton, A., Tilbrook, B., Law, R. M., Bakker, D., Doney, S. C., Gruber, N., Ishii, M., Hoppema, M., Lovenduski, N. S., Matear, R. J., McNeil, B. I., Metzl, N., Mikaloff Fletcher, S. E., Monteiro, P. M. S., Rödenbeck, C., Sweeney, C., and Takahashi, T. (2013). Sea-air CO₂ fluxes in the Southern Ocean for the period 1990–2009. *Biogeosciences*, 10(6):4037–4054.
- Michel, S. (2006). *Téledétection de la salinité à la surface des océans : variabilité de la salinité de surface d'après un modèle global de couche mélangée océanique*. Paris 7.
- Moussa, H., Benallal, M. A., Goyet, C., Lefevre, N., EL Jai, M. C., Guglielmi, V., and Touratier, F. (2015). A comparison of Multiple Non-linear regression and neural network techniques for sea surface salinity estimation in the tropical Atlantic ocean based on satellite data. *ESAIM: Proceedings and Surveys*, 49:65–77.
- Rintoul, S. R., Donguy, J. R., and Roemmich, D. H. (1997). Seasonal evolution of upper ocean thermal structure between Tasmania and Antarctica. *Deep Sea Research Part I: Oceanographic Research Papers*, 44(7):1185–1202.
- Willmott, C. J. and Matsuura, K. (2005). Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Climate research*, 30(1):79.

E.4 Publication 4 : Satellite-derived CO₂ flux in the surface seawater of the austral ocean south of Australia

Satellite-derived CO₂ flux in the surface seawater of the austral ocean south of Australia

M.A. BENALLAL^{1,2}, H. MOUSSA^{1,2}, J.M. Lencina-Avila^{1,2}, F. TOURATIER^{1,2},
C. GOYET^{1,2}, N. POISSON^{1,3}, A. POISSON^{1,4}

International Journal of Remote Sensing (draft pre-refereeing)

¹IMAGES ESPACE-DEV, Université de Perpignan Via Domitia, 52 Avenue Paul Alduy, 66860 Perpignan Cedex 9, France.

²ESPACE-DEV, UG UA UR UM IRD, Maison de la télédétection, 500 Rue Jean-François Breton, 34093 Montpellier Cedex 5, France.

³Agence de l'Environnement et de la Maîtrise de l'Énergie (ADEME), Paris, France.

⁴Laboratoire d'Océanographie et du Climat : Expérimentations et Approches Numériques (LOCEAN/IPSL), Université Pierre et Marie Curie, Paris, France.

Mohamed.benallal@univ-perp.fr

Abstract

A step by step algorithm for air-sea CO₂ flux ($F(\text{CO}_2)$) calculation from satellite parameters is presented in this study. Parameters used for $F(\text{CO}_2)$ calculation are : (1) sea surface temperature (SST) and chlorophyll-*a* (chl-*a*) from Moderate Resolution Imaging Spectroradiometer Aqua (MODIS Aqua), (2) sea surface salinity (SSS) estimated from MODIS Aqua SST using multiple linear regression (MLR), (3) seawater CO₂ fugacity ($f(\text{CO}_2^{\text{sw}})$) estimated by MODIS Aqua SST and chl-*a* using feedforward neural networks (FNN), (4) atmospheric CO₂ fugacity ($f(\text{CO}_2^{\text{atm}})$) from the Cape Grim station and (5) wind speed (WS) from the Quick and Advanced Scatterometer (QSCAT and ASCAT). *In situ* data provided by (1) the SURVOSTRAL (SURveillance de l'Océan AuSTRAL - Monitoring the Southern Ocean) project, (2) IMOS (The Integrated Marine Observed System) project and (3) MINERVE (Mesures à l'INterface Eau-air de la Variabilité des Échanges de CO₂) projet collected on the research vessel (RV) L'Astrolabe, are used to establish and validate the models. These models are then tested using remote sensing data. This work focus on the southern ocean from the south of Australia to the Antarctica coasts (between 43.5 °S and 67 °S), on the period of spring and austral summer (from October to March, between 2002/2003 and 2014/2015). In each step, the result's precision of experiments was determined by the Root Mean Square Error (RMSE). Results show (1) an improvement of satellite SSS estimation with a precision of ± 0.16 using SST and latitude, (2) an estimation of $f(\text{CO}_2^{\text{sw}})$ by satellite data with a good accuracy of ± 9.45 μatm (0.96 Pa) and (3) a calculation of $F(\text{CO}_2)$ using satellite data with a global RMSE of about ± 3 $\text{mmol CO}_2 \text{ m}^{-2} \text{ d}^{-1}$. The interpolated $F(\text{CO}_2)$ shows that, in the period of austral spring and summer, this region absorbs the atmospheric CO₂, and becomes a stronger sink of CO₂ throughout the years (from an overall average absorption of about 2 $\text{mmol CO}_2 \text{ m}^{-2} \text{ d}^{-1}$ in 2002/2003 to about 7 $\text{mmol CO}_2 \text{ m}^{-2} \text{ d}^{-1}$ in 2014/2015).

Submitted 18 May 2016, In press.

Key words: Southern Ocean; CO₂ flux; MODIS Aqua; Multiple Linear Regression; Neural networks.

1 Introduction

It has long been recognized that the ocean represents a sink of carbon dioxide for the atmosphere. About 30–40% of the anthropogenic carbon released into the atmosphere is absorbed by world oceans (Brewer, 1978; Chen and Millero, 1979; Siegenthaler and Sarmiento, 1993; Feely et al., 2001; Sarmiento and Gruber, 2002; Takahashi et al., 2002; Sabine et al., 2004; Le Quéré et al., 2014, 2015). Human activities are the much responsible of this anthropogenic CO₂ increase. The atmospheric CO₂ fugacity ($f(\text{CO}_2^{\text{atm}})$) has increased from 277 μatm in 1750 (Pre-industrial period) to 407 μatm in April 2016 (Dlugokencky and Tans, 2016; Le Quéré et al., 2015) (according to Bureau International Des Poids Et Mesures (1955), 1 atm = 101325

Pa). Understanding the present-day carbon budget is directly related to determining the role of the Southern Ocean. Sallée et al. (2012) reported that more than 40% of the anthropogenic CO₂ in the ocean has entered south of 40 °S. Yet an accurate and global estimation of the air-sea CO₂ flux ($F(\text{CO}_2)$) in this region, is crucial to study effects on climate change and especially on ocean acidification.

Until a few years ago, the estimation of $F(\text{CO}_2)$ has been obtained using ship board measurements data (*in situ* data). Even if these data are considered as references to provide models (due to their good precision), these data are limited in spatial and temporal coverage. Satellite data seems to be a good alternative since they have proved their worth in several previous

works on estimating seawater CO₂ fugacity ($f(\text{CO}_2^{\text{sw}})$) and $F(\text{CO}_2)$ over many areas of the ocean (Stephens et al., 1995; McNeil et al., 2007; Zhu et al., 2009; Moussa et al., 2016). The southern ocean, especially the Antarctic zone, is a very difficult area to access by oceanographic research vessels (RV) (far away from any coast, storms, strong winds, ice, ...). This causes a lack of *in situ* observations and a restriction for research cruises during the period from october to march (austral spring and summer). Satellites are more effective since they provide data every day of the year and throughout the seasons, moreover they cover much larger areas.

This work is based on databases from three projects; (1) the SURVOSTRAL project (SURVeillance de l'Océan AuSTRAL - Monitoring the Southern Ocean), which aims to provide temperature and salinity informations in the upper ocean, in order to use these valuable data for monitoring climate variations and validating computer simulations and projections for the Southern ocean (<http://www.legos.obs-mip.fr/recherches/>

projets-en-cours/projet-survostral), (2) IMOS (The Integrated Marine Observed System), which is designed to be a fully-integrated national system, observing at ocean-basin and regional scales and covering physical, chemical and biological variables (<http://imos.aodn.org.au/>) and (3) MINERVE (Mesures à l'INterface Eau-air de la Variabilité des Échanges de CO₂- measurements in the sea-air interface of the CO₂ exchange variability) projet which aims to observe and understand the seasonal and inter-annual variability of the partial pressure of CO₂ ($p(\text{CO}_2^{\text{sw}})$) in surface waters using hydrological and bio-geo-chemical data in Antarctic Ocean South of Australia (<http://caraus.ipsl.jussieu.fr/>). These programs provide data of three cruises during each austral summer, between Hobart (Tasmania, Australia, 42.88 °S, 147.33 °E) and Dumont D'Urville (Terre Adelie, Antarctica, 66.40 °S, 140.00 °E) (Figure 1(a)). Cruises are conducted on board the RV L'Astrolabe, for which logistics routes are presented in Figure 1(b) (SURVOSTRAL data was used to plot these cruise tracks)

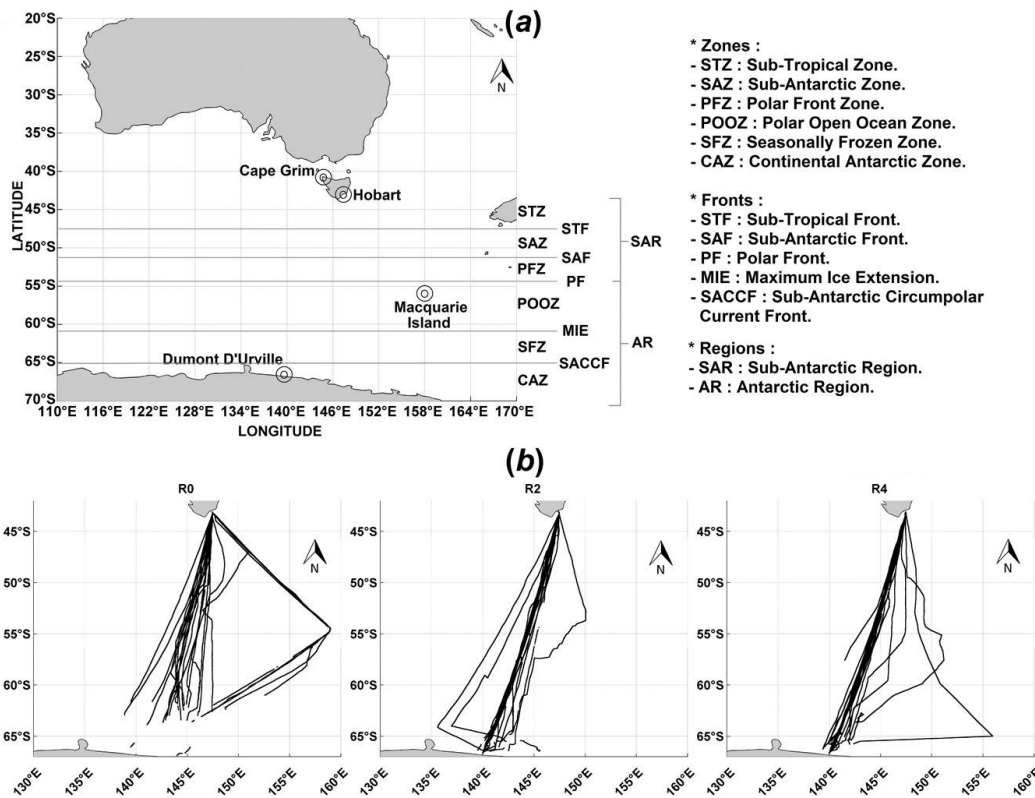


Figure 1: Regional map of the studied area : (a) southern ocean regions, zones and fronts. (b) cruise tracks of the RV L'Astrolabe (where data are available) (R0, R2 and R4 are the three rotations)

In this paper, we present a set of methods detailed in a general algorithm in Figure 2. We start by analyzing all required parameters, after that we establish models, map $F(\text{CO}_2)$, and finally we discuss the ob-

tained results. The three main steps of this algorithm are the following : (1) modelling and estimating sea surface salinity (SSS¹) ((i) and (ii) respectively in Figure 2), (2) modelling and estimating $f(\text{CO}_2^{\text{sw}})$ ((iii)

¹According to IAPSO Working Group on Symbols, Units, and Nomenclature in Physical Oceanography. (1985), SSS has no unit.

and (iv) respectively in Figure 2) and (3) use all the resulting parameters to estimate $F(\text{CO}_2)$ and then, compare the resulting data with $F(\text{CO}_2)$ calculated using *in situ* parameters ((v) and (vi) respectively in Figure 2).

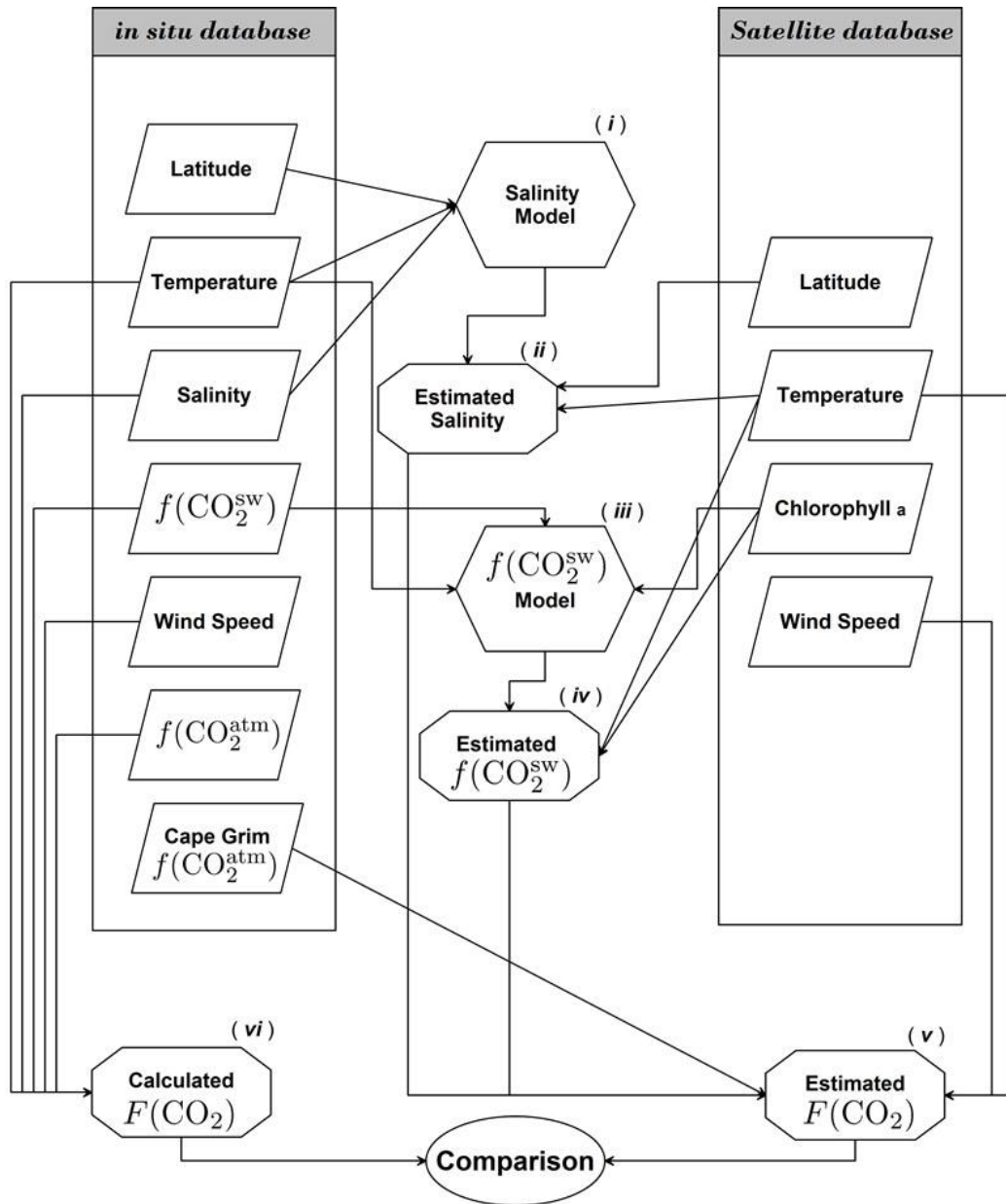


Figure 2: The general followed algorithm : (i) : SSS model construction using *in situ* data, (ii) : SSS estimation with satellite data, (iii) : $f(\text{CO}_2^{\text{sw}})$ model construction using *in situ* data, (iv) : $f(\text{CO}_2^{\text{sw}})$ estimation with satellite data, (v) : Calculate $F(\text{CO}_2)$ using satellite data and estimated parameters, (vi) : Calculate $F(\text{CO}_2)$ using *in situ* data.

2 Data and Methods

The difference between $f(\text{CO}_2^{\text{sw}})$ and $f(\text{CO}_2^{\text{atm}})$ (expressed in μatm) indicates the rate of CO_2 exchange in the air-sea interface and its direction. This exchange is represented by the CO_2 flux ($F(\text{CO}_2)$, expressed in $\text{mmol CO}_2 \text{ m}^{-2} \text{ d}^{-1}$) which can be computed as follows :

$$F(\text{CO}_2) = \beta_0 \times k \times (f(\text{CO}_2^{\text{sw}}) - f(\text{CO}_2^{\text{atm}})) \quad (1)$$

wherein β_0 represents the solubility coefficient of CO_2 in seawater ($\text{mol l}^{-1} \text{ atm}^{-1}$), determined according to Weiss (1974) as a function of sea surface temperature (SST; expressed in $^\circ\text{C}$) and SSS. k represents the gas transfer velocity (cm h^{-1}), calculated as function of wind speed (WS; u expressed in m s^{-1}) and the Schmidt number Sc (no unit) of CO_2 in sea water. Positive $F(\text{CO}_2)$ value indicates oceanic source of CO_2 and negative value indicates an oceanic CO_2 sink, this

sign is determined by the term $f(\text{CO}_2^{\text{sw}}) - f(\text{CO}_2^{\text{atm}})$ commonly noted as $\Delta f\text{CO}_2$. In order to calculate $F(\text{CO}_2)$, the following parameters are then required : SST, SSS, $f(\text{CO}_2^{\text{sw}})$, $f(\text{CO}_2^{\text{atm}})$ and WS.

2.1 *In situ* data

All data collected for this study are presented in this section. *In situ* data used in this paper are available as follow :

- The SURVOSTRAL project provides data of SST and SSS from October 2002/March 2003 to October 2014/March 2015. Distribution and interannual variabilities of these two parameters was presented in Chaigneau and Morrow (2002), Morrow et al. (2008) and Morrow and Kestenare (2014).
- The IMOS program provides of SST, SSS, $f(\text{CO}_2^{\text{sw}})$, $f(\text{CO}_2^{\text{atm}})$ and WS data from October 2008/March 2009 to October 2012/March 2013 (databases are available for download in http://cdiac.ornl.gov/ftp/oceans/VOS_L_Astrolabe) (Tilbrook et al., 2013).
- The MINERVE project provides data of SST, SSS and $f(\text{CO}_2^{\text{sw}})$ for some rotations (R0 for 2005/2006, R2 for 2004/2005 and 2006/2007, R4 for 2004/2005).

R0 rotations are done in the austral spring, R4 in the austral summer and R2 in the transition between these two seasons. For each rotation, a first model was developed to interpolate SSS (Section 2.4) and a second one for interpolating $f(\text{CO}_2^{\text{sw}})$ (Section 2.5).

In addition to this databases, data of $f(\text{CO}_2^{\text{atm}})$ from the Cape Grim station (40.68 °S 144.68 °E, Figure 1(a)) was used for estimating $F(\text{CO}_2)$. The Cape Grim station is located just south of the isolated north-west tip of Tasmania (Figure 1(a)). The sampled air arrives at Cape Grim after long trajectories over the Southern ocean, under conditions described as "baseline". This baseline air is representative of a large area of the Southern hemisphere. Cape Grim is an important site because it is unaffected by regional pollution sources (there are no nearby cities or industry that would contaminate the air quality)(Cape Grim Greenhouse Gas Data, greenhouse-gases, <http://www.csiro.au/>).

2.2 Satellite data

2.2.1 MODIS Aqua

The Moderate Resolution Imaging Spectro-radiometer Aqua (MODIS Aqua), is a part of the Aqua Earth-Observing Satellite Mission (<http://aqua.nasa.gov/>) which is a part of the Earth Observing System (EOS)(<http://eosps.nasa.gov/>). The purpose of this major program is to provide the data collection on the water cycle of the earth, including evaporation from the oceans, water vapor in the atmosphere,

clouds, precipitation, soil moisture, sea ice, ice, earth and snow cover on the land and ice (NASA Goddard Space Flight Center, Ocean Ecology Laboratory, Ocean Biology Processing Group, 2014).

We used the Level 3 Standard Mapped Image (SMI) product of SST and chlorophyll-*a* (chl-*a*). These products are a 8640×4320 pixels images which correspond to a resolution of $4 \text{ km} \times 4 \text{ km}$. All MODIS Aqua products are available at different time scales : one day image, 8-day composite image, one month, one season, and one year. Additional time scales are available for some periods.

2.2.2 QSCAT and ASCAT

QSCAT and ASCAT are diminutives of Quick and Advanced Scatterometer. The two satellites provide WS data near the ocean surface (about 10 m height). They are complementary in term of time; QSCAT was launched in June 1999 and operated until November 2009, and ASCAT was launched on the October 2006 and became fully operational since May 2007 (up until now).

Data provided by the two instruments are available for users in different time scales (one day, one week, one month, and one year) with a spatial resolution of $25 \text{ km} \times 25 \text{ km}$ which correspond to a 1440×720 pixels images (<http://www.remss.com/>).

2.3 Methods

2.3.1 Multiple Linear Regression (MLR)

The aim of the MLR method is to find coefficients ($\beta_0, \beta_1, \dots, \beta_n$) to determine a relationship between an output variable (y) and two or more input variables (x_1, x_2, \dots, x_n). The most popular method of Least Squares is used to find these coefficients by minimizing the summed square of residuals (differences between observations and regression line). We note this equation as :

$$y = \beta_0 + \beta_1 \times x_1 + \beta_2 \times x_2 + \dots + \beta_n \times x_n = \beta_0 + \sum_{i=1}^n \beta_i \times x_i \quad (2)$$

Where n is the number of input observations.

2.3.2 Feedforward Neural Network (FNN)

The FNN is the simplest type of neural networks (NNs) which is composed of three layers or more : one for the input, one for the output and one or more for the the hidden layers.

Input neurons correspond to the input parameter(s), the output neurons numbers correspond to the target output(s), the hidden layers and neurons numbers needs to be set according to the complexity of the problem. Several researchers proved that the three-layers FNN (with one hidden layer) is sufficient for non-linear function approximation and is represented in the class of universal approximators (Irie

and Miyake, 1988; Hornik et al., 1989; De Villiers and Barnard, 1993; Tamura and Tateishi, 1997), however, the number of hidden neurons must be chosen by increasing it and testing every time.

2.4 Salinity using satellite data

Instead of using (the not so accurate) SSS provided by satellites, we choose to follow the approach used in Benallal et al. (2016), Moussa et al. (2015) and Moussa et al. (2016). In this approach, SSS models are constructed using the MLR method, based on *in situ* latitude and SST, satellite data of latitude and SST from MODIS Aqua are then used to estimate SSS. This gives (1) more precision in term of resolution compared with SSS provided by satellite as SMOS (Soil Moisture and Ocean Salinity, <https://smos.cnes.fr/>) and AQUARIUS (<http://oceancolor.gsfc.nasa.gov/cms/data/aquarius>) (25 km \times 25 km and 150 km \times 150 km respectively) and (2) MODIS Aqua time coverage (since July 2002) is larger than that of SMOS (since November 2009) and AQUARIUS (between June 2011 and June 2015).

The SSS models were developed using the SURVOSTRAL database (section 2.1). It contains 502844 points which are divided into two subsets to have as much data in each subset : the first one contains 234165 points, and it is used to establish the models (75703 for R0, 86949 for R2 and 71713 for R4, from 2002/2003 to 2007/2008), the second one contains 291298 points, and is used for the validation processes (94993 for R0, 94532 for R2 and 101773 for R4, from 2008/2009 to 2014/2015). Finally, the models were tested using satellite SST data provided by MODIS Aqua, which correspond to the validation data points (in terms of time, latitude and longitude). In order to provide SSS with good accuracy, for each rotation, two experiments based on two types of MODIS Aqua data were performed :

- E1 : One day SST data.
- E2 : 8-days composite SST data. Each day of this image is represented by the 8-days composite image.

To evaluate each experiment, four error parameters were calculated : (1) the Mean Bias Error (MBE), (2) the Mean Absolute Error (MAE), (3) the Root Mean Square Error (RMSE) and (4) the Percentage of Lost Data (PLD) when extracting satellite data points that correspond to the *in situ* data (see appendix A for more details).

All these experiments are presented in the general algorithm of Figure 2 as "salinity model". As we see in this Figure, we used *in situ* data of latitude, SST and SSS to develop SSS models for each rotation. After that, we used latitude and SST from MODIS Aqua to estimate SSS. Results will be presented and discussed in Section 3.1.

2.5 $f(\text{CO}_2^{\text{sw}})$ using satellite data

After estimating SSS, the next challenge was to estimate $f(\text{CO}_2^{\text{sw}})$ using satellite data. We use $f(\text{CO}_2^{\text{sw}})$ instead of $p(\text{CO}_2^{\text{sw}})$ (partial pressure of CO_2^{sw}) because the CO_2^{sw} is a non-perfect gas (a perfect gas is a theoretical gas composed of many randomly moving particles that do not interact). The ratio between the partial pressure and the fugacity of CO_2 varies in the interval $[0.996 - 0.997]$, with a SST in the range $[0^\circ\text{C} - 30^\circ\text{C}]$ and normal atmospheric pressure of 1 atm. For example, a $p(\text{CO}_2^{\text{sw}})$ of 360 μatm at 25 $^\circ\text{C}$ represents a difference of 1 μatm with $f(\text{CO}_2^{\text{sw}})$, this difference is negligible given the associated error of the measurement (2 to 5 μatm)(?).

In order to estimate $f(\text{CO}_2^{\text{sw}})$, we have to determine which parameters affect the $f(\text{CO}_2^{\text{sw}})$, and to choose a method to develop the models. Ocean properties are driven by ocean circulation and $f(\text{CO}_2^{\text{sw}})$ depends mainly on solubility and marine biology. The solubility is driven by SST, earlier works studied the SST- $f(\text{CO}_2^{\text{sw}})$ relationship (Gordon and Jones, 1973; Copin-Montegut, 1988; Goyet and Poisson, 1989; Goyet et al., 1993). This relationship was also used to map $f(\text{CO}_2^{\text{sw}})$ in many regions of the globe (Stephens et al., 1995; Lefèvre and Taylor, 2002; Inoue et al., 2003; Olsen et al., 2004; Lefèvre et al., 2005). Marine biology is also a parameter that modifies the total inorganic carbon in the ocean chemistry and causes direct variations of $f(\text{CO}_2^{\text{sw}})$ (mainly through photosynthesis, respiration and decomposition processes). Laika et al. (2009) investigated the relative importance of biological and SST effects on $f(\text{CO}_2^{\text{sw}})$ variation in this region, and found that distribution of $f(\text{CO}_2^{\text{sw}})$ is mainly associated with SST and chl-*a*.

The frequently used method for estimating $f(\text{CO}_2^{\text{sw}})$ is the MLR (Section 2.3.1), and recently the neural networks (NN) techniques. The MLR method was mainly used to provide empirical equation to estimate and map $f(\text{CO}_2^{\text{sw}})$ (for example Jamet et al. (2007) in the North Atlantic and Chierici et al. (2012) in the high-latitude Pacific sector of the Southern ocean). NN techniques are generally used to solve different problems, like Self-Organisation Map (SOM) by Kohonen (2001) for analysing and clustering high dimensional data space (Flexer, 1999; Xu and Wunsch, 2005; Lefèvre et al., 2005; Friedrich and Oschlies, 2009; Telszewski et al., 2009; Nakaoka et al., 2013), and FNN (Section 2.3.2) which has been widely used for non-linear regression, time-series prediction (Barreto and Araújo, 2002; Barreto, 2007) and $f(\text{CO}_2^{\text{sw}})$ interpolation (Landschützer et al., 2013; Zeng et al., 2014; Moussa et al., 2016). Lefèvre et al. (2005) and Moussa et al. (2016) reported that NN techniques predict $f(\text{CO}_2^{\text{sw}})$ with a better accuracy than MLR. In this work we choose to use SST and chl-*a* as parameters and FNN as method.

The aim is to develop $f(\text{CO}_2^{\text{sw}})$ model using *in situ* data, and after that, use satellite data to test it and

map the estimated $f(\text{CO}_2^{\text{sw}})$. We used the IMOS *in situ* database for this experiment. This step is presented as " $f(\text{CO}_2^{\text{sw}})$ model" in Figure 2. To estimate $f(\text{CO}_2^{\text{sw}})$, we performed a FNN for each rotation. The input parameters are SST and chl-*a* (note that 8-days composite images of chl-*a* from MODIS Aqua were used for both training and testing steps; no *in situ* chl-*a* data were available for this work), and the output

parameter is the $f(\text{CO}_2^{\text{sw}})$ (Section 2.3.2). This FNN was constructed with two input neurons for SST and chl-*a*, one output neuron for $f(\text{CO}_2^{\text{sw}})$, and one hidden layer with number of neurons that increase from 1 to 50. The general algorithm followed in order to choose the best FNN is presented in Figure 3(a).

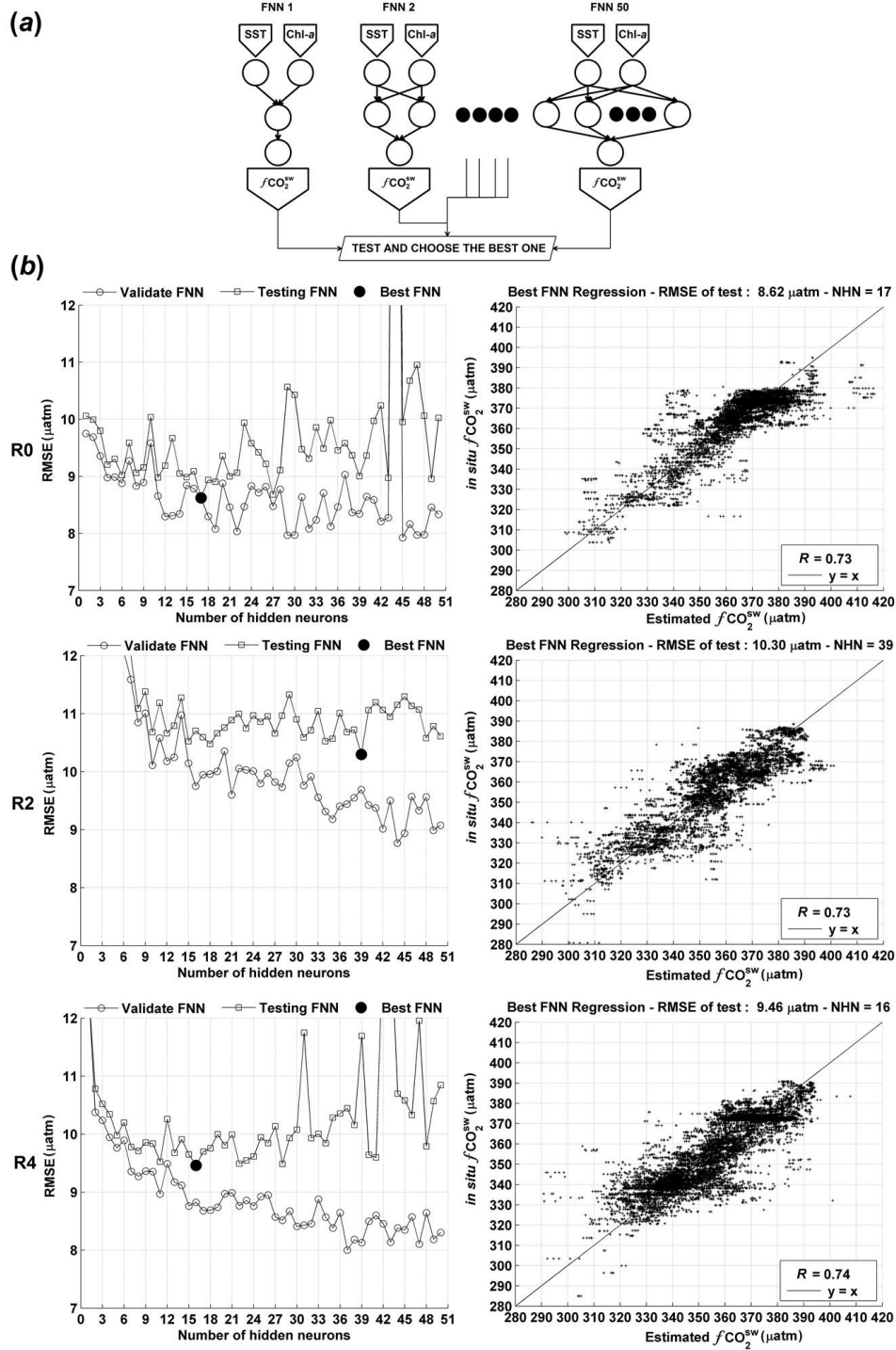


Figure 3: $f(\text{CO}_2^{\text{sw}})$ modelling : (a) the followed algorithm for choosing the best FNN. (b) results of validation and test step for each rotation (left) and the regression of $f(\text{CO}_2^{\text{sw}})$ for the best FNN (right). R represents the coefficient of correlation.

To find the best FNN, for each rotation, fifty FNNs were performed corresponding to FNN with one to fifty hidden neurons respectively. Firstly, we begin by dividing the database into two uniform datasets which have the same time and space scales (from 2008/2009 to 2012/2013 for both datasets; 15864 for R0, 14034 for R2 and 24413 for R4). Then, for each FNN, we pass through 3 steps :

1. Train the FNN using the first dataset (we name it training dataset).
2. Validate the FNN using the second dataset (we name it validation dataset). We also performed a second *in situ* validation with MINERVE data points. Note that these data numbers are lower than IMOS (2452 for R0, 7508 for R2 and 2490 for R4).
3. Test the FNN with satellite data. All satellite data points correspond to the validation database points (in term of time, latitude and longitude).

The best FNN is chosen among the fifty FNNs according to the minimum of RMSE value in the test step. Results of $f(\text{CO}_2^w)$ modelling are presented in section 3.2.

		MBE			MAE			RMSE			PLD (%)		
		R0	R2	R4	R0	R2	R4	R0	R2	R4	R0	R2	R4
E1	V.	-0.04	-0.06	-0.02	0.12	0.14	0.14	0.18	0.19	0.20	94	89	93
	G.		-0.03			0.13			0.19			92	
E2	V.	-0.05	-0.02	-0.02	0.13	0.11	0.11	0.17	0.17	0.15	57	44	51
	G.		-0.03			0.12			0.16			51	

Table 1: Salinity modelling results. E1: One day satellite data as distributed online by the NASA, E2: 8-days composite satellite data as distributed online by the NASA, V: Value, G: Global result considering the three rotations.

A general RMSE of 0.16 was found for the three SSS model estimation. In order to confirm models accuracy, we performed a second test in which we provided a pairwise comparison between the modelled SSS ($\text{SSS}_{\text{Model}}$) starting with MODIS Aqua SST and the two SSS products from SMOS (SSS_{SMOS}) and AQUARIUS ($\text{SSS}_{\text{AQUARIUS}}$). Note that we performed an intersection of available data between MODIS Aqua vs. SMOS and MODIS Aqua vs. AQUARIUS data (a dataset between 2011/2012 and 2014/2015 was found respectively). Results are presented in Table 2, these

3 Results and Discussions

3.1 Salinity

The results of the two experiments of the SSS estimation step (Section 2.4) are shown in Table 1. To choose the best type of data we compared the two experiments by their four error parameters. We choose the Experiment number 2 (E2, Table 1) as the best one because we noted that it represents the best compromise between the four error parameters:

- The MBE inform about the model shift and in our case, it is close to zero (-0.03 and in the same range as in E1).
- The MAE, which inform about the mean level of errors, is at its minimum in E2 (0.12).
- The RMSE, which is the most important error parameter, gives the model precision and is also at its minimum in E2 (0.16). Even if we look at the three rotations one by one, the RMSE of E2 in each rotation still better than E1.
- We see clearly in E1 that the use of one day satellite data causes a important PLD which involve a lack of comparable data (92% of loss, which is not representative of all spatio-temporal variations of SSS over all the studied area). The 51% of loosing data in E2 seems to be an acceptable loss that don't have a significant effect on the model and make satellite data available in all periods and areas as well as for *in situ* data.

results confirm that our $\text{SSS}_{\text{Model}}$ gives, for the whole season, a lower RMSE than SSS provided by satellite: (1) 0.16 and 1.27 respectively for $\text{SSS}_{\text{Model}}$ vs. SSS_{SMOS} and (2) 0.13 and 0.48 respectively for $\text{SSS}_{\text{Model}}$ vs. $\text{SSS}_{\text{AQUARIUS}}$. These results show that models provide SSS with a better accuracy than that with satellite in this region, and represent a good alternative to get SSS maps from satellite SST. Figure 4(b) shows an example of mapping SSS starting with MODIS Aqua SST.

	SSS _{Model} vs. SSS _{SMOS}		SSS _{Model} vs. SSS _{AQUARIUS}	
	RMSE (SSS _{Model})	RMSE (SSS _{SMOS})	RMSE (SSS _{Model})	RMSE (SSS _{AQUARIUS})
R0	0.17	1.30	0.17	0.55
R2	0.15	1.29	0.10	0.42
R4	0.17	1.25	0.13	0.50
Global R0+R2+R4	0.16	1.27	0.13	0.48

Table 2: Pairwise comparison between modelled SSS (SSS_{Model}) and the two satellite SSS products (SSS_{SMOS} and SSS_{AQUARIUS}).

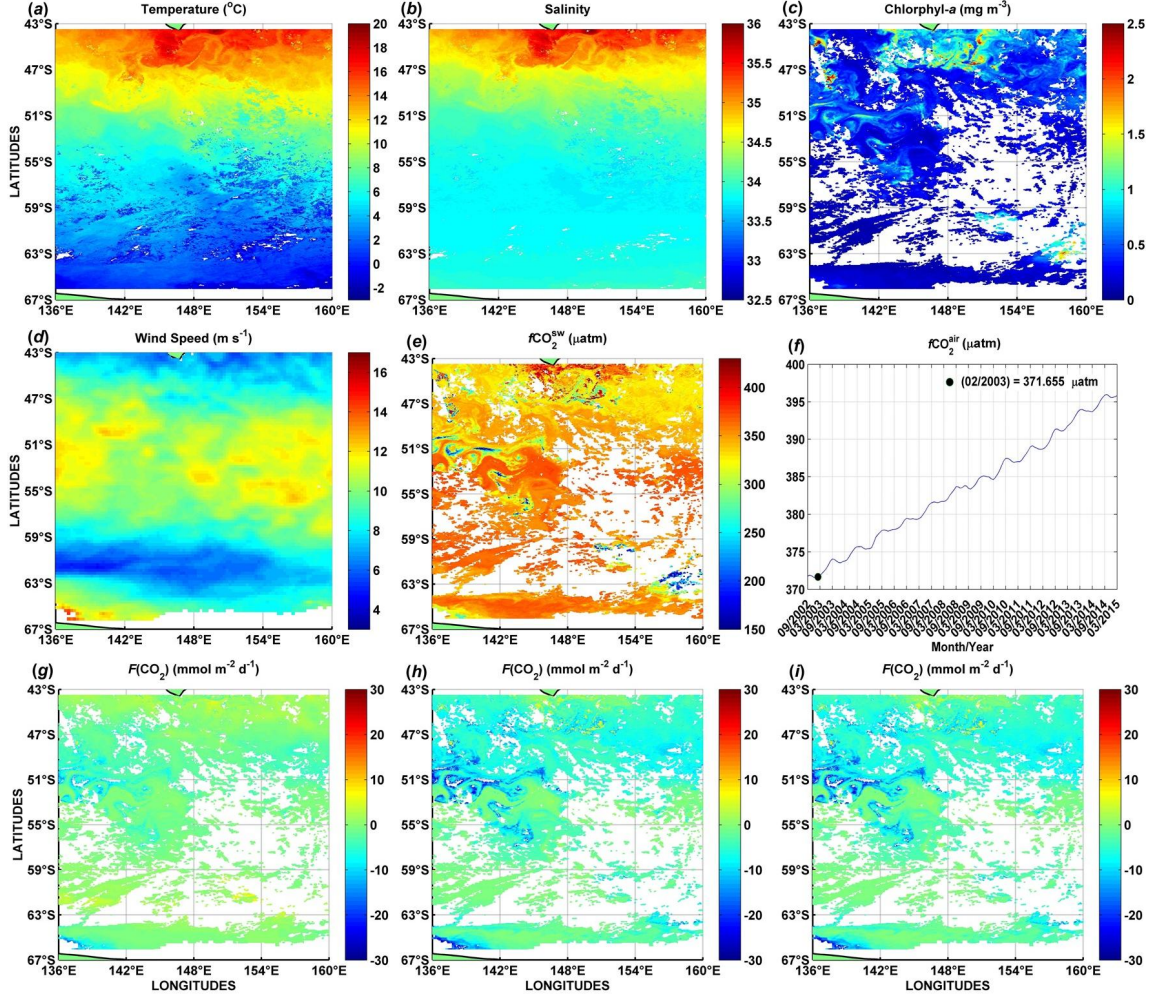


Figure 4: Example of mapping each parameter used to estimate $F(\text{CO}_2)$ for days between 02 Feb. 2003 and 09 Feb. 2003 : (a) 8-days composite image of MODIS Aqua SST. (b) Estimated SSS from MODIS Aqua SST using correspondent model. (c) 8-days composite image of MODIS Aqua Chlorophyll- a . (d) 8-days average WS corresponding to 8-days MODIS Aqua composite image. (e) Estimated $f(\text{CO}_2^{\text{sw}})$ using MODIS Aqua SST and chl- a . (f) $f(\text{CO}_2^{\text{atm}})$ from Cape Grim station (black point correspond to the title value). (g) Estimated $F(\text{CO}_2)$ using Liss and Merlivat (1986) equation. (h) Estimated $F(\text{CO}_2)$ using Ho et al. (2006) equation. (i) Estimated $F(\text{CO}_2)$ using Wanninkhof (2014) equation.

3.2 CO_2 fugacity ($f(\text{CO}_2^{\text{sw}})$)

Results of the $f(\text{CO}_2^{\text{sw}})$ modelling, validating and testing (as described in Section 2.5) are presented in Table 3. Figure 3(b) presents RMSE values of the validation and the test steps, as well as the regressions of the best FNNs (using the IMOS database). The best FNNs for R0, R2 and R4 were respectively FNNs with 17, 39

and 16 hidden neurons. Training the FNNs gives a RMSE of 8.62, 9.70 and 8.82 μatm for R0, R2 and R4 respectively (Considering all rotations, we obtain a general RMSE of about 9.00 μatm). *In situ* validation gives almost the same RMSEs as for training (8.61, 9.67 and 8.81 μatm for R0, R2 and R4 respectively and a global RMSE of 8.99 μatm). The MINERVE

$f(\text{CO}_2^{\text{sw}})$ *in situ* data used as a second *in situ* validation (a different database from that used for the FNNs trainings) gives a global RMSE of 13.56 μatm (13.63, 13.43 and 13.82 μatm for R0, R2 and R4 respectively) which is acceptable regarding to the difference between

the two databases; actually, this database is still scarce in term of time coverage (only three rotations between 2004/2005 and 2006/2007, Section 2.1).

	Training	<i>in situ</i> validation		satellite test
	results	IMOS data	MINERVE data	using MODIS Aqua data
R0	8.62	8.61	13.63	8.62
R2	9.70	9.67	13.43	10.30
R4	8.81	8.81	13.82	9.46
Global R0+R2+R4	9.00	8.99	13.56	9.45

Table 3: fCOSW modelling results (RMSEs expressed in μatm) : training, validation with *in situ* data of MINERVE and IMOS, and testing with satellite of MODIS Aqua.

The test with satellite data is the most important step which determine the precision of the FNNs. Results presented in Figure 3(b) show that even if the RMSE of validation step (with *in situ* data) decreases when adding neurons to the hidden layer, this may not improve the RMSE in the test step (with satellite data). A good RMSE of about 8.62, 10.30 and 9.46 μatm for R0, R2, and R4 respectively was obtained (considering all rotations, we obtain a general RMSE of about 9.45 μatm). The coefficients of correlation (R) were 0.73, 0.73 and 0.74 respectively for R0, R2 and R4 (considering all rotations, a general R of 0.86 was obtained). Figure 4(e) shows an example of mapping $f(\text{CO}_2^{\text{sw}})$ starting with MODIS Aqua SST and chl- a .

3.3 CO₂ air-sea flux ($F(\text{CO}_2)$)

The purpose of this study is to map $F(\text{CO}_2)$ using satellite data. The result's precision is estimated using *in situ* data as reference. We calculated the $F(\text{CO}_2)$ using equation 1 for both *in situ* and satellite data.

As cited previously, the k parameter was developed by several researchers, since the eighties until today (Table 4 presents a list of the most popular equations), k is directly affected by WS. Figure 5 shows

this relationship using the different gas transfer velocity equations (note that these equations were developed in different regions, with different conditions). As we can see, because of certain equations similarity, their plots are very close each other (for example, the two Sweeney et al. (2007) and Ho et al. (2011) (a) equations). The most used in literature is the Wanninkhof (1992) equation, which was recently improved in Wanninkhof (2014). We clearly see that at high WS (≥ 20 m s⁻¹) the difference between references increase especially for Wanninkhof and McGillis (1999) and equation (b) of Ho et al. (2011). The reason is that these two equations were developed for steady wind with a speed lower than 20 m s⁻¹. In this paper, we compare results calculated with three references :

- Liss and Merlivat (1986) : the equation was developed for the three WS cases (low, high and medium), and it is one of the first used methods.
- Ho et al. (2006) : the equation was developed in the southern ocean near to the studied area (few hundred km from the coast of New Zealand).
- Wanninkhof (2014) : the equation is the last version of the most used equation of Wanninkhof (1992).

Reference	Correspondent Equation
1 Liss and Merlivat (1986)	if $u < 3.6$ then $k = 0.170 \times u \times (Sc/600)^{-2/3}$ else if $3.6 < u \leq 13$ then $k = (2.85 \times u - 9.65) \times (Sc/600)^{-0.5}$ else $k = (5 \times u - 49.3) \times (Sc/600)^{-0.5}$
2 Wanninkhof (1992)	if $u \leq 6$ then $k = 0.31 \times u^2 \times (Sc/660)^{-0.5}$ else $k = 0.39 \times u^2 \times (Sc/660)^{-0.5}$
3 Wanninkhof and McGillis (1999)	$k = 0.0283 \times u^3 \times (Sc/660)^{-0.5}$
4 Nightingale et al. (2000)	$(0.222 \times u^2 + 0.333 \times u) \cdot (Sc/600)^{-0.5}$
5 Jean-Baptiste et al. (2002)	$k = 1.45 \times u^{1.5} \times (Sc/310)^{-0.5}$
6 Ho et al. (2006)	$k = 0.266 \times u^2 \times (Sc/600)^{-0.5}$
7 Sweeney et al. (2007)	$k = 0.27 \times u^2 \times (Sc/660)^{-0.5}$
8 Wanninkhof et al. (2009)	$k = 3 + 0.1 \times u + 0.064 \times u^2 + 0.011 \times u^3 \cdot (Sc/660)^{-0.5}$
9 Ho et al. (2011) equation (a)	$k = 0.286 \times u^2 \times (Sc/600)^{-0.5}$
10 Ho et al. (2011) equation (b)	$k = 0.0298 \times u^3 \times (Sc/600)^{-0.5}$
11 Wanninkhof (2014)	$k = 0.251 \times u^2 \times (Sc/660)^{-0.5}$

Table 4: List of the most popular equations used to calculate the gas transfer velocity (k). u represent the WS (in m s⁻¹) et Sc the Shmidt number.

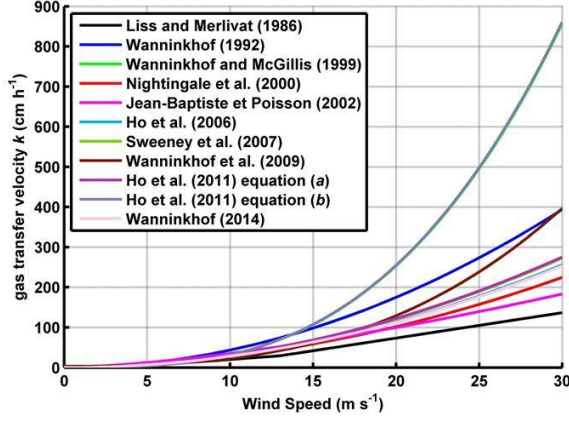


Figure 5: Gas Transfer Velocity (k) and WS relationship.

For each rotation, we present in Figure 6 a comparison between the calculated $F(\text{CO}_2)$ with both *in situ* and satellite data, and the regression between the two. Table 5 shows, for each reference, the RMSE of calculating $F(\text{CO}_2)$ between *in situ* and satellite data, as well as the R . The RMSE varies between 1.99 and 3.10 $\text{mmol CO}_2 \text{ m}^{-2} \text{ d}^{-1}$ and the R between 0.81 and 0.82. These results show that (1) using satellite data and interpolated data (using developed models of SSS and $f(\text{CO}_2^{\text{sw}})$) gives the ability to map $F(\text{CO}_2)$ over the region south of Australia and (2) this approach can be used with all k equations (The global R is almost the same for all equations, see appendix B for more details). Comparing with *in situ* data, this results allow us to provide $F(\text{CO}_2)$ maps during all the period from October to March using satellite data. Figure 4(g),(h) and (i) present an example of the mapped $F(\text{CO}_2)$ using all parameters (Figure 4(a) to (f)).

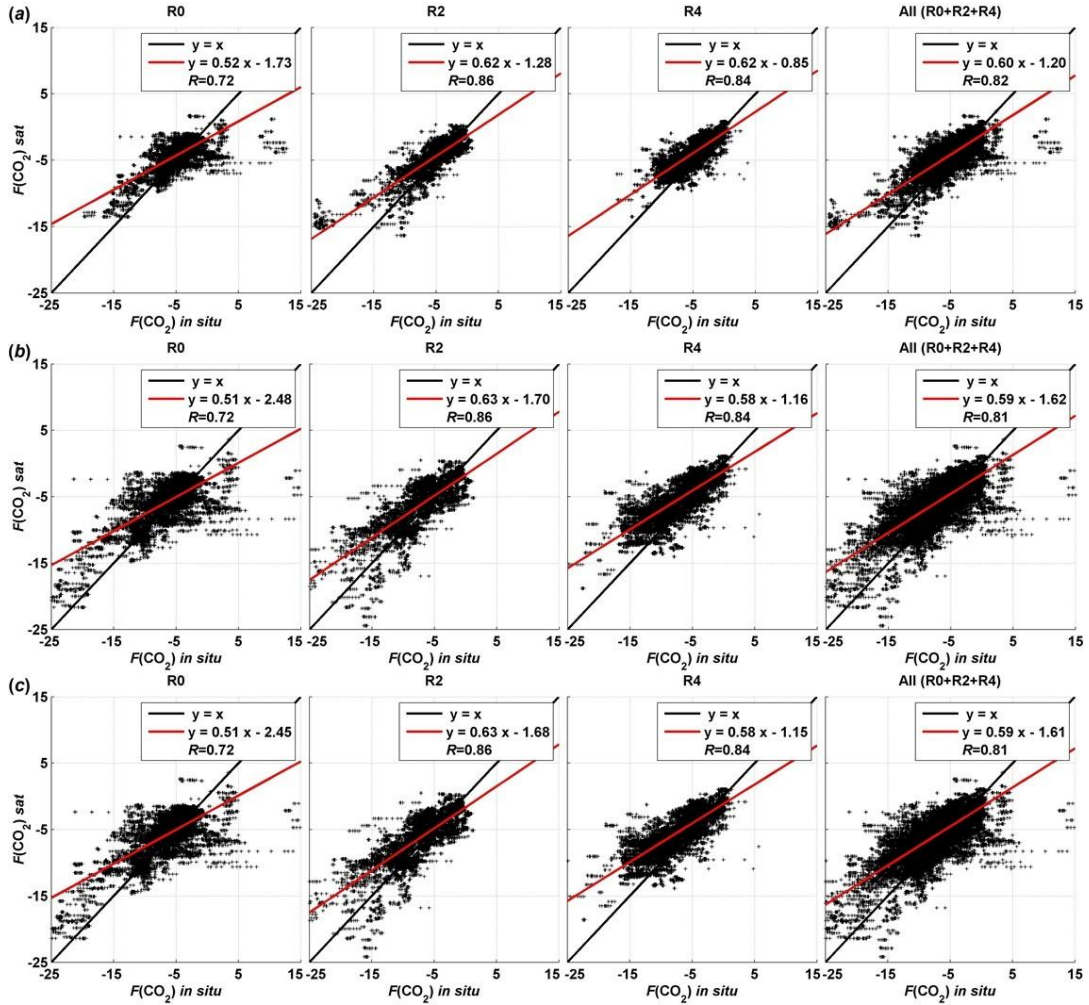


Figure 6: Comparison between the three references for $F(\text{CO}_2)$ calculated with *in situ* data ($F(\text{CO}_2) \text{ in situ}$) and with the estimated parameters using developed models ($F(\text{CO}_2) \text{ sat}$): (a) Using Liss and Merlivat (1986) equation. (b) Using Ho et al. (2006) equation. (c) Using Wanninkhof (2014) equation.

Equation Reference	RMSE (mmol CO ₂ m ⁻² d ⁻¹)			Global RMSE (mmol CO ₂ m ⁻² d ⁻¹)	Global <i>R</i>
	R0	R2	R4		
Liss and Merlivat (1986)	2.19	2.18	1.74	1.99	0.82
Ho et al. (2006)	3.39	3.30	2.76	3.10	0.81
Wanninkhof (2014)	3.36	3.27	2.73	3.07	0.81
Global R0+R2+R4	9.00	8.99	13.56	9.45	

Table 5: $f\text{CO}_2\text{SW}$ modelling results (RMSEs expressed in μatm) : training, validation with *in situ* data of MINERVE and IMOS, and testing with satellite of MODIS Aqua.

Figure ?? presents results of estimating $F(\text{CO}_2)$. The interannual average distribution of $F(\text{CO}_2)$ between October and March from 2002/2003 to 2014/2015 show that, for the three references, this region of the austral ocean acts as a sink of CO_2 . This aptitude of absorbing CO_2 is becoming higher through the years, especially if we look at the two references of Ho et al. (2006) and Wanninkhof (2014) (which are almost the same because of the similarity of their equations, $F(\text{CO}_2)$ is slightly underestimated by Ho et al. (2006) equation). We also note that the sub-Antarctic region (SAR, north of the polar front (PF)) is a stronger

sink than the Antarctic Region (AR, south of the PF). The reason is that the surface $f(\text{CO}_2^{\text{w}})$ is higher in the AR and this causes a lesser absorption of $f(\text{CO}_2^{\text{atm}})$.

If we look at the anomalies of the two Ho et al. (2006) and Wanninkhof (2014) (which are more appropriate to the studied region) we note a global anomaly of $-0.37 \text{ mmol CO}_2 \text{ m}^{-2} \text{ d}^{-1}$ with a small difference between the AR and SAR anomalies (-0.34 and $-0.39 \text{ mmol CO}_2 \text{ m}^{-2} \text{ d}^{-1}$ respectively).

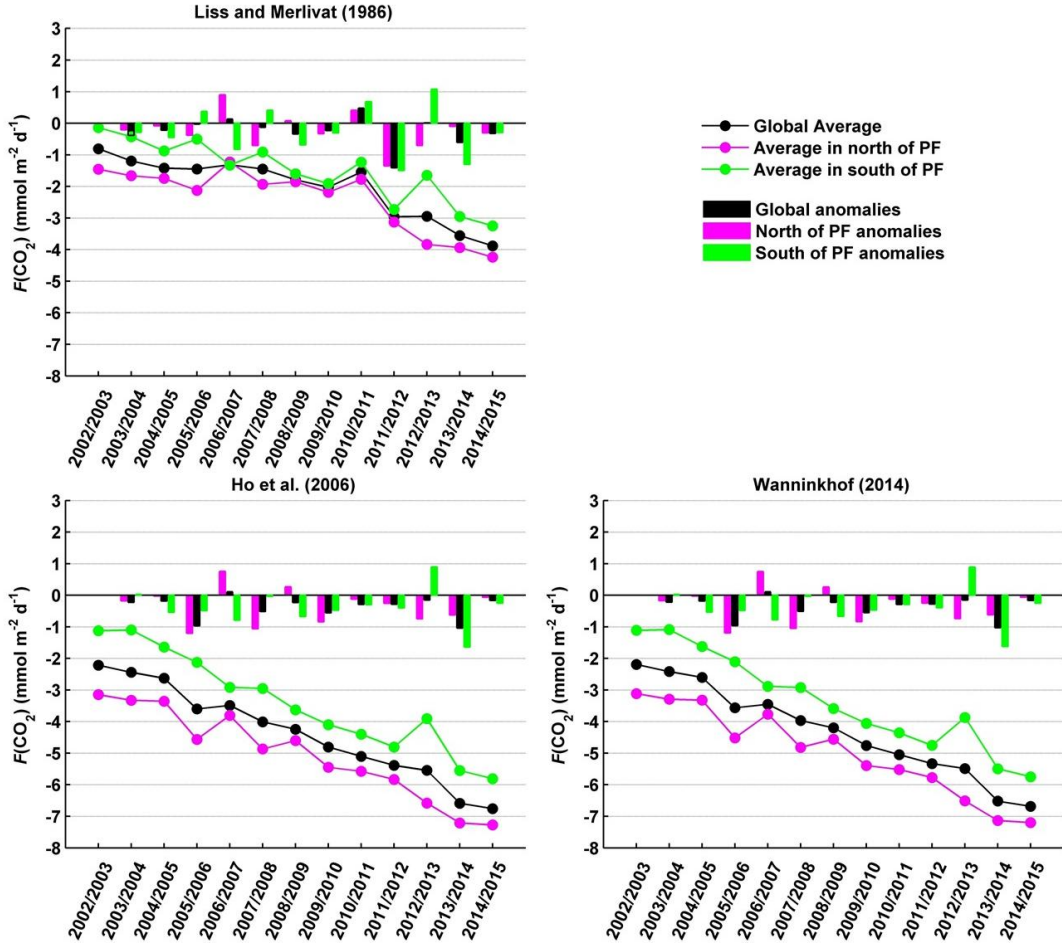


Figure 7: Inter-annual distributions and anomalies of estimated $F(\text{CO}_2)$ in the period of October-March.

This results show that the ocean react due to the increase of $f(\text{CO}_2^{\text{atm}})$ (as we see with Cape Grim data in Figure 4(f)). We note a global average absorption of about $2 \text{ mmol CO}_2 \text{ m}^{-2} \text{ d}^{-1}$ in 2002/2003 to about $7 \text{ mmol CO}_2 \text{ m}^{-2} \text{ d}^{-1}$ in 2014/2015. Table 6 presents details about maximums, means and minimums of $F(\text{CO}_2)$ according to Wanninkhof (2014). We note the decreasing of the three values over the years, which confirm the fact that this region is steadily absorbing more and more atmospheric CO_2 , thus acidifying this ocean area.

October to March	maximum of $F(\text{CO}_2)$	mean of $F(\text{CO}_2)$	minimum of $F(\text{CO}_2)$
2002/2003	-0.49	-2.19	-4.46
2003/2004	-0.56	-2.42	-5.35
2004/2005	-1.39	-2.60	-7.18
2005/2006	-1.24	-3.56	-7.21
2006/2007	-2.02	-3.45	-5.14
2007/2008	-1.86	-3.97	-6.32
2008/2009	-2.25	-4.20	-6.39
2009/2010	-2.61	-4.76	-7.08
2010/2011	-3.33	-5.05	-6.90
2011/2012	-3.48	-5.33	-7.60
2012/2013	-3.81	-5.49	-7.67
2013/2014	-4.62	-6.52	-8.04
2014/2015	-4.35	-6.69	-9.06

Table 6: Maximum, mean and minimum of $F(\text{CO}_2)$ (expressed in $\text{mmol CO}_2 \text{ m}^{-2} \text{ d}^{-1}$) presented following Wanninkhof (2014) equation with a precision of $\pm 3.07 \text{ mmol CO}_2 \text{ m}^{-2} \text{ d}^{-1}$.

Figure 8 presents the latitudinal distribution of the estimated $F(\text{CO}_2)$ using the last developed equation of Wanninkhof (2014) for the entire studied period. The first remark is that the increasing of CO_2 absorption is noticeable in all latitudes and zones throughout the years. From the north to the south, we see that the STZ is a stronger sink of CO_2 than the SAZ (from about $-4.45 \pm 2 \text{ mmol CO}_2 \text{ m}^{-2} \text{ d}^{-1}$ for the STZ to $-2.75 \pm 2 \text{ mmol CO}_2 \text{ m}^{-2} \text{ d}^{-1}$ for the SAZ in 2002/2003, and from about $-8.14 \pm 3 \text{ mmol CO}_2 \text{ m}^{-2} \text{ d}^{-1}$ for the STZ to $-6.86 \pm 3 \text{ mmol CO}_2 \text{ m}^{-2} \text{ d}^{-1}$ for the SAZ in 2014/2015). In Borges et al. (2007) the study of the STZ $F(\text{CO}_2)$ showed a direct relationship with the $f(\text{CO}_2^{\text{sw}})$ in this zone, which is in with our results. This is mainly caused by the abundance of chl- a reflecting a higher phytoplankton activity. This causes a higher $\Delta f\text{CO}_2$ by consuming sea-water CO_2 , enhancing the atmospheric CO_2 absorption. The PFZ is the zone that absorbs less atmospheric CO_2 ; in this zone, the PF (about 45.5°S) causes an increasing of $f(\text{CO}_2^{\text{sw}})$ which reduces the absorption, however we note a global average $F(\text{CO}_2)$ values of about $-0.55 \pm 1 \text{ mmol CO}_2 \text{ m}^{-2} \text{ d}^{-1}$ in 2002/2003 to $-4.5 \pm 2 \text{ mmol CO}_2 \text{ m}^{-2} \text{ d}^{-1}$ in 2014/2015. The AR is marked by a decreasing values of $F(\text{CO}_2)$ from the POOZ to the SFZ (from about -1.72 ± 2 for the POOZ to $-2.75 \pm 2 \text{ mmol CO}_2 \text{ m}^{-2} \text{ d}^{-1}$ for the SFZ in 2002/2003, and from about $-4.92 \pm 3 \text{ mmol CO}_2 \text{ m}^{-2} \text{ d}^{-1}$ for the POOZ to $-6.25 \pm 3 \text{ mmol CO}_2 \text{ m}^{-2} \text{ d}^{-1}$ for the SFZ in 2014/2015). In the

south of SFZ and the CAZ, we note a larger decreasing of the global $F(\text{CO}_2)$ average (from about $-4.00 \pm 5 \text{ mmol CO}_2 \text{ m}^{-2} \text{ d}^{-1}$ in 2002/2003 to $-9.00 \pm 6 \text{ mmol CO}_2 \text{ m}^{-2} \text{ d}^{-1}$ in 2014/2015). This higher absorption of atmospheric CO_2 is caused by the coupled effect of low temperature and the increasing chl- a (Moore and Abbott, 2000). However, we note a high fluctuation (higher standard deviations) which is mainly caused by region's dynamic and phytoplankton blooms episodes, presented by the upwelling of the Antarctic Bottom Water which is characterized by high $f(\text{CO}_2^{\text{sw}})$ values. These results confirm the $F(\text{CO}_2)$ trend of previous works in the southern ocean (Metzl et al., 1995, 1999, 2006; McNeil et al., 2007; Boutin et al., 2008).

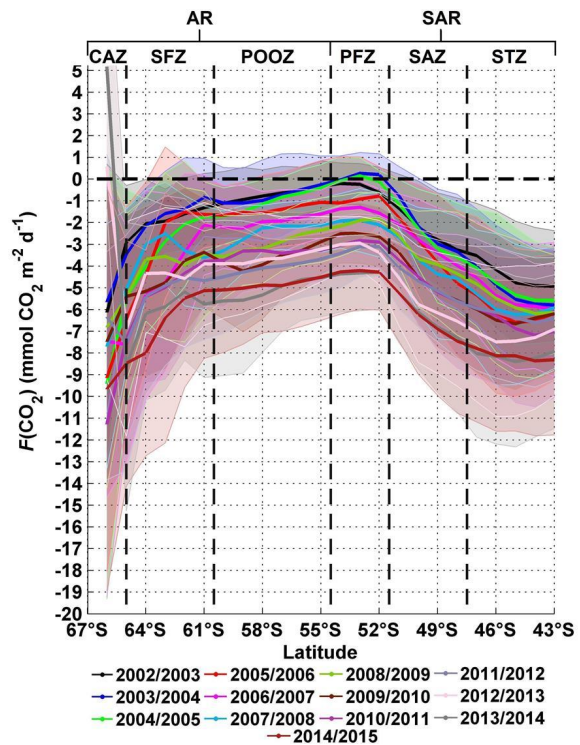


Figure 8: Average latitudinal distributions of the estimated $F(\text{CO}_2)$ in the period of October-March for the last developed equation of Wanninkhof (2014), shaded colors represent the standard deviations.

4 Conclusion

In this study, an algorithm was presented in order to develop models to estimate $F(\text{CO}_2)$ calculation parameters using satellite data. Figure 4 shows an example of mapping each parameter used to calculate $F(\text{CO}_2)$. This Figure represents results of each step presented in the general algorithm (Figure 2).

We started by the estimation of SSS over the period of October to March. The best type of image for this purpose was the 8-days composite images from MODIS Aqua. A RMSE of 0.16 was found for estimating SSS. We showed how we can use the MLR method (based on latitude and SST) to provide SSS with better accu-

racy than satellites like SMOS and AQUARIUS, and with a bigger spatio-temporal coverage (4 km × 4 km, since 2002). This result represents a good alternative to provide SSS maps from satellite data when satellite SST is available (As long as satellite SSS precision still low in this region).

The next step was estimating $f(\text{CO}_2^{\text{sw}})$, we showed how we used the FNN method to develop a neural network for each rotation. A global RMSE of about 9.45 μatm was found using satellite data. We used these neural networks to map the $f(\text{CO}_2^{\text{sw}})$ using exclusively satellite data (SST and chl-*a* from MODIS Aqua).

We further showed how we proceed to map $F(\text{CO}_2)$ using satellite data. We map $F(\text{CO}_2)$ at a resolution of 4 km × 4 km with a precision of about ± 3 $\text{mmol CO}_2 \text{ m}^{-2} \text{ d}^{-1}$ in the austral ocean south of Australia. Results show that this region is becoming an increasing sink of atmospheric CO_2 over the years.

This approach allows us to provide all parameters for $F(\text{CO}_2)$ calculation using remote sensing data of SST and chl-*a* (only $f(\text{CO}_2^{\text{atm}})$ for the tests was provided by the Cape Grim Station). *In situ* measurements provided by all projects are still very important to improve models and to calibrate remote sensing data. As a perspective for this work, all models of MLR and FNN, source codes, estimated parameters and maps will be compiled in a database to be published and related to this work in order to provide materials for comparison with other scientific results in this region. This will contribute on the global estimation of ocean CO_2 exchange to improve our knowledge of ocean reaction due to air CO_2 increasing, and its impact on ocean acidification.

Acknowledgements

We would like to thanks all data providers without whom, this work would not have been possible :

- Australian under-way CO_2 data ($f(\text{CO}_2^{\text{sw}})$) was sourced from the Integrated Marine Observing System (IMOS) which is supported by the Australian Government through the National Collaborative Research Infrastructure Strategy and the Super Science Initiative.
- Temperature and Salinity was provided by the SURVOSTRAL project which is the result of a partnership between several institutions : (1) The IPEV (Institut polaire français Paul Emile Victor) that supports the administrative organization and provides the logistics framework, human resources, and technical and financial resources to the development of French research in the Polar Regions, (2) The LEGOS (Laboratoire d'Etudes en Géophysique et Océanographie Spatiales) laboratory, which is responsible for analysing and distribution thermosalinograph data, (3) The CSIRO which provide technical

support, including the development of software for recording and analysis data and (4) The SIO (Scripps Institution of Oceanography), which participate in the development of deep ocean observing technologies and provides instruments for the program.

- Data of SST and $f(\text{CO}_2^{\text{sw}})$ provided by the MINERVE project also supported by the IPEV.
- $f(\text{CO}_2^{\text{atm}})$ data from the Cape Grim station was provide by the CSIRO Marine and Atmospheric Research and the Australian Bureau of Meteorology (Cape Grim Baseline Air Pollution Station).
- Satellite Data from MODIS Aqua and web resources were provided by the Ocean Biology Distributed Active Archive Center (OB.DAAC).
- ASCAT and QSCAT data were produced by Remote Sensing Systems with thanks to the NASA Ocean Vector Winds Science Team for funding and support.

I am particularly grateful to the RV *L'Astrolabe* crew and all volunteers who participate to research cruises on-board it. A special thanks to PhD funders : (1) the Algerian Ministry of Higher Education and Scientific Research (MESRS) for M.A. BENAL-LAL and H. MOUSSA, and (2) the the Brazilian National Council for Scientific and Technological Development (CNPq) for J.M. Lencina-Avila. Thanks to them for giving us the opportunity to undertake research at the IMAGES-ESPACE_DEV laboratory.

Appendices

A Definition of SSS error parameters

Here we present the error parameters used to choose the best type of SST satellite data in SSS estimation. e_i and n represent error and number of data points respectively :

- The Mean Bias Error (MBE) computed as :

$$MBE = \left[\frac{1}{n} \sum_{i=1}^n e_i \right] \quad (3)$$

- The Mean Absolute Error (MAE) computed as :

$$MAE = \left[\frac{1}{n} \sum_{i=1}^n |e_i| \right] \quad (4)$$

- The Root Mean Square Error (RMSE) computed as :

$$RMSE = \left[\frac{1}{n} \sum_{i=1}^n e_i^2 \right]^{\frac{1}{2}} \quad (5)$$

- The Percentage of Lost Data (PLD) represent the proportion (%) of *in situ* data with no corresponding satellite data.

B All $F(\text{CO}_2)$ results

Table 7 presents the application of the approach with all k equations presented in Table 4. This results

confirm that the R remains high when using different equations; the approach presented in this paper is than independent and can be used for all equations.

	Equation	R0		R2		R4		Glob.	Glob.
		RMSE	R	RMSE	R	RMSE	R	RMSE	R
1	Liss and Merlivat (1986)	2.19	0.72	2.18	0.86	1.74	0.84	2.00	0.82
2	Wanninkhof (1992)	5.22	0.72	5.07	0.86	4.25	0.84	4.76	0.81
3	Wanninkhof and McGillis (1999)	5.88	0.66	5.80	0.86	3.93	0.82	5.07	0.78
4	Nightingale et al. (2000)	3.11	0.73	3.03	0.86	2.56	0.84	2.85	0.81
5	Jean-Baptiste et al. (2002)	3.44	0.75	3.38	0.85	3.02	0.84	3.24	0.82
6	Ho et al. (2006)	3.39	0.72	3.30	0.86	2.76	0.84	3.10	0.81
7	Sweeney et al. (2007)	3.61	0.72	3.51	0.86	2.94	0.84	3.30	0.81
8	Wanninkhof et al. (2009)	4.17	0.66	4.15	0.82	2.97	0.81	3.67	0.77
9	Ho et al. (2011) (a)	3.65	0.72	3.55	0.86	2.97	0.84	3.33	0.81
10	Ho et al. (2011) (b)	5.90	0.66	5.82	0.86	3.94	0.82	5.09	0.78
11	Wanninkhof (2014)	3.36	0.72	3.27	0.86	2.73	0.84	3.07	0.81

Table 7: $F(\text{CO}_2)$ results for all equations. The RMSE (expressed in $\text{mmol CO}_2 \text{ m}^{-2} \text{ d}^{-1}$) is calculated using the $F(\text{CO}_2)$ calculated with *in situ* data and with estimated parameters and the R is the coefficient of correlation between the two. Glob. : Global.

References

- Barreto, G. A. (2007). Time Series Prediction with the Self-Organizing Map: A Review. In Hammer, D. B. and Hitzler, D. P., editors, *Perspectives of Neural-Symbolic Integration*, volume 77, pages 135–158. Springer Berlin Heidelberg.
- Barreto, G. A. and Araújo, A. F. R. (2002). Non-linear Modeling of Dynamic Systems with the Self-Organizing Map. In Dorronsoro, J. R., editor, *Artificial Neural Networks — ICANN 2002*, pages 975–980. Springer Berlin Heidelberg.
- Benallal, M., Moussa, H., Touratier, F., Goyet, C., and Poisson, A. (2016). Ocean salinity from satellite-derived temperature in the Antarctic Ocean. *Antarctic Science*, 28(02):127–134.
- Borges, A. V., Tilbrook, B., Metzl, N., Lenton, A., and Delille, B. (2007). Inter-annual variability of the carbon dioxide oceanic sink south of Tasmania. *Biogeosciences Discussions*, 4(5):3639–3671.
- Boutin, J., Merlivat, L., Hénoq, C., Martin, N., and Sallée, J. B. (2008). Air-sea CO_2 flux variability in frontal regions of the Southern Ocean from Carbon Interface Ocean Atmosphere drifters. *Limnology and Oceanography*, 53(5part2):2062–2079.
- Brewer, P. G. (1978). Direct observation of the oceanic CO_2 increase. *Geophysical Research Letters*, 5(12):997–1000.
- Bureau International Des Poids Et Mesures (1955). Dictionnaire de l'unité de mesure des poids et mesures. Technical report, Bureau International Des Poids Et Mesures, Quai des Grands-Augustins, 55, Paris.
- Chaigneau, A. and Morrow, R. (2002). Surface temperature and salinity variations between Tasmania and Antarctica, 1993–1999. *Journal of Geophysical Research: Oceans* (1978–2012), 107(C12):SRF–22.
- Chen, G.-T. and Millero, F. J. (1979). Gradual increase of oceanic CO_2 . *Nature*, 277(5693):205–206.
- Chierici, M., Signorini, S. R., Mattsdotter-Björk, M., Fransson, A., and Olsen, A. (2012). Surface water fCO_2 algorithms for the high-latitude Pacific sector of the Southern Ocean. *Remote Sensing of Environment*, 119:184–196.
- Copin-Montegut, C. (1988). A new formula for the effect of temperature on the partial pressure of CO_2 in seawater. *Marine Chemistry*, 25(1):29–37.
- De Villiers, J. and Barnard, E. (1993). Backpropagation neural nets with one and two hidden layers. *IEEE Transactions on Neural Networks*, 4(1):136–141.
- Dlugokencky and Tans, P. (2016). Trends in Atmospheric Carbon Dioxide.
- Feely, R., Sabine, C., Takahashi, T., and Wanninkhof, R. (2001). Uptake and Storage of Carbon Dioxide in the Ocean: The Global CO_2 Survey. *Oceanography*, 14(4):18–32.
- Flexer, A. (1999). On the Use of Self-Organizing Maps for Clustering and Visualization. In Żytkow, J. M. and Rauch, J., editors, *Principles of Data Mining and Knowledge Discovery*, number 1704 in Lecture Notes in Computer Science, pages 80–88. Springer Berlin Heidelberg. DOI: 10.1007/978-3-540-48247-5_9.
- Friedrich, T. and Oschlies, A. (2009). Neural network-based estimates of North Atlantic surface pCO_2 from satellite data: A methodological study. *Journal of Geophysical Research: Oceans*, 114(C3):C03020.

- Gordon, L. I. and Jones, L. B. (1973). The effect of temperature on carbon dioxide partial pressures in seawater. *Marine Chemistry*, 1(4):317–322.
- Goyet, C., Millero, F. J., Poisson, A., and Shafer, D. K. (1993). Temperature dependence of CO₂ fugacity in seawater. *Marine Chemistry*, 44(2):205–219.
- Goyet, C. and Poisson, A. (1989). New determination of carbonic acid dissociation constants in seawater as a function of temperature and salinity. *Deep Sea Research Part A. Oceanographic Research Papers*, 36(11):1635–1654.
- Ho, D. T., Law, C. S., Smith, M. J., Schlosser, P., Harvey, M., and Hill, P. (2006). Measurements of air-sea gas exchange at high wind speeds in the Southern Ocean: Implications for global parameterizations. *Geophysical Research Letters*, 33(16).
- Ho, D. T., Wanninkhof, R., Schlosser, P., Ullman, D. S., Hebert, D., and Sullivan, K. F. (2011). Toward a universal relationship between wind speed and gas exchange: Gas transfer velocities measured with ³He/SF₆ during the Southern Ocean Gas Exchange Experiment. *Journal of Geophysical Research: Oceans*, 116(C4):C00F04.
- Hornik, K., Stinchcombe, M., and White, H. (1989). Multilayer feedforward networks are universal approximators. *Neural Networks*, 2(5):359–366.
- IAPSO Working Group on Symbols, Units, and Nomenclature in Physical Oceanography. (1985). *The International System of Units (SI) in oceanography : report of IAPSO Working Group on Symbols, Units, and Nomenclature in Physical Oceanography (SUN)*. Unesco, Paris.
- Inoue, H. Y., Ishii, M., Matsueda, H., Kawano, T., Murata, A., and Takasugi, Y. (2003). Distribution of the partial pressure of CO₂ in surface water (pCO_{2w}) between Japan and the Hawaiian Islands: pCO_{2w}-SST relationship in the winter and summer. *Tellus B*, 55(2):456–465.
- Irie, B. and Miyake, S. (1988). Capabilities of three-layered perceptrons. In *IEEE International Conference on Neural Networks, 1988*, pages 641–648 vol.1.
- Jamet, C., Moulin, C., Lefèvre, N., and others (2007). Estimation of the oceanic pCO₂ in the North Atlantic from VOS lines in-situ measurements: parameters needed to generate seasonally mean maps. In *Annales Geophysicae*, volume 25, pages 2247–2257.
- Jean-Baptiste, P., Fourré, E., and Poisson, A. (2002). Gas Transfer Velocities for 3he in a Lake at High Wind Speeds. In Donelan, M. A., Mennan, W., Saltzman, E. S., and Wanninkhof, R., editors, *Gas Transfer at Water Surfaces*, pages 233–238. American Geophysical Union.
- Kohonen, T. (2001). *Self-Organizing Maps*, volume 30 of *Springer Series in Information Sciences*. Springer Berlin Heidelberg, Berlin, Heidelberg.
- Laika, H., Goyet, C., Vouve, F., Poisson, A., and Touratier, F. (2009). Interannual properties of the CO₂ system in the Southern Ocean south of Australia. *Antarctic Science*, 21(06):663.
- Landschützer, P., Gruber, N., Bakker, D. C. E., Schuster, U., Nakaoka, S., Payne, M. R., Sasse, T. P., and Zeng, J. (2013). A neural network-based estimate of the seasonal to inter-annual variability of the Atlantic Ocean carbon sink. *Biogeosciences*, 10(11):7793–7815.
- Le Quéré, C., Moriarty, R., Andrew, R. M., Canadell, J. G., Sitch, S., Korsbakken, J. I., Friedlingstein, P., Peters, G. P., Andres, R. J., Boden, T. A., Houghton, R. A., House, J. I., Keeling, R. F., Tans, P., Arneeth, A., Bakker, D. C. E., Barbero, L., Bopp, L., Chang, J., Chevallier, F., Chini, L. P., Ciais, P., Fader, M., Feely, R. A., Gkritzalis, T., Harris, I., Hauck, J., Ilyina, T., Jain, A. K., Kato, E., Kitidis, V., Klein Goldewijk, K., Koven, C., Landschützer, P., Lauvset, S. K., Lefèvre, N., Lenton, A., Lima, I. D., Metzl, N., Millero, F., Munro, D. R., Murata, A., Nabel, J. E. M. S., Nakaoka, S., Nojiri, Y., O’Brien, K., Olsen, A., Ono, T., Pérez, F. F., Pfeil, B., Pierrot, D., Poulter, B., Rehder, G., Rödenbeck, C., Saito, S., Schuster, U., Schwinger, J., Séférian, R., Steinhoff, T., Stocker, B. D., Sutton, A. J., Takahashi, T., Tilbrook, B., van der Laan-Luijkx, I. T., van der Werf, G. R., van Heuven, S., Vandemark, D., Viovy, N., Wiltshire, A., Zaehle, S., and Zeng, N. (2015). Global Carbon Budget 2015. *Earth System Science Data*, 7(2):349–396.
- Le Quéré, C., Moriarty, R., Andrew, R. M., Peters, G. P., Ciais, P., Friedlingstein, P., Jones, S. D., Sitch, S., Tans, P., Arneeth, A., Boden, T. A., Bopp, L., Bozec, Y., Canadell, J. G., Chevallier, F., Cosca, C. E., Harris, I., Hoppema, M., Houghton, R. A., House, J. I., Jain, A., Johannessen, T., Kato, E., Keeling, R. F., Kitidis, V., Klein Goldewijk, K., Koven, C., Landa, C. S., Landschützer, P., Lenton, A., Lima, I. D., Marland, G., Mathis, J. T., Metzl, N., Nojiri, Y., Olsen, A., Ono, T., Peters, W., Pfeil, B., Poulter, B., Raupach, M. R., Regnier, P., Rödenbeck, C., Saito, S., Salisbury, J. E., Schuster, U., Schwinger, J., Séférian, R., Segsneider, J., Steinhoff, T., Stocker, B. D., Sutton, A. J., Takahashi, T., Tilbrook, B., van der Werf, G. R., Viovy, N., Wang, Y.-P., Wanninkhof, R., Wiltshire, A., and Zeng, N. (2014). Global carbon budget 2014. *Earth System Science Data Discussions*, 7(2):521–610.
- Lefèvre, N. and Taylor, A. (2002). Estimating pCO₂ from sea surface temperatures in the Atlantic gyres. *Deep Sea Research Part I: Oceanographic Research Papers*, 49(3):539–554.
- Lefèvre, N., Watson, A. J., and Watson, A. R. (2005). A comparison of multiple regression and neural network techniques for mapping in situ pCO₂ data. *Tellus B*, 57(5):375–384.

- Liss, P. S. and Merlivat, L. (1986). Air-Sea Gas Exchange Rates: Introduction and Synthesis. In Buat-Ménard, P., editor, *The Role of Air-Sea Exchange in Geochemical Cycling*, number 185 in NATO ASI Series, pages 113–127. Springer Netherlands.
- McNeil, B. I., Metzl, N., Key, R. M., Matear, R. J., and Corbiere, A. (2007). An empirical estimate of the Southern Ocean air-sea CO₂ flux. *Global Biogeochemical Cycles*, 21(3):GB3011.
- Metzl, N., Brunet, C., Jabaud-Jan, A., Poisson, A., and Schauer, B. (2006). Summer and winter air-sea CO₂ fluxes in the Southern Ocean. *Deep Sea Research Part I: Oceanographic Research Papers*, 53(9):1548–1563.
- Metzl, N., Poisson, A., Louanchi, F., Brunet, C., Schauer, B., and Bres, B. (1995). Spatio-temporal distributions of air-sea fluxes of CO₂ in the Indian and Antarctic oceans. *Tellus B*, 47(1-2):56–69.
- Metzl, N., Tilbrook, B., and Poisson, A. (1999). The annual fCO₂ cycle and the air-sea CO₂ flux in the sub-Antarctic Ocean. *Tellus B*, 51(4):849–861.
- Moore, J. K. and Abbott, M. R. (2000). Phytoplankton chlorophyll distributions and primary production in the Southern Ocean. *Journal of Geophysical Research: Oceans (1978–2012)*, 105(C12):28709–28722.
- Morrow, R. and Kestenare, E. (2014). Nineteen-year changes in surface salinity in the Southern Ocean south of Australia. *Journal of Marine Systems*, 129:472–483.
- Morrow, R., Valladeau, G., and Sallee, J.-B. (2008). Observed subsurface signature of Southern Ocean sea level rise. *Progress in Oceanography*, 77(4):351–366.
- Moussa, H., Benallal, M. A., Goyet, C., Lefevre, N., EL Jai, M. C., Guglielmi, V., and Touratier, F. (2015). A comparison of Multiple Non-linear regression and neural network techniques for sea surface salinity estimation in the tropical Atlantic ocean based on satellite data. *ESAIM: Proceedings and Surveys*, 49:65–77.
- Moussa, H., Benallal, M. A., Goyet, C., and Lefèvre, N. (2016). Satellite-derived CO₂ fugacity in surface seawater of the tropical Atlantic Ocean using a feedforward neural network. *International Journal of Remote Sensing*, 37(3):580–598.
- Nakaoka, S., Telszewski, M., Nojiri, Y., Yasunaka, S., Miyazaki, C., Mukai, H., and Usui, N. (2013). Estimating temporal and spatial variation of ocean surface CO₂ in the North Pacific using a self-organizing map neural network technique. *Biogeosciences*, 10(9):6093–6106.
- NASA Goddard Space Flight Center, Ocean Ecology Laboratory, Ocean Biology Processing Group (2014). MODIS-Aqua Ocean Color Data. Technical report, NASA Goddard Space Flight Center, Ocean Ecology Laboratory, Ocean Biology Processing Group.
- Nightingale, P. D., Malin, G., Law, C. S., Watson, A. J., Liss, P. S., Liddicoat, M. I., Boutin, J., and Upstill-Goddard, R. C. (2000). In situ evaluation of air-sea gas exchange parameterizations using novel conservative and volatile tracers. *Global Biogeochemical Cycles*, 14(1):373–387.
- Olsen, A., Triñanes, J. A., and Wanninkhof, R. (2004). Sea-air flux of CO₂ in the Caribbean Sea estimated using in situ and remote sensing data. *Remote Sensing of Environment*, 89(3):309–325.
- Sabine, C. L., Feely, R. A., Gruber, N., Key, R. M., Lee, K., Bullister, J. L., Wanninkhof, R., Wong, C. S., Wallace, D. W. R., Tilbrook, B., Millero, F. J., Peng, T.-H., Kozyr, A., Ono, T., and Rios, A. F. (2004). The Oceanic Sink for Anthropogenic CO₂. *Science*, 305(5682):367–371.
- Sallée, J.-B., Matear, R. J., Rintoul, S. R., and Lenton, A. (2012). Localized subduction of anthropogenic carbon dioxide in the Southern Hemisphere oceans. *Nature Geoscience*, 5(8):579–584.
- Sarmiento, J. L. and Gruber, N. (2002). Sinks for anthropogenic carbon. *Physics Today*, 55(8):30–36.
- Siegenthaler, U. and Sarmiento, J. L. (1993). Atmospheric carbon dioxide and the ocean. *Nature*, 365(6442):119–125.
- Stephens, M. P., Samuels, G., Olson, D. B., Fine, R. A., and Takahashi, T. (1995). Sea-air flux of CO₂ in the North Pacific using shipboard and satellite data. *Journal of Geophysical Research: Oceans*, 100(C7):13571–13583.
- Sweeney, C., Gloor, E., Jacobson, A. R., Key, R. M., McKinley, G., Sarmiento, J. L., and Wanninkhof, R. (2007). Constraining global air-sea gas exchange for CO₂ with recent bomb 14c measurements. *Global Biogeochemical Cycles*, 21(2):GB2015.
- Takahashi, T., Sutherland, S. C., Sweeney, C., Poisson, A., Metzl, N., Tilbrook, B., Bates, N., Wanninkhof, R., Feely, R. A., Sabine, C., and others (2002). Global sea-air CO₂ flux based on climatological surface ocean pCO₂, and seasonal biological and temperature effects. *Deep Sea Research Part II: Topical Studies in Oceanography*, 49(9):1601–1622.
- Tamura, S. and Tateishi, M. (1997). Capabilities of a four-layered feedforward neural network: four layers versus three. *IEEE Transactions on Neural Networks*, 8(2):251–255.
- Telszewski, M., Chazottes, A., Schuster, U., Watson, A. J., Moulin, C., Bakker, D. C. E., González-Dávila, M., Johannessen, T., Körtzinger, A., Lüger, H., Olsen, A., Omar, A., Padin, X. A., Ríos, A. F., Steinhoff, T., Santana-Casiano, M., Wallace, D.

- W. R., and Wanninkhof, R. (2009). Estimating the monthly pCO₂ distribution in the North Atlantic using a self-organizing neural network. *Biogeosciences*, 6(8):1405–1421.
- Tilbrook, B., Neill, C., and Akl, J. (2013). Integrated Marine Observing System (IMOS) sea surface and atmospheric fCO₂ measurements in the Southern Ocean onboard MV L’Astrolabe in 2013. *Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, US Department of Energy, Oak Ridge, Tennessee*.
- Wanninkhof, R. (1992). Relationship between wind speed and gas exchange over the ocean. *Journal of Geophysical Research*, 97(C5):7373.
- Wanninkhof, R. (2014). Relationship between wind speed and gas exchange over the ocean revisited: Gas exchange and wind speed over the ocean. *Limnology and Oceanography: Methods*, 12(6):351–362.
- Wanninkhof, R., Asher, W. E., Ho, D. T., Sweeney, C., and McGillis, W. R. (2009). Advances in Quantifying Air-Sea Gas Exchange and Environmental Forcing*. *Annual Review of Marine Science*, 1(1):213–244.
- Wanninkhof, R. and McGillis, W. R. (1999). A cubic relationship between air-sea CO₂ exchange and wind speed. *Geophysical Research Letters*, 26(13):1889–1892.
- Weiss (1974). Carbon dioxide in water and seawater: the solubility of a non-ideal gas. *Marine Chemistry*, 2:203–215.
- Xu, R. and Wunsch, D., I. (2005). Survey of clustering algorithms. *IEEE Transactions on Neural Networks*, 16(3):645–678.
- Zeng, J., Nojiri, Y., Landschützer, P., Telszewski, M., and Nakaoka, S. (2014). A Global Surface Ocean fCO₂ Climatology Based on a Feed-Forward Neural Network. *Journal of Atmospheric and Oceanic Technology*, 31(8):1838–1849.
- Zhu, Y., Shang, S., Zhai, W., and Dai, M. (2009). Satellite-derived surface water pCO₂ and air-sea CO₂ fluxes in the northern South China Sea in summer. *Progress in Natural Science*, 19(6):775–779.

E.5 Publication 5 : Evolution of the human impact on oceans (en cours de préparation)

Evolution of the human impact on oceans

Catherine Goyet¹ Mohamed Anis Benallal² Amandine Bijoux³
Véronique Guglielmi⁴ Moussa Hadjer⁵ Anne-Cécile Ribou⁶
Franck Touratier⁷

CONTRIBUTED VOLUME / Springer ed.

Contribution to :

Co-viability of Social and Ecological Systems: Reconnect Man to the
Biosphere in an Era of Global Change

- ¹ Oceanographer, Professor of oceanography, IMAGES laboratory, UMR ESPACE-DEV, Perpignan University, Via Domitia, 52 avenue Paul Alduy, 66860 Perpignan, France, email : cgoyet@univ-perp.fr
- ² Computer scientist, Ph.D student, IMAGES laboratory, UMR ESPACE-DEV, Perpignan University, Via Domitia, 52 avenue Paul Alduy, 66860 Perpignan, France, email : mohamed.benallal@univ-perp.fr
- ³ Chemist, Ph.D student, OBS-Banyuls, Quai Racovizta, 66650 Banyuls-sur-Mer, France, email : bijoux@obs-banyuls.fr
- ⁴ Digital processing, Professor associate, IMAGES laboratory, UMR ESPACE-DEV, université de Perpignan Via Domitia, 52 avenue Paul Alduy, 66860 Perpignan, France, email : veronique.guglielmi@univ-perp.fr
- ⁵ Computer scientist, Ph.D student, IMAGES laboratory, UMR ESPACE-DEV, Perpignan University, Via Domitia, 52 avenue Paul Alduy, 66860 Perpignan, France, email : hadjer.moussa@univ-perp.fr
- ⁶ Environmental Chemistry, Professor associate, IMAGES, UMR ESPACE-DEV, université de Perpignan Via Domitia, 52 avenue Paul Alduy, 66860 Perpignan, France, email : ribou@univ-perp.fr
- ⁷ Oceanographer, Professor associate, IMAGES, UMR ESPACE-DEV, université de Perpignan Via Domitia, 52 avenue Paul Alduy, 66860 Perpignan, France, email : touratier@univ-perp.fr

Abstract

Since an important degree of the carbon dioxide (CO₂) that mankind emits in the atmosphere ends up in the ocean as a result of gas exchange through the air-sea interface, the ocean acidifies itself (because of the reaction CO₂ produces when it contacts water). Ocean acidification is one of the perturbing anthropogenic effects (because of human activity), not only for maritime ecosystems but also for Man himself as he will have to adapt. He will have to deal with variations in fisheries and other costal touristic activities (reduction of corals, and so on). However, the CO₂ anthropogenic penetration in the ocean is not uniform, and the cold Polar Regions will be affected by the change faster than the warm tropical regions. The study that we are conducting addresses several points of the quantification, in time and space, of CO₂ anthropogenic penetration in the ocean. One of them has permitted the identification of breaking points. We demonstrate that four points of rupture can be located in order to quantify risks of ocean acidification, which may have as a result the dissolution of calcium carbonates that are indispensable to maritime ecosystems (shells, corals, and so on). Thus, man will have to act rapidly to reduce deleterious effects and to find long-lasting solutions.

Annexe F

Copyright, Figure 2.6

ELSEVIER LICENSE TERMS AND CONDITIONS

Jun 13, 2016

This Agreement between Hadjer MOUSSA ("You") and Elsevier ("Elsevier") consists of your license details and the terms and conditions provided by Elsevier and Copyright Clearance Center.

License Number	3887050965517
License date	Jun 13, 2016
Licensed Content Publisher	Elsevier
Licensed Content Publication	Deep Sea Research Part I: Oceanographic Research Papers
Licensed Content Title	Near-surface circulation in the Tropical Atlantic Ocean
Licensed Content Author	Rick Lumpkin, Silvia L. Garzoli
Licensed Content Date	March 2005
Licensed Content Volume Number	52
Licensed Content Issue Number	3
Licensed Content Pages	24
Start Page	495
End Page	518
Type of Use	reuse in a thesis/dissertation
Portion	figures/tables/illustrations
Number of figures/tables/illustrations	1
Format	both print and electronic
Are you the author of this Elsevier article?	No
Will you be translating?	No
Order reference number	
Original figure numbers	Figure 1
Title of your thesis/dissertation	Traitement automatique de données océanographiques pour l'interpolation de la fCO ₂ de surface dans l'océan atlantique tropical, en utilisant les données satellitaires
Expected completion date	Sep 2016
Estimated size (number of pages)	350
Elsevier VAT number	GB 494 6272 12
Requestor Location	Hadjer MOUSSA 18 Bis AV ALFRED SAUVY RCE CANIGOU APT 8C106 PERPIGNAN, 66100 France Attn: Hadjer MOUSSA
Total	0.00 EUR
Terms and Conditions	

INTRODUCTION

1. The publisher for this copyrighted material is Elsevier. By clicking "accept" in connection with completing this licensing transaction, you agree that the following terms and conditions apply to this transaction (along with the Billing and Payment terms and conditions established by Copyright Clearance Center, Inc. ("CCC"), at the time that you opened your Rightslink account and that are available at any time at <http://myaccount.copyright.com>).

GENERAL TERMS

2. Elsevier hereby grants you permission to reproduce the aforementioned material subject to the terms and conditions indicated.
3. Acknowledgement: If any part of the material to be used (for example, figures) has appeared in our publication with credit or acknowledgement to another source, permission must also be sought from that source. If such permission is not obtained then that material may not be included in your publication/copies. Suitable acknowledgement to the source must be made, either as a footnote or in a reference list at the end of your publication, as follows:
"Reprinted from Publication title, Vol /edition number, Author(s), Title of article / title of chapter, Pages No., Copyright (Year), with permission from Elsevier [OR APPLICABLE SOCIETY COPYRIGHT OWNER]."
Also Lancet special credit - "Reprinted from The Lancet, Vol. number, Author(s), Title of article, Pages No., Copyright (Year), with permission from Elsevier."
4. Reproduction of this material is confined to the purpose and/or media for which permission is hereby given.
5. Altering/Modifying Material: Not Permitted. However figures and illustrations may be altered/adapted minimally to serve your work. Any other abbreviations, additions, deletions and/or any other alterations shall be made only with prior written authorization of Elsevier Ltd. (Please contact Elsevier at permissions@elsevier.com)
6. If the permission fee for the requested use of our material is waived in this instance, please be advised that your future requests for Elsevier materials may attract a fee.
7. Reservation of Rights: Publisher reserves all rights not specifically granted in the combination of (i) the license details provided by you and accepted in the course of this licensing transaction, (ii) these terms and conditions and (iii) CCC's Billing and Payment terms and conditions.
8. License Contingent Upon Payment: While you may exercise the rights licensed immediately upon issuance of the license at the end of the licensing process for the transaction, provided that you have disclosed complete and accurate details of your proposed use, no license is finally effective unless and until full payment is received from you (either by publisher or by CCC) as provided in CCC's Billing and Payment terms and conditions. If full payment is not received on a timely basis, then any license preliminarily granted shall be deemed automatically revoked and shall be void as if never granted. Further, in the event that you breach any of these terms and conditions or any of CCC's Billing and Payment terms and conditions, the license is automatically revoked and shall be void as if never granted. Use of materials as described in a revoked license, as well as any use of the materials beyond the scope of an unrevoked license, may constitute copyright infringement and publisher reserves the right to take any and all action to protect its copyright in the materials.
9. Warranties: Publisher makes no representations or warranties with respect to the licensed material.
10. Indemnity: You hereby indemnify and agree to hold harmless publisher and CCC, and their respective officers, directors, employees and agents, from and against any and all claims arising out of your use of the licensed material other than as specifically authorized pursuant to this license.
11. No Transfer of License: This license is personal to you and may not be sublicensed, assigned, or transferred by you to any other person without publisher's written permission.
12. No Amendment Except in Writing: This license may not be amended except in a writing signed by both parties (or, in the case of publisher, by CCC on publisher's behalf).
13. Objection to Contrary Terms: Publisher hereby objects to any terms contained in any purchase order, acknowledgment, check endorsement or other writing prepared by you, which terms are inconsistent with these terms and conditions or CCC's Billing and Payment terms and conditions. These terms and conditions, together with CCC's Billing and Payment terms and conditions (which are incorporated herein), comprise the entire agreement between you and publisher (and CCC) concerning this licensing transaction. In the event of any conflict between your obligations established by these terms and conditions and those established by

CCC's Billing and Payment terms and conditions, these terms and conditions shall control.

14. **Revocation:** Elsevier or Copyright Clearance Center may deny the permissions described in this License at their sole discretion, for any reason or no reason, with a full refund payable to you. Notice of such denial will be made using the contact information provided by you. Failure to receive such notice will not alter or invalidate the denial. In no event will Elsevier or Copyright Clearance Center be responsible or liable for any costs, expenses or damage incurred by you as a result of a denial of your permission request, other than a refund of the amount(s) paid by you to Elsevier and/or Copyright Clearance Center for denied permissions.

LIMITED LICENSE

The following terms and conditions apply only to specific license types:

15. **Translation:** This permission is granted for non-exclusive world **English** rights only unless your license was granted for translation rights. If you licensed translation rights you may only translate this content into the languages you requested. A professional translator must perform all translations and reproduce the content word for word preserving the integrity of the article.

16. **Posting licensed content on any Website:** The following terms and conditions apply as follows:
Licensing material from an Elsevier journal: All content posted to the web site must maintain the copyright information line on the bottom of each image; A hyper-text must be included to the Homepage of the journal from which you are licensing at <http://www.sciencedirect.com/science/journal/xxxxx> or the Elsevier homepage for books at <http://www.elsevier.com>; Central Storage: This license does not include permission for a scanned version of the material to be stored in a central repository such as that provided by Heron/XanEdu.

Licensing material from an Elsevier book: A hyper-text link must be included to the Elsevier homepage at <http://www.elsevier.com>. All content posted to the web site must maintain the copyright information line on the bottom of each image.

Posting licensed content on Electronic reserve: In addition to the above the following clauses are applicable: The web site must be password-protected and made available only to bona fide students registered on a relevant course. This permission is granted for 1 year only. You may obtain a new license for future website posting.

17. **For journal authors:** the following clauses are applicable in addition to the above:

Preprints:

A preprint is an author's own write-up of research results and analysis, it has not been peer-reviewed, nor has it had any other value added to it by a publisher (such as formatting, copyright, technical enhancement etc.). Authors can share their preprints anywhere at any time. Preprints should not be added to or enhanced in any way in order to appear more like, or to substitute for, the final versions of articles however authors can update their preprints on arXiv or RePEc with their Accepted Author Manuscript (see below).

If accepted for publication, we encourage authors to link from the preprint to their formal publication via its DOI. Millions of researchers have access to the formal publications on ScienceDirect, and so links will help users to find, access, cite and use the best available version. Please note that Cell Press, The Lancet and some society-owned have different preprint policies. Information on these policies is available on the journal homepage.

Accepted Author Manuscripts: An accepted author manuscript is the manuscript of an article that has been accepted for publication and which typically includes author-incorporated changes suggested during submission, peer review and editor-author communications.

Authors can share their accepted author manuscript:

- – immediately
 - via their non-commercial person homepage or blog
 - by updating a preprint in arXiv or RePEc with the accepted manuscript
 - via their research institute or institutional repository for internal institutional uses or as part of an invitation-only research collaboration work-group
 - directly by providing copies to their students or to research collaborators for their personal use
 - for private scholarly sharing as part of an invitation-only work group on commercial sites with which Elsevier has an agreement
- – after the embargo period

- via non-commercial hosting platforms such as their institutional repository
- via commercial sites with which Elsevier has an agreement

In all cases accepted manuscripts should:

- – link to the formal publication via its DOI
- – bear a CC-BY-NC-ND license - this is easy to do
- – if aggregated with other manuscripts, for example in a repository or other site, be shared in alignment with our hosting policy not be added to or enhanced in any way to appear more like, or to substitute for, the published journal article.

Published journal article (JPA): A published journal article (PJA) is the definitive final record of published research that appears or will appear in the journal and embodies all value-adding publishing activities including peer review co-ordination, copy-editing, formatting, (if relevant) pagination and online enrichment.

Policies for sharing publishing journal articles differ for subscription and gold open access articles:

Subscription Articles: If you are an author, please share a link to your article rather than the full-text.

Millions of researchers have access to the formal publications on ScienceDirect, and so links will help your users to find, access, cite, and use the best available version.

Theses and dissertations which contain embedded PJAs as part of the formal submission can be posted publicly by the awarding institution with DOI links back to the formal publications on ScienceDirect.

If you are affiliated with a library that subscribes to ScienceDirect you have additional private sharing rights for others' research accessed under that agreement. This includes use for classroom teaching and internal training at the institution (including use in course packs and courseware programs), and inclusion of the article for grant funding purposes.

Gold Open Access Articles: May be shared according to the author-selected end-user license and should contain a [CrossMark logo](#), the end user license, and a DOI link to the formal publication on ScienceDirect. Please refer to Elsevier's [posting policy](#) for further information.

18. **For book authors** the following clauses are applicable in addition to the above: Authors are permitted to place a brief summary of their work online only. You are not allowed to download and post the published electronic version of your chapter, nor may you scan the printed edition to create an electronic version. **Posting to a repository:** Authors are permitted to post a summary of their chapter only in their institution's repository.

19. **Thesis/Dissertation:** If your license is for use in a thesis/dissertation your thesis may be submitted to your institution in either print or electronic form. Should your thesis be published commercially, please reapply for permission. These requirements include permission for the Library and Archives of Canada to supply single copies, on demand, of the complete thesis and include permission for Proquest/UMI to supply single copies, on demand, of the complete thesis. Should your thesis be published commercially, please reapply for permission. Theses and dissertations which contain embedded PJAs as part of the formal submission can be posted publicly by the awarding institution with DOI links back to the formal publications on ScienceDirect.

Elsevier Open Access Terms and Conditions

You can publish open access with Elsevier in hundreds of open access journals or in nearly 2000 established subscription journals that support open access publishing. Permitted third party re-use of these open access articles is defined by the author's choice of Creative Commons user license. See our [open access license policy](#) for more information.

Terms & Conditions applicable to all Open Access articles published with Elsevier:

Any reuse of the article must not represent the author as endorsing the adaptation of the article nor should the article be modified in such a way as to damage the author's honour or reputation. If any changes have been made, such changes must be clearly indicated.

The author(s) must be appropriately credited and we ask that you include the end user license and a DOI link to the formal publication on ScienceDirect.

If any part of the material to be used (for example, figures) has appeared in our publication with credit or acknowledgement to another source it is the responsibility of the user to ensure their reuse complies with the terms and conditions determined by the rights holder.

Additional Terms & Conditions applicable to each Creative Commons user license:

CC BY: The CC-BY license allows users to copy, to create extracts, abstracts and new works from the Article, to alter and revise the Article and to make commercial use of the Article (including reuse and/or resale of the Article by commercial entities), provided the user gives appropriate credit (with a link to the formal publication through the relevant DOI), provides a link to the license, indicates if changes were made and the licensor is not represented as endorsing the use made of the work. The full details of the license are available at <http://creativecommons.org/licenses/by/4.0>.

CC BY NC SA: The CC BY-NC-SA license allows users to copy, to create extracts, abstracts and new works from the Article, to alter and revise the Article, provided this is not done for commercial purposes, and that the user gives appropriate credit (with a link to the formal publication through the relevant DOI), provides a link to the license, indicates if changes were made and the licensor is not represented as endorsing the use made of the work. Further, any new works must be made available on the same conditions. The full details of the license are available at <http://creativecommons.org/licenses/by-nc-sa/4.0>.

CC BY NC ND: The CC BY-NC-ND license allows users to copy and distribute the Article, provided this is not done for commercial purposes and further does not permit distribution of the Article if it is changed or edited in any way, and provided the user gives appropriate credit (with a link to the formal publication through the relevant DOI), provides a link to the license, and that the licensor is not represented as endorsing the use made of the work. The full details of the license are available at <http://creativecommons.org/licenses/by-nc-nd/4.0>. Any commercial reuse of Open Access articles published with a CC BY NC SA or CC BY NC ND license requires permission from Elsevier and will be subject to a fee.

Commercial reuse includes:

- – Associating advertising with the full text of the Article
- – Charging fees for document delivery or access
- – Article aggregation
- – Systematic distribution via e-mail lists or share buttons

Posting or linking by commercial companies for use by customers of those companies.

20. Other Conditions:

v1.8

Questions? customercare@copyright.com or +1-855-239-3415 (toll free in the US) or +1-978-646-2777.

Table des figures

1.1	Illustration du cycle du carbone en présentant les principaux processus affectant ce cycle	3
1.2	Représentation des trois processus biologiques affectant la $f\text{CO}_2$	8
1.3	Circulation thermohaline présentée par les courants marins chauds (en surface) en orange, et en bleu les courants froids (en profondeur)	9
1.4	Illustration du processus de l'Upwelling, ou remontée d'eau froide	11
1.5	Graphe représentant l'augmentation du CO_2 dans l'atmosphère associée à une augmentation de la $f\text{CO}_2$ de l'océan, et donc une diminution du pH (Station Mauna Loa)	12
2.1	Carte du monde représentant les lignes imaginaires de la terre	16
2.2	Flux de CO_2 dans l'océan Atlantique. Les valeurs négatives représentent les puits de CO_2 . Le carré noir représente l'embouchure du fleuve de l'Amazonie	16
2.3	Relation entre $f\text{CO}_2$ et le flux air-mer de CO_2 dans l'ouest de l'océan Atlantique tropical	17
2.4	La zone de convergence intertropicale	18
2.5	La position de l'ITCZ pendant le mois de juillet et le mois de janvier	18
2.6	Les principaux courants marins de la zone Atlantique tropical	19
2.7	Le Golfe de Guinée qui se situ dans la côte Ouest de l'Afrique	21
2.8	La différence de SSS entre janvier et juillet (maximum moins minimum) dans l'océan Atlantique tropical, les deux lignes noires en gras représentent les deux régions où la différence de SSS dépasse 1,5 unité	23
2.9	Nombre de données de $f\text{CO}_2$ de surface	26
3.1	Les principales étapes de la télédétection par satellite	29
3.2	Les deux modes de télédétection : passive et active	30
3.3	Les composante d'une onde électromagnétique	31
3.4	Spectre électromagnétique : longueur d'onde et fréquence de chaque bande	31
3.5	Signature spectrale de l'eau, du sol, et de la végétation	33
3.6	Types de rayonnements absorbés et réfléchis par la végétation et par l'eau	33
3.7	Moyenne mensuelle de Chl- <i>a</i> de Mai 2015 fournie par MODIS Aqua	37
3.8	Moyenne mensuelle de SST de Mai 2015 fournie par MODIS Aqua	37
3.9	Moyenne mensuelle de SSS de Mai 2015 fournie par Aquarius	38
4.1	Illustration d'un réseau de neurones avec : 1 couche d'entrée (input layer), 1 couche cachée (hidden layer), et 1 couche de sortie (output layer).	44

4.2	Représentation basique d'un neurone artificiel	45
4.3	Deux types d'architecture de réseau de neurones.	47
4.4	Architecture d'un RN de type MLP	48
4.5	Partitionnement de données en 3 classes	50
4.6	Déroulement de l'algorithme <i>K</i> -means en 6 itérations, pour une répartition en 3 classes	51
5.1	Algorithme général résumant les principales étapes de cette étude.	54
5.2	Variation de la fCO_2 en fonction de la bathymétrie pendant les quatre saisons.	56
5.3	Extrait de la BD <i>in situ</i> où l'on peut voir de très faibles valeurs de SSS (cercle noir).	57
5.4	Variation de la fCO_2 en fonction de la bathymétrie pendant les quatre saisons, après l'application des deux premiers filtres.	58
5.5	Variabilité de fCO_2 pendant la saison de l'été (Juin-Août) : les zones contenant de fortes valeurs de fCO_2 dans (a), et de faibles valeurs de fCO_2 dans (b).	59
5.6	Variabilité de fCO_2 , SST, et Chl- <i>a</i> pendant la saison de l'été, dans la zone entourée par un cercle noir dans la Figure 5.5a	59
5.7	Variabilité de fCO_2 et de Chl- <i>a</i> pendant la saison de l'été, dans la zone entourée par un cercle noir dans la Figure 5.5b	60
5.8	fCO_2 pendant la saison de l'été, après le premier prétraitement ; (a) : <i>variabilité en fonction de la distance à la masse terrestre</i> . (b) : <i>boîte à dispersion de fCO_2</i>	60
5.9	Distribution spatiale des données de fCO_2 de surface dans l'Atlantique tropical par année par mois	62
5.10	Représentation de la variabilité de SST <i>in situ</i> pendant la saison du printemps. (a) <i>Moyenne et écarts-type de SST de chaque année</i> , (b) <i>Distribution spatiale de SST</i> , (c) <i>Variation de SST en fonction de la latitude</i> , (d) <i>Variation de SST en fonction de la longitude</i>	64
5.11	Représentation de la variabilité de SSS <i>in situ</i> pendant la saison du printemps. (a) <i>Moyenne et écarts-type de SSS de chaque année</i> , (b) <i>Distribution spatiale de SSS</i> , (c) <i>Variation de SSS en fonction de la latitude</i> , (d) <i>Variation de SSS en fonction de la longitude</i>	66
5.12	Représentation de la variabilité de fCO_2 <i>in situ</i> pendant la saison du printemps. (a) <i>Moyenne et écarts-type de fCO_2 de chaque année</i> , (b) <i>Distribution spatiale de fCO_2</i> , (c) <i>Variation de fCO_2 en fonction de la latitude</i> , (d) <i>Variation de fCO_2 en fonction de la longitude</i>	68
5.13	Représentation de la variabilité de SST <i>in situ</i> pendant la saison de l'hiver. (a) <i>Moyenne et écarts-type de SST de chaque année</i> , (b) <i>Distribution spatiale de SST</i> , (c) <i>Variation de SST en fonction de la latitude</i> , (d) <i>Variation de SST en fonction de la longitude</i>	70
5.14	Représentation de la variabilité de SSS <i>in situ</i> pendant la saison de l'hiver. (a) <i>Moyenne et écarts-type de SSS de chaque année</i> , (b) <i>Distribution spatiale de SSS</i> , (c) <i>Variation de SSS en fonction de la latitude</i> , (d) <i>Variation de SSS en fonction de la longitude</i>	72

5.15	Représentation de la variabilité de $f\text{CO}_2$ <i>in situ</i> pendant la saison de l'hiver. (a) Moyenne et écarts-type de $f\text{CO}_2$ de chaque année, (b) Distribution spatiale de $f\text{CO}_2$, (c) Variation de $f\text{CO}_2$ en fonction de la latitude, (d) Variation de $f\text{CO}_2$ en fonction de la longitude	74
5.16	Données satellitaires de Chl-a mensuelles, fournies par le capteur MODIS-A , dans la zone 23,5°S-70°O à 23,5°N-15°E, entre décembre 2012 et novembre 2013, avec une résolution spatiale de 4 km × 4 km.	76
5.17	Histogramme des données de Chl-a extraites de la BD de la saison du printemps. (a) Avant application de la transformation 'Log10', (b) Après application de la transformation 'Log10'	77
5.18	Représentation de la variabilité de $\text{Log}_{10}(\text{Chl-a}$ satellitaire) pendant la saison du printemps. (a) Moyenne et écarts-type de Chl-a de chaque année, (b) Distribution spatiale de Chl-a, (c) Variation de Chl-a en fonction de la latitude, (d) Variation de Chl-a en fonction de la longitude	78
5.19	Représentation de la variabilité de $\text{Log}_{10}(\text{Chl-a}$ satellitaire) pendant la saison de l'hiver. (a) Moyenne et écarts-type de Chl-a de chaque année, (b) Distribution spatiale de Chl-a, (c) Variation de Chl-a en fonction de la latitude, (d) Variation de Chl-a en fonction de la longitude	80
5.20	Données satellitaires de SST mensuelles, fournies par le capteur MODIS-A , dans la zone 23,5°S-70°O à 23,5°N-15°E, entre décembre 2012 et novembre 2013, avec une résolution spatiale de 4 km × 4 km.	82
5.21	Données satellitaires de SSS mensuelles, fournies par le capteur SMOS , dans la zone 23,5°S-70°O à 23,5°N-15°E, entre décembre 2012 et novembre 2013, avec une résolution spatiale de 25 km × 25 km.	83
5.22	Extrait de la matrice MATLAB représentant la BD de construction des RNs. (a) matrice globale, (b) détail de chaque cellule de la colonne 'data' dans (a). Bathy. : Bathymétrie, Dist. mass. : Distance à la masse terrestre	85
5.23	L'application de la méthode <i>K-means</i> sur chaque ensemble de données.	87
5.24	Histogrammes de toutes les données de SST satellitaire par saison, avec des séparations entre les 3 classes (C1, C2, et C3).	88
5.25	Représentation spatiale des données <i>in situ</i> de chaque saison, divisées en trois classes.	89
5.26	L'étape de division de la BD de construction des RNs en 12 matrices.	90
6.1	Illustration de la 1e, la 2e, et la 20e itération (IT.) du déroulement de l'algorithme proposé pour choisir le RN optimal. N. bin. : neurone binaire	92
6.2	Apprentissage de la classe C1 de la saison du printemps : $f\text{CO}_2$ estimée par le RN en fonction de $f\text{CO}_2$ <i>in situ</i>	95
6.3	Apprentissage de la classe C1 de la saison du printemps : (a) histogramme de $f\text{CO}_2$, (b) relation entre les résidus et les valeurs de $f\text{CO}_2$, (c) relation entre le pourcentage de données (par rapport à la BD d'apprentissage) de chaque intervalle de l'histogramme, et les RMSE correspondants.	96
6.4	Validation de la classe C1 de la saison du printemps : $f\text{CO}_2$ estimée par le RN en fonction de $f\text{CO}_2$ <i>in situ</i>	98
6.5	Test de la classe C1 de la saison du printemps : $f\text{CO}_2$ calculée par le RN à partir des données satellitaire, en fonction de $f\text{CO}_2$ <i>in situ</i>	99

6.6	Test de la classe C1 de la saison du printemps : (a) <i>histogramme de fCO_2</i> , (b) <i>relation entre les résidus et les valeurs de fCO_2</i> , (c) <i>relation entre le pourcentage de données (par rapport à la BD de test) de chaque intervalle de l'histogramme, et les RMSE correspondants.</i>	100
6.7	Apprentissage de la classe C2 de la saison du printemps : fCO_2 estimée par le RN en fonction de fCO_2 <i>in situ.</i>	102
6.8	Apprentissage de la classe C2 de la saison du printemps : Résidus du jour 06/05/2009. (a) <i>positions géographiques des points</i> , (b) <i>variation de fCO_2 en fonction des heures</i> , (c) <i>relation entre les résidus et les valeurs de fCO_2</i>	103
6.9	Apprentissage de la classe C2 : (a) <i>positions géographiques des points (mars 2013)</i> , (b) <i>moyennes et écart-types de fCO_2 (pendant les mois de mars de toutes les années) dans la zone $21,5^\circ N-70^\circ O$ à $23,5^\circ N-65^\circ O$.</i>	104
6.10	Apprentissage de la classe C2 de la saison du printemps : (a) <i>histogramme de fCO_2</i> , (b) <i>relation entre les résidus et les valeurs de fCO_2</i> , (c) <i>relation entre le pourcentage de données (par rapport à la BD d'apprentissage) de chaque intervalle de l'histogramme, et les RMSE correspondants.</i>	105
6.11	Validation de la classe C2 de la saison du printemps : fCO_2 calculée par le RN en fonction de fCO_2 <i>in situ.</i>	107
6.12	Test de la classe C2 de la saison du printemps : fCO_2 calculée par le RN à partir des données satellitaires, en fonction de fCO_2 <i>in situ.</i>	108
6.13	Test de la classe C2 de la saison du printemps : (a) <i>histogramme de fCO_2</i> , (b) <i>relation entre les résidus et les valeurs de fCO_2</i> , (c) <i>relation entre le pourcentage de données (par rapport à la BD de test) de chaque intervalle de l'histogramme, et les RMSE correspondants.</i>	109
6.14	Apprentissage de la classe C3 de la saison du printemps : fCO_2 estimée par le RN en fonction de fCO_2 <i>in situ.</i>	111
6.15	Apprentissage de la classe C3 de la saison du printemps : (a) <i>histogramme de fCO_2</i> , (b) <i>relation entre les résidus et les valeurs de fCO_2</i> , (c) <i>relation entre le pourcentage de données (par rapport à la BD d'apprentissage) de chaque intervalle de l'histogramme, et les RMSE correspondants.</i>	112
6.16	Validation de la classe C3 de la saison du printemps : fCO_2 calculée par le RN en fonction de fCO_2 <i>in situ.</i>	114
6.17	Test de la classe C3 de la saison du printemps : fCO_2 calculée par le RN à partir des données satellitaires, en fonction de fCO_2 <i>in situ.</i>	115
6.18	Test de la classe C3 de la saison du printemps : (a) <i>histogramme de fCO_2</i> , (b) <i>relation entre les résidus et les valeurs de fCO_2</i> , (c) <i>relation entre le pourcentage de données (par rapport à la BD de test) de chaque intervalle de l'histogramme, et les RMSE correspondants.</i>	116
6.19	Apprentissage de la classe C1 de la saison de l'hiver : fCO_2 estimée par le RN en fonction de fCO_2 <i>in situ.</i>	118
6.20	Apprentissage de la classe C1 de la saison de l'hiver : (a) <i>histogramme de fCO_2</i> , (b) <i>relation entre les résidus et les valeurs de fCO_2</i> , (c) <i>relation entre le pourcentage de données (par rapport à la BD d'apprentissage) de chaque intervalle de l'histogramme, et les RMSE correspondants.</i>	119

6.21	Test de la classe C1 de la saison de l'hiver : fCO ₂ calculée par le RN à partir des données satellitaires, en fonction de fCO ₂ <i>in situ</i>	121
6.22	Test de la classe C1 de la saison de l'hiver : (a) <i>histogramme de fCO₂</i> , (b) <i>relation entre les résidus et les valeurs de fCO₂</i> , (c) <i>relation entre le pourcentage de données (par rapport à la BD de test) de chaque intervalle de l'histogramme, et les RMSE correspondants</i>	122
6.23	Apprentissage de la classe C2 de la saison de l'hiver : fCO ₂ estimée par le RN en fonction de fCO ₂ <i>in situ</i>	124
6.24	Apprentissage de la classe C2 de la saison de l'hiver : (a) <i>histogramme de fCO₂</i> , (b) <i>relation entre les résidus et les valeurs de fCO₂</i> , (c) <i>relation entre le pourcentage de données (par rapport à la BD d'apprentissage) de chaque intervalle de l'histogramme, et les RMSE correspondants</i>	125
6.25	Validation de la classe C2 de la saison de l'hiver : fCO ₂ estimée par le RN en fonction de fCO ₂ <i>in situ</i>	127
6.26	Test de la classe C2 de la saison de l'hiver : fCO ₂ calculée par le RN à partir des données satellitaires, en fonction de fCO ₂ <i>in situ</i>	128
6.27	Test de la classe C2 de la saison de l'hiver : (a) <i>relation entre les résidus et l'histogramme de fCO₂</i> , (b) <i>relation entre le pourcentage de données (par rapport à la BD de test) de chaque intervalle de l'histogramme, et les RMSE correspondants</i>	129
6.28	Apprentissage de la classe C3 de la saison d'hiver : fCO ₂ estimée par le RN en fonction de fCO ₂ <i>in situ</i>	131
6.29	Apprentissage de la classe C3 de la saison de l'hiver, entre le 30 et le 31 janvier 2012 : (a) <i>positions géographiques des points</i> , (b) <i>variabilités de fCO₂, SST, Chl-a, et de SSS en fonction des heures de la journée</i>	132
6.30	Apprentissage de la classe C3 de la saison de l'hiver : (a) <i>histogramme de fCO₂</i> , (b) <i>relation entre les résidus et les valeurs de fCO₂</i> , (c) <i>relation entre le pourcentage de données (par rapport à la BD d'apprentissage) de chaque intervalle de l'histogramme, et les RMSE correspondants</i>	133
6.31	Validation de la classe C3 de la saison d'hiver : fCO ₂ calculée par le RN en fonction de fCO ₂ <i>in situ</i>	135
6.32	Test de la classe C3 de la saison de l'hiver : fCO ₂ calculée par le RN à partir des données satellitaires, en fonction de fCO ₂ <i>in situ</i>	136
6.33	Test de la classe C3 de la saison de l'hiver : (a) <i>relation entre les résidus et l'histogramme de fCO₂</i> , (b) <i>relation entre le pourcentage de données (par rapport à la BD de test) de chaque intervalle de l'histogramme, et les RMSE correspondants</i>	137
7.1	Les deux principales étapes suivies pour le calcul de fCO ₂ à partir des données satellitaires (et assimilées).	140
7.2	Moyennes mensuelles de SST, SSS, Chl-a, et fCO ₂ entre juillet 2002 et décembre 2013.	141
7.3	Présentation de la variabilité annuelle de SST, SSS, Chl-a, et fCO ₂ , pour les années allant de 2002 à 2007.	143
7.4	Présentation de la variabilité annuelle de SST, SSS, Chl-a, et fCO ₂ , pour les années allant de 2008 à 2013.	144

7.5	Cartes mensuelles de SST satellitaires, fournies par le capteur MODIS-A , dans la zone 23,5°S-70°O à 23,5°N-15°E, entre décembre 2012 et novembre 2013, avec une résolution spatiale de 4 km × 4 km.	147
7.6	Cartes mensuelles de Log ₁₀ (Chl- <i>a</i>) satellitaires, fournies par le capteur MODIS-A , dans la zone 23,5°S-70°O à 23,5°N-15°E, entre décembre 2012 et novembre 2013, avec une résolution spatiale de 4 km × 4 km.	148
7.7	Cartes mensuelles de SSS satellitaires, fournies par le capteur SMOS , dans la zone 23,5°S-70°O à 23,5°N-15°E, entre décembre 2012 et novembre 2013, avec une résolution spatiale de 25 km × 25 km.	149
7.8	Résultats de l'interpolation de fCO ₂ mensuelles à partir des données satellitaires, produites par les RNs construits, dans la zone 23,5°S-70°O à 23,5°N-15°E, entre décembre 2012 et novembre 2013, avec une résolution spatiale de 4 km × 4 km.	150
A.1	Version MATLAB utilisée, et spécifications techniques du PC.	174
A.2	Exemple de structuration par année de fichiers satellitaires.	178
B.1	Représentation de la variabilité de SST <i>in situ</i> pendant la saison de l'été. (a) Moyenne et écarts-type de SST de chaque année, (b) Distribution spatiale de SST, (c) Variation de SST en fonction de la latitude, (d) Variation de SST en fonction de la longitude	204
B.2	Représentation de la variabilité de SSS <i>in situ</i> pendant la saison de l'été. (a) Moyenne et écarts-type de SSS de chaque année, (b) Distribution spatiale de SSS, (c) Variation de SSS en fonction de la latitude, (d) Variation de SSS en fonction de la longitude	205
B.3	Représentation de la variabilité de fCO ₂ <i>in situ</i> pendant la saison de l'été. (a) Moyenne et écarts-type de fCO ₂ de chaque année, (b) Distribution spatiale de fCO ₂ , (c) Variation de fCO ₂ en fonction de la latitude, (d) Variation de fCO ₂ en fonction de la longitude	206
B.4	Représentation de la variabilité de Log ₁₀ (Chl- <i>a</i> satellitaire) pendant la saison de l'été. (a) Moyenne et écarts-type de Chl- <i>a</i> de chaque année, (b) Distribution spatiale de Chl- <i>a</i> , (c) Variation de Chl- <i>a</i> en fonction de la latitude, (d) Variation de Chl- <i>a</i> en fonction de la longitude	207
B.5	Représentation de la variabilité de SST <i>in situ</i> pendant la saison de l'automne. (a) Moyenne et écarts-type de SST de chaque année, (b) Distribution spatiale de SST, (c) Variation de SST en fonction de la latitude, (d) Variation de SST en fonction de la longitude	211
B.6	Représentation de la variabilité de SSS <i>in situ</i> pendant la saison de l'automne. (a) Moyenne et écarts-type de SSS de chaque année, (b) Distribution spatiale de SSS, (c) Variation de SSS en fonction de la latitude, (d) Variation de SSS en fonction de la longitude	212
B.7	Représentation de la variabilité de fCO ₂ <i>in situ</i> pendant la saison de l'automne. (a) Moyenne et écarts-type de fCO ₂ de chaque année, (b) Distribution spatiale de fCO ₂ , (c) Variation de fCO ₂ en fonction de la latitude, (d) Variation de fCO ₂ en fonction de la longitude	213

B.8	Représentation de la variabilité de $\text{Log}_{10}(\text{Chl-}a \text{ satellitaire})$ pendant la saison de l'automne. (a) Moyenne et écarts-type de <i>Chl-a</i> de chaque année, (b) Distribution spatiale de <i>Chl-a</i> , (c) Variation de <i>Chl-a</i> en fonction de la latitude, (d) Variation de <i>Chl-a</i> en fonction de la longitude	214
C.1	Apprentissage de la classe C1C2 de la saison de l'été : fCO_2 estimée par le RN en fonction de fCO_2 <i>in situ</i>	216
C.2	Apprentissage de la classe C1C2 de la saison de l'été : (a) histogramme de fCO_2 , (b) relation entre les résidus et les valeurs de fCO_2 , (c) relation entre le pourcentage de données (par rapport à la BD d'apprentissage) de chaque intervalle de l'histogramme, et les RMSE correspondants.	217
C.3	Validation de la classe C1C2 de la saison de l'été : fCO_2 estimée par le RN en fonction de fCO_2 <i>in situ</i>	219
C.4	Test de la classe C1C2 de la saison de l'été : fCO_2 calculée par le RN à partir des données satellitaire, en fonction de fCO_2 <i>in situ</i>	220
C.5	Test de la classe C1C2 de la saison de l'été : (a) histogramme de fCO_2 , (b) relation entre les résidus et les valeurs de fCO_2 , (c) relation entre le pourcentage de données (par rapport à la BD d'apprentissage) de chaque intervalle de l'histogramme, et les RMSE correspondants.	221
C.6	Apprentissage de la classe C3 de la saison de l'été : fCO_2 estimée par le RN en fonction de fCO_2 <i>in situ</i>	224
C.7	Apprentissage de la classe C3 de la saison de l'été : (a) histogramme de fCO_2 , (b) relation entre les résidus et les valeurs de fCO_2 , (c) relation entre le pourcentage de données (par rapport à la BD d'apprentissage) de chaque intervalle de l'histogramme, et les RMSE correspondants.	225
C.8	Validation de la classe C3 de la saison de l'été : fCO_2 estimée par le RN en fonction de fCO_2 <i>in situ</i>	226
C.9	Test de la classe C3 de la saison de l'été : fCO_2 calculée par le RN à partir des données satellitaire, en fonction de fCO_2 <i>in situ</i>	227
C.10	Test de la classe C3 de la saison de l'été : (a) histogramme de fCO_2 , (b) relation entre les résidus et les valeurs de fCO_2 , (c) relation entre le pourcentage de données (par rapport à la BD d'apprentissage) de chaque intervalle de l'histogramme, et les RMSE correspondants.	228
C.11	Apprentissage de la classe C1C2 de la saison de l'automne : fCO_2 estimée par le RN en fonction de fCO_2 <i>in situ</i>	231
C.12	Apprentissage de la classe C1C2 de la saison de l'automne : (a) histogramme de fCO_2 , (b) relation entre les résidus et les valeurs de fCO_2 , (c) relation entre le pourcentage de données (par rapport à la BD d'apprentissage) de chaque intervalle de l'histogramme, et les RMSE correspondants.	232
C.13	Validation de la classe C1C2 de la saison de l'automne : fCO_2 estimée par le RN en fonction de fCO_2 <i>in situ</i>	233
C.14	Test de la classe C1C2 de la saison de l'automne : fCO_2 calculée par le RN à partir des données satellitaire, en fonction de fCO_2 <i>in situ</i>	234

C.15	Test de la classe C1C2 de la saison de l'automne : (a) <i>histogramme de fCO_2</i> , (b) <i>relation entre les résidus et les valeurs de fCO_2</i> , (c) <i>relation entre le pourcentage de données (par rapport à la BD d'apprentissage) de chaque intervalle de l'histogramme, et les RMSE correspondants.</i>	235
C.16	Apprentissage de la classe C3 de la saison de l'automne : fCO_2 estimée par le RN en fonction de fCO_2 <i>in situ.</i>	238
C.17	Apprentissage de la classe C3 de la saison de l'automne : (a) <i>histogramme de fCO_2</i> , (b) <i>relation entre les résidus et les valeurs de fCO_2</i> , (c) <i>relation entre le pourcentage de données (par rapport à la BD d'apprentissage) de chaque intervalle de l'histogramme, et les RMSE correspondants.</i>	239
C.18	Validation de la classe C3 de la saison de l'automne : fCO_2 estimée par le RN en fonction de fCO_2 <i>in situ.</i>	241
C.19	Test de la classe C3 de la saison de l'automne : fCO_2 calculée par le RN à partir des données satellitaire, en fonction de fCO_2 <i>in situ.</i>	242
C.20	Test de la classe C3 de la saison de l'automne : (a) <i>histogramme de fCO_2</i> , (b) <i>relation entre les résidus et les valeurs de fCO_2</i> , (c) <i>relation entre le pourcentage de données (par rapport à la BD d'apprentissage) de chaque intervalle de l'histogramme, et les RMSE correspondants.</i>	243

Liste des tableaux

4.1	Présentation des trois fonctions d'activations les plus utilisées. * <i>tangente-sigmoïde</i>	46
5.1	Nombre de données de fCO ₂ de surface (extraites de la BD SOCAT V.3, de 23,5°S-70°O à 23,5°N-15°E) par mois et par saison	61
5.2	Paramètres statistiques de SST de toutes les années pendant la saison du printemps.	63
5.3	Paramètres statistiques de SSS de toutes les années pendant la saison du printemps.	65
5.4	Paramètres statistiques de fCO ₂ de toutes les années pendant la saison du printemps.	67
5.5	Paramètres statistiques de SST de toutes les années pendant la saison de l'hiver.	69
5.6	Paramètres statistiques de SSS de toutes les années pendant la saison de l'hiver.	71
5.7	Paramètres statistiques de fCO ₂ de toutes les années pendant la saison de l'hiver.	73
5.8	Paramètres statistiques de Log ₁₀ (Chl- <i>a</i> satellitaire) de toutes les années pendant la saison du printemps.	77
5.9	Paramètres statistiques de Chl- <i>a</i> satellitaire de toutes les années pendant la saison du printemps.	79
5.10	Paramètres statistiques de Log ₁₀ (Chl- <i>a</i> satellitaire) de toutes les années pendant la saison de l'hiver.	79
5.11	Paramètres statistiques de Chl- <i>a</i> satellitaire de toutes les années pendant la saison de l'hiver.	81
5.12	Les proportions des données <i>in situ</i> perdues après l'utilisation des données satellitaires de Chl- <i>a</i> pour la construction des RNs.	85
5.13	Moyennes et écarts-types de la SST satellitaire (de toutes les années; 2002/2003 -> 2014) par saison.	86
5.14	Récapitulatif des résultats de la classification des données satellitaires. Les écarts-types maximaux sont en rouge. *Nombre de points	87
5.15	Nombre de points de données de chaque classe de chaque saison.	89
6.1	Apprentissage de la classe C1 de la saison du printemps : RMSE et biais correspondants à chaque année. * nombre de points, ** pourcentage de données par rapport à la BD d'apprentissage	94

6.2	Apprentissage de la classe C1 de la saison du printemps : relation entre le nombre de points et le RMSE, pour chaque intervalle de l'histogramme de fCO ₂ . * nombre de points ** pourcentage de données par rapport au nombre total de données de la BD d'apprentissage	97
6.3	Validation de la classe C1 de la saison du printemps : RMSE est biais correspondant à chaque année. * nombre de points, ** pourcentage de données par rapport à la BD de validation	97
6.4	Test de la classe C1 de la saison du printemps : RMSE et biais (de fCO ₂ , SST, et SSS) correspondants à chaque année. * nombre de points, ** pourcentage de données par rapport à la BD de test	98
6.5	Test de la classe C1 de la saison du printemps : relation entre le nombre de points et le RMSE, pour chaque intervalle de l'histogramme de fCO ₂ . * nombre de points ** pourcentage de données par rapport au nombre total de données de la BD de test	101
6.6	Apprentissage de la classe C2 de la saison du printemps : RMSE et biais correspondants à chaque année. * nombre de points, ** pourcentage de données par rapport à la BD d'apprentissage	102
6.7	Apprentissage de la classe C2 de la saison du printemps : relation entre le nombre de points et le RMSE, pour chaque intervalle de l'histogramme de fCO ₂ . * nombre de points ** pourcentage de données par rapport au nombre total de données de la BD d'apprentissage	106
6.8	Validation de la classe C2 de la saison du printemps : RMSE et biais correspondants à chaque année. * nombre de points de données ** pourcentage de données par rapport au nombre total de données de la BD de validation	106
6.9	Test de la classe C2 de la saison du printemps : RMSE et biais (de fCO ₂ , SST, et SSS) correspondant à chaque année.	108
6.10	Test de la classe C2 de la saison du printemps : relation entre le nombre de points et le RMSE, pour chaque intervalle de l'histogramme de fCO ₂ . * nombre de points ** pourcentage de données par rapport au nombre total de données de la BD de test	110
6.11	Apprentissage de la classe C3 de la saison du printemps : RMSE et biais correspondants à chaque année. * pourcentage de données par rapport à la taille de la BD d'apprentissage	111
6.12	Apprentissage de la classe C3 de la saison du printemps : relation entre le nombre de points et le RMSE, pour chaque intervalle de l'histogramme de fCO ₂ . * nombre de points ** pourcentage de données par rapport au nombre total de données de la BD d'apprentissage	113
6.13	Validation de la classe C3 de la saison du printemps : RMSE et biais correspondants à chaque année. * nombre de points de données ** pourcentage de données par rapport au nombre total de données de la BD de validation	113
6.14	Test de la classe C3 de la saison du printemps : RMSE et biais (de fCO ₂ , SST, et SSS) correspondant à chaque année.	115

6.15	Test de la classe C3 de la saison du printemps : relation entre le nombre de points et le RMSE, pour chaque intervalle de l'histogramme de fCO ₂ . * nombre de points ** pourcentage de données par rapport au nombre total de données de la BD de test	117
6.16	Apprentissage de la classe C1 de la saison de l'hiver : RMSE et biais correspondants à chaque année. * nombre de points, ** pourcentage de données par rapport à la BD d'apprentissage	117
6.17	Apprentissage de la classe C1 de la saison de l'hiver : relation entre le nombre de points et le RMSE, pour chaque intervalle de l'histogramme de fCO ₂ . * nombre de points ** pourcentage de données par rapport au nombre total de données de la BD d'apprentissage	120
6.18	Test de la classe C1 de la saison de l'hiver : RMSE et biais (de fCO ₂ , SST, et SSS) correspondants à chaque année. * nombre de points, ** pourcentage de données par rapport à la BD de test	120
6.19	Test de la classe C1 de la saison de l'hiver : relation entre le nombre de points et le RMSE, pour chaque intervalle de l'histogramme de fCO ₂ . * nombre de points ** pourcentage de données par rapport au nombre total de données de la BD de test	123
6.20	Apprentissage de la classe C2 de la saison de l'hiver : RMSE et biais correspondants à chaque année. * nombre de points, ** pourcentage de données par rapport à la BD d'apprentissage	123
6.21	Apprentissage de la classe C2 de la saison de l'hiver : relation entre le nombre de points et le RMSE, pour chaque intervalle de l'histogramme de fCO ₂ . * nombre de points ** pourcentage de données par rapport au nombre total de données de la BD d'apprentissage	126
6.22	Validation de la classe C2 de la saison de l'hiver : RMSE et biais correspondants à chaque année. * nombre de points de données ** pourcentage de données par rapport au nombre total de données de la BD de validation	126
6.23	Test de la classe C2 de la saison de l'hiver : RMSE et biais (de fCO ₂ , SST, et SSS) correspondants à chaque année. * nombre de points, ** pourcentage de données par rapport à la BD de test	128
6.24	Test de la classe C2 de la saison de l'hiver : relation entre le nombre de points et le RMSE, pour chaque intervalle de l'histogramme de fCO ₂ . * nombre de points ** pourcentage de données par rapport au nombre total de données de la BD de test	130
6.25	Apprentissage de la classe C3 de la saison de l'hiver : RMSE et biais correspondants à chaque année. * pourcentage de données par rapport à la taille de la BD d'apprentissage	130
6.26	Apprentissage de la classe C3 de la saison de l'hiver : relation entre le nombre de points et le RMSE, pour chaque intervalle de l'histogramme de fCO ₂ . * nombre de points ** pourcentage de données par rapport au nombre total de données de la BD d'apprentissage	134
6.27	Validation de la classe C3 de la saison de l'hiver : RMSE et biais correspondants à chaque année. * nombre de points de données ** pourcentage de données par rapport au nombre total de données de la BD de validation	134

6.28	Test de la classe C3 de la saison de l'hiver : RMSE et biais (de fCO ₂ , SST, et SSS) correspondants à chaque année. * nombre de points, ** pourcentage de données par rapport à la BD de test	136
6.29	Test de la classe C3 de la saison de l'hiver : relation entre le nombre de points et le RMSE, pour chaque intervalle de l'histogramme de fCO ₂ . * nombre de points ** pourcentage de données par rapport au nombre total de données de la BD de test	138
7.1	Nombre de points de données propre à chaque saison, pour la période allant de juillet 2002 à décembre 2013.	141
7.2	Récapitulatif des résultats obtenus dans cette étude. La colonne 'Test (seuils)' contient les RMSEs recalculés en appliquant un seuil de 0,5 °C sur la SST satellitaire, et un seuil de 0,4 sur la SSS satellitaire.	154
B.1	Paramètres statistiques de SST de toutes les années pendant la saison de l'été.	201
B.2	Paramètres statistiques de SSS de toutes les années pendant la saison de l'été.	202
B.3	Paramètres statistiques de fCO ₂ de toutes les années pendant la saison de l'été.	202
B.4	Paramètres statistiques de Log ₁₀ (Chl- <i>a</i> satellitaire) de toutes les années pendant la saison de l'été.	203
B.5	Paramètres statistiques de Chl- <i>a</i> satellitaire de toutes les années pendant la saison de l'été.	203
B.6	Paramètres statistiques de SST de toutes les années pendant la saison de l'automne.	208
B.7	Paramètres statistiques de SSS de toutes les années pendant la saison de l'automne.	209
B.8	Paramètres statistiques de fCO ₂ de toutes les années pendant la saison de l'automne.	209
B.9	Paramètres statistiques de Log ₁₀ (Chl- <i>a</i> satellitaire) de toutes les années pendant la saison de l'automne.	210
B.10	Paramètres statistiques de Chl- <i>a</i> satellitaire de toutes les années pendant la saison de l'automne.	210
C.1	Apprentissage de la classe C1C2 de la saison de l'été : RMSE et biais correspondants à chaque année. * nombre de points, ** pourcentage de données par rapport à la BD d'apprentissage	216
C.2	Apprentissage de la classe C1C2 de la saison de l'été : relation entre le nombre de points et le RMSE, pour chaque intervalle de l'histogramme de fCO ₂ . * nombre de points ** pourcentage de données par rapport au nombre total de données de la BD d'apprentissage	218
C.3	Validation de la classe C1C2 de la saison de l'été : RMSE et biais correspondant à chaque année. * nombre de points, ** pourcentage de données par rapport à la BD de validation	219

C.4	Test de la classe C1C2 de la saison de l'été : RMSE et biais (de fCO ₂ , SST, et SSS) correspondants à chaque année. * nombre de points, ** pourcentage de données par rapport à la BD de test	220
C.5	Test de la classe C1C2 de la saison de l'été : relation entre le nombre de points et le RMSE, pour chaque intervalle de l'histogramme de fCO ₂ . * nombre de points ** pourcentage de données par rapport au nombre total de données de la BD d'apprentissage	222
C.6	Apprentissage de la classe C3 de la saison de l'été : RMSE et biais correspondants à chaque année. * nombre de points, ** pourcentage de données par rapport à la BD d'apprentissage	223
C.7	Apprentissage de la classe C3 de la saison de l'été : relation entre le nombre de points et le RMSE, pour chaque intervalle de l'histogramme de fCO ₂ . * nombre de points ** pourcentage de données par rapport au nombre total de données de la BD d'apprentissage	223
C.8	Validation de la classe C3 de la saison de l'été : RMSE est biais correspondant à chaque année. * nombre de points, ** pourcentage de données par rapport à la BD de validation	226
C.9	Test de la classe C3 de la saison de l'été : RMSE et biais (de fCO ₂ , SST, et SSS) correspondants à chaque année. * nombre de points, ** pourcentage de données par rapport à la BD de test	227
C.10	Test de la classe C3 de la saison de l'été : relation entre le nombre de points et le RMSE, pour chaque intervalle de l'histogramme de fCO ₂ . * nombre de points ** pourcentage de données par rapport au nombre total de données de la BD d'apprentissage	229
C.11	Apprentissage de la classe C1C2 de la saison de l'automne : RMSE et biais correspondants à chaque année. * nombre de points, ** pourcentage de données par rapport à la BD d'apprentissage	230
C.12	Apprentissage de la classe C1C2 de la saison de l'automne : relation entre le nombre de points et le RMSE, pour chaque intervalle de l'histogramme de fCO ₂ . * nombre de points ** pourcentage de données par rapport au nombre total de données de la BD d'apprentissage	231
C.13	Validation de la classe C1C2 de la saison de l'automne : RMSE est biais correspondant à chaque année. * nombre de points, ** pourcentage de données par rapport à la BD de validation	233
C.14	Test de la classe C1C2 de la saison de l'automne : RMSE et biais (de fCO ₂ , SST, et SSS) correspondants à chaque année. * nombre de points, ** pourcentage de données par rapport à la BD de test	234
C.15	Test de la classe C1C2 de la saison de l'automne : relation entre le nombre de points et le RMSE, pour chaque intervalle de l'histogramme de fCO ₂ . * nombre de points ** pourcentage de données par rapport au nombre total de données de la BD d'apprentissage	236
C.16	Apprentissage de la classe C3 de la saison de l'automne : RMSE et biais correspondants à chaque année. * nombre de points, ** pourcentage de données par rapport à la BD d'apprentissage	237

C.17 Apprentissage de la classe C3 de la saison de l'automne : relation entre le nombre de points et le RMSE, pour chaque intervalle de l'histogramme de fCO ₂ . * nombre de points ** pourcentage de données par rapport au nombre total de données de la BD d'apprentissage	238
C.18 Validation de la classe C3 de la saison de l'automne : RMSE est biais correspondant à chaque année. * nombre de points, ** pourcentage de données par rapport à la BD de validation	240
C.19 Test de la classe C3 de la saison de l'automne : RMSE et biais (de fCO ₂ , SST, et SSS) correspondants à chaque année. * nombre de points, ** pourcentage de données par rapport à la BD de test	242
C.20 Test de la classe C3 de la saison de l'automne : relation entre le nombre de points et le RMSE, pour chaque intervalle de l'histogramme de fCO ₂ . * nombre de points ** pourcentage de données par rapport au nombre total de données de la BD d'apprentissage	244
D.1 Paramètres statistiques de SST satellitaire pendant la saison de l'hiver entre 2002 et 2013 (de 23,5°S-70°O à 23,5°N-15°E).	245
D.2 Paramètres statistiques de SSS satellitaire pendant la saison de l'hiver entre 2002 et 2013 (de 23,5°S-70°O à 23,5°N-15°E).	246
D.3 Paramètres statistiques de Chl- <i>a</i> satellitaire pendant la saison de l'hiver entre 2002 et 2013 (de 23,5°S-70°O à 23,5°N-15°E).	246
D.4 Paramètres statistiques de fCO ₂ satellitaire pendant la saison de l'hiver entre 2002 et 2013 (de 23,5°S-70°O à 23,5°N-15°E).	247
D.5 Paramètres statistiques de SST satellitaire pendant la saison du printemps entre 2003 et 2013 (de 23,5°S-70°O à 23,5°N-15°E).	247
D.6 Paramètres statistiques de SSS satellitaire pendant la saison du printemps entre 2003 et 2013 (de 23,5°S-70°O à 23,5°N-15°E).	248
D.7 Paramètres statistiques de Chl- <i>a</i> satellitaire pendant la saison du printemps entre 2003 et 2013 (de 23,5°S-70°O à 23,5°N-15°E).	248
D.8 Paramètres statistiques de fCO ₂ satellitaire pendant la saison du printemps entre 2003 et 2013 (de 23,5°S-70°O à 23,5°N-15°E).	249
D.9 Paramètres statistiques de SST satellitaire pendant la saison de l'été entre 2002 et 2013 (de 23,5°S-70°O à 23,5°N-15°E).	249
D.10 Paramètres statistiques de SSS satellitaire pendant la saison de l'été entre 2002 et 2013 (de 23,5°S-70°O à 23,5°N-15°E).	250
D.11 Paramètres statistiques de Chl- <i>a</i> satellitaire pendant la saison de l'été entre 2002 et 2013 (de 23,5°S-70°O à 23,5°N-15°E).	250
D.12 Paramètres statistiques de fCO ₂ satellitaire pendant la saison de l'été entre 2002 et 2013 (de 23,5°S-70°O à 23,5°N-15°E).	251
D.13 Paramètres statistiques de SST satellitaire pendant la saison de l'automne entre 2002 et 2013 (de 23,5°S-70°O à 23,5°N-15°E).	251
D.14 Paramètres statistiques de SSS satellitaire pendant la saison de l'automne entre 2002 et 2013 (de 23,5°S-70°O à 23,5°N-15°E).	252
D.15 Paramètres statistiques de Chl- <i>a</i> satellitaire pendant la saison de l'automne entre 2002 et 2013 (de 23,5°S-70°O à 23,5°N-15°E).	252

D.16 Paramètres statistiques de fCO₂ satellitaire pendant la saison de l'automne
entre 2002 et 2013 (de 23,5°S-70°O à 23,5°N-15°E). 253

TRAITEMENT AUTOMATIQUE DE DONNEES OCEANOGRAPHIQUES POUR L'INTERPOLATION DE LA fCO₂ DE SURFACE DANS L'OCEAN ATLANTIQUE TROPICAL, EN UTILISANT LES DONNEES SATELLITAIRES.

Doctorant : MOUSSA Hadjer / **Discipline :** Informatique
Directeur de thèse : GOYET Catherine

Résumé :

Ce travail de thèse consiste à utiliser les données satellitaires de SST (température de surface), SSS (salinité de surface), et Chl-*a* (chlorophylle-*a*), pour l'interpolation de la fugacité du CO₂ (fCO₂) dans la couche de surface de l'océan Atlantique tropical, pour les saisons de la période 2002-2013. Trois types de données ont été utilisés : *in situ* (BD (base de données) SOCAT V.3) ; satellitaires (capteurs : MODIS-A, Sea-WIFS, et SMOS) ; et assimilées (BD SODA V.2.2.4). La première étape était la classification des données en se basant sur la SST. La deuxième étape était l'interpolation de la fCO₂ (pour chaque classe de chaque saison), en utilisant des RNs (réseaux de neurones artificiels) de type *feedforward*, avec un apprentissage de type *backpropagation*. Les résultats obtenus (RMSEs (*root mean square error*) variant de 8,8 à 15,7 µatm) permettent de confirmer l'importance de : traiter les saisons séparément, classifier les données, et choisir le meilleur RN en fonction des résultats de la généralisation. Ceci a permis l'élaboration de 138 fichiers CSV (*Comma-separated values*) de fCO₂ mensuelle, avec une résolution de 4 km x 4 km, pour la période allant de juillet 2002 à décembre 2013.

Mots-clés : Fugacité, CO₂, Atlantique, température, salinité, chlorophylle, satellite, interpolation, feedforward, classification.

AUTOMATIC PROCESSING OF OCEANOGRAPHIC DATA IN ORDER TO INTERPOLATE SEA SURFACE fCO₂ IN THE TROPICAL ATLANTIC OCEAN, USING SATELLITE DATA.

PhD student: MOUSSA Hadjer / **Discipline:** Computer Science
Supervisor: GOYET Catherine

Abstract:

This thesis work consists of using satellite data of SST (sea surface temperature), SSS (sea surface salinity), and Chl-*a* (chlorophyll-*a*), in order to interpolate the CO₂ fugacity (fCO₂) in the surface of the tropical Atlantic ocean, for seasons of the period 2002-2013. Three data types were used: *in situ* (SOCAT V.3 DB (database)); satellite (MODIS-A, Sea-WIFS, and SMOS sensors); and assimilated (SODA V.2.2.4 DB). The first step was the data classification based on SST. The second step was the fCO₂ interpolation (for each class of each season), using *feedforward* NNs (artificial neural networks) with a *backpropagation* learning method. Obtained results (RMSEs (*root mean square error*) between 8,8 and 15,7 µatm) confirm the importance of: process each season separately, pass through data classification step, and choose the best NN on the basis of generalization step results. This allowed the development of 138 monthly fCO₂ CSV (*Comma-separated values*) file, with 4 km x 4 km spatial resolution, for the period from July 2002 to December 2013.

Key-words: Fugacity, CO₂, Atlantic, temperature, salinity, chlorophyll, satellite, interpolation, feedforward, classification.