

UNIVERSITÉ CHEIKH ANTA DIOP DE DAKAR



ÉCOLE DOCTORALE :

MATHÉMATIQUES ET INFORMATIQUE

SPÉCIALITÉ :

MODÉLISATION DES SYSTÈMES COMPLEXES

Année : 2013 - N° d'ordre : TH1843

Thèse de Doctorat Unique

Pour obtenir le grade de Docteur de :

L'UNIVERSITÉ CHEIKH ANTA DIOP DE DAKAR

Sujet de la Thèse :

Classification neuronale de données satellitaires

pour la caractérisation de la variabilité de l'upwelling Sénégal-Mauritanien

Présentée et soutenue par : FARIKOU OUSMANE

Date de soutenance : 14 Décembre 2013 à Dakar

JURY

Président :	Benjamin MAMPASSI	Professeur titulaire	UCAD - Dakar
Rapporteurs :	Amadou Tahirou DIAW	Professeur titulaire	UCAD - Dakar
	Papa NGOM	Maître de conférences	UCAD - Dakar
	Bassirou DIAW	Chercheur	CROD - Dakar
Examineur :	Michel CREPON	Directeur de recherche	CNRS Paris - France
Directeurs de thèse :	Salam SAWADOGO	Maître de conférences	EPT – Thiès
	Awa NIANG	Maître de conférences	ESP - UCAD - Dakar

Remerciements :

Tout d'abord je loue **DIEU LE TOUT PUISSANT** de m'avoir donné la force, la patience et le courage d'accomplir ce travail. J'exprime toute ma gratitude à toutes les personnes qui m'ont permis de réaliser ce travail.

Ce travail de thèse a été élaboré en alternance entre le Laboratoire de Traitement de l'Information (LTI) de l'École Supérieure Polytechnique (ESP) de l'Université Cheikh Anta de Dakar, Laboratoire d'Océanographie et du Climat, Expérimentations et approches Numériques (LOCEAN) de Paris 6 (France) et l'Institut des Sciences et Techniques d'Abéché (IUSTA) au Tchad.

En premier lieu, je tiens à remercier mon directeur de thèse Salam SAWADOGO d'avoir accepté de diriger ce travail et de m'avoir mis en contact par le biais de la collaboration qui existe entre Les Laboratoires de Dakar (Sénégal) et LOCEAN de Paris (France) avec l'équipe MMSA du LOCEAN. Je remercie aussi Awa NIANG du Laboratoire de Traitement de l'Information (LTI) pour ses critiques et conseils pertinents.

Je tiens à remercier ici tout le personnel de l'équipe MMSA du LOCEAN qui m'ont apporté idées, conseils et aides dans la réalisation de ce travail. Plus particulièrement, je souhaite remercier solennellement Sylvie THYRIA qui m'a accueilli dans son équipe pendant mes deux séjours à Paris, par la qualité de son apport, sa grande disponibilité et l'intérêt qu'elle a porté à ce travail.

Je tiens à remercier officiellement Mr Michel CREPON qui m'a, non seulement accueilli au LOCEAN au sein de l'équipe MMSA et mis à ma disposition des outils de travail de qualité mais aussi qui a toujours apporté son œil expert et rigoureux tout au long de ce travail. Il a beaucoup contribué à l'interprétation géophysique des résultats obtenus, orienté les démarches scientifiques en apportant sa grande connaissance en océanographie et son expérience inépuisable. Si vous lisez ce manuscrit, c'est par ce que Mr Michel CREPON s'est investi personnellement pour l'aboutissement de ce travail, j'en suis très reconnaissant et j'espère continuer avec lui d'autres aventures scientifiques.

Je tiens à remercier le Professeur Benjamin MAMPASSI pour m'avoir fait l'honneur d'être le président du jury de ma thèse.

Je tiens à remercier les Professeurs Amadou Tahirou DIAW, Papa NGOM et le Docteur Bassirou DIAW pour m'avoir fait l'honneur d'être les rapporteurs et membres du jury de ma thèse. Veuillez trouver la manifestation de mes remerciements et de mon plus profond respect.

Merci aux personnes avec qui j'ai travaillé et collaboré et qui m'ont permis d'avancer et d'apprendre : Carlos MEJIA et Julien BRAJARD.

Un grand merci à tous les collaborateurs rencontrés au LOCEAN : Julien VINCENT, Yves DANDONNEAU, Severine ALVAIN.

Un grand merci à tous les membres du laboratoire LTI, en particulier son directeur le Professeur Claude LISHOU. Merci à tous les responsables et personnels de l'ESP.

Je remercie l'Institut des Sciences et Techniques d'Abéché (IUSTA) qui a financé cette thèse dont le coût estimatif de ces 5 ans avoisine 7.000.000F : coût de transports aérien(N'djamèna-Dakar-Paris), coût d'hébergement(Dakar-Paris), coût de restauration(Dakar-Paris), frais d'inscription de ces cinq années ($300.000F + 4 \times 150.000F$) et plus d'autres frais divers.

Je remercie les responsables et personnels de l'IUSTA en particulier le DG MAHAMOUD YOUSOUF KHAYAL qui m'a poussé dans cette aventure et aussi pour son soutien tout au long de ce travail de thèse.

Je remercie les étudiants de thèse et stagiaires avec lesquels j'ai collaboré comme Abdou Karim Guèye, Mbaye Babacar Guèye, Doudou Gaye, Papa Baba Fall, Daouda Diouf, Abdou Rahmane Mayaki, Fily Diallo, Ablaye Thiam, Paul Batista pour leurs échanges fructueux.

Un grand merci à tous les Tchadiens rencontrés à Dakar : Dr Adoum Hassane, Youssouf Moussa, Abderamane Haliki, Dr Youssouf Mahamat, Dr Ahmat, Moustapha Abatcha, Adoum Yacoup, Mahamat Cherif, Mahamat Adam, Papa camara et sa famille. etc

Un grand merci à Monsieur le Pr Salif GAYE, Vice-Recteur de l'Université de Thiés, mon parent et collègue pour ses conseils et soutiens moraux et pour m'avoir permis d'entrer en contact avec mon directeur de thèse.

Je remercie mes amis et collègues du département d'Informatique Industrielle et de Gestion (IIG), qui n'ont ménagé aucun effort pour m'aider et participer à la réalisation de cette œuvre en particulier à Dr Kocsis le chef de département, Yaya Dagal, Mahamat habib, Dr Mahamat Abdoulaye Bechir, Dr Oumar Bada.

Je remercie mes amis du quartier Ridina en particulier le commissaire de la police judiciaire Idriss Adoum Daye pour son encouragement et soutien depuis mon entrée en 1ère année de MPC à l'université de Farcha où il m'avait acheté un vélo pour faciliter mes déplacements, je ne l'oublierai jamais. Un grand merci aussi à Djibrime(Bébé) pour son service rendu et soutien.

Je n'oublie pas de remercier Moctar BARRY Guinée et sa famille qui m'ont considéré plus que comme un parent.

Je remercie toutes les personnes que j'ai oubliées.

Je remercie toute ma famille ma maman Haoua Oumar et mon papa Ousmane Mahamat, tous défunts. Je remercie également la famille de Oumar Abbasse (Grand-père) avec qui j'ai grandi et devenu ce que je suis. En particulier Oubbo avec ses enfants, oncles avec ses enfants, amina avec ses enfants, zenabou avec ses enfants et toutes les autres familles.

Grand merci à Ali Baba avec son épouse Safia, Grand merci à Seid Baba et sa famille et Grand merci à Abdelsalam pour les services rendus pendant mon absence. Grand merci aussi à toutes mes nièces en particulier à Haoua Alioum.

Grand merci à ma femme Haoua Mahamat Saleh pour sa patience, aussi à mes deux filles Haoua et Didio.

Enfin, je remercie toute ma famille et belle-famille Hinda qui m'ont beaucoup aidé et soutenu durant ces cinq années.

Résumé : L'objectif de cette thèse est la classification neuronale de données satellitaires pour la caractérisation de l'upwelling Sénégal-mauritanien. Dans cette thèse, les données satellitaires (SST et Chl-a) ont permis de mener une étude de la variabilité et de la dynamique de l'upwelling côtier et de ses courants associés. Cette étude a la particularité de couvrir quasiment toute la côte d'Ouest-Africaine. L'approche utilisée a permis de décomposer les eaux de la zone Sénégal-Mauritanien en 6 classes à travers SOM2 puis SOM3. Les six classes obtenues sont bien définies et correspondent chacune à une zone géographique bien délimitée montrant la robustesse de cette approche. Les six classes peuvent être considérées comme des indicateurs statistiques pour analyser la variabilité spatio-temporelle de l'upwelling Sénégal-Mauritanien. L'analyse de ces classes ont permis de mettre en évidence que cette zone présente une très forte variabilité saisonnière qui est due à la fois au vent, à l'ensoleillement, et aussi à la circulation océanique de grande échelle (NECC). Nous avons par ailleurs introduit des données nLw (normalized water leaving radiance) afin de compléter les analyses des eaux atypiques (classe 5). L'utilisation des réseaux multicouches associés aux développements des algorithmes ont permis d'aboutir à des modèles spécifiques de normalisation des données de la zone d'étude, et leurs utilisations ont permis d'atténuer les effets de la chlorophylle-a afin de caractériser les eaux atypiques. Cette approche combinée de modèle de normalisation et les réseaux de neurones constituent l'apport de cette thèse, et a permis de caractériser non seulement les eaux des classes mais également quelques espèces biogéochimiques de la zone Sénégal-Mauritanien.

Mots-clés : upwelling Sénégal-Mauritanien, couleur de l'océan, température de surface, classification neuronale (SOM), variabilité spatio-temporelle.

Abstract : The aim of this thesis was the neural classification of satellite data for the characterization of the Senegalese-Mauritanian upwelling. In this thesis, satellite data have led to a study of the variability and dynamics of coastal upwelling and its associated currents. This study is unique in that concern almost the entire coast of West Africa. The approach allowed to decompose water Senegalese-Mauritanian zone in 6 classes through SOM2 then SOM3. The six classes obtained are well defined and each correspond to a well-defined geographical area showing the robustness of this approach. The analysis of these classes have to highlight that this area has a very strong seasonal variability is due to both wind, sunshine, and also to the large-scale ocean circulation (NECC). We have also introduced data nLw(normalized water leaving radiance) to complete the water analysis atypical (class 5). The developments of algorithms have resulted in specific Models normalization data of the study area, and their uses have helped reduce the effects of chlorophyll-a to the caracateriser water atypical. This combined approach of normalization model and neural networks are the contribution of this thesis, and helped to characterize not only the water but also classes of some species biogeochemical Senegalese-Mauritanian zone.

Keywords : senegalo-Mauritanian upwelling, ocean color, sea surface temperature, neural classification (SOM), spatio-temporal variability

Table des matières

Introduction Générale	9
1 Systèmes d' Upwelling de Bord Est	12
1.1 Introduction	12
1.2 Topographie et Bathymétrie de la zone d'étude	13
1.3 Dynamique de l'upwelling côtier	16
1.3.1 Transport et pompage d'Ekman	16
1.3.2 Aspect physique	16
1.3.3 Pompage	18
1.3.4 Application aux bords Est des bassins océaniques	18
1.3.5 Aspect biologique	19
1.3.6 Variabilité et saisonnalité de l'upwelling	20
1.4 Conclusion	22
2 Présentation des réseaux de neurones	23
2.1 Introduction	23
2.2 Historique	23
2.3 Définition	24
2.4 Apprentissage	24
2.5 Généralisation	25
2.6 Les Perceptrons Multicouches	26
2.6.1 Neurone artificiel	27
2.7 Le Perceptron	28
2.7.1 Architecture et mise en équation	28
2.7.2 Fonctionnement	29
2.7.3 Réseau à couches	31
2.7.4 Réseau monocouche, présentation	31
2.7.5 Réseaux multicouches	31
2.7.6 Apprentissage	32
2.7.7 Différents types d'apprentissage	32
2.7.8 Apprentissage non-supervisé	32
2.7.9 Apprentissage supervisé	32
2.7.10 Apprentissage supervisé du réseau monocouche	32
2.7.11 principe	32

2.7.12	Méthode de gradient totale	32
2.7.13	Accélération de la convergence	35
2.8	Cartes topologiques auto organisatrices de Kohonen	36
2.8.1	Principe de la modélisation	36
2.8.2	Architecture des cartes topologiques auto organisatrices de Kohonen	36
2.8.3	Apprentissage à l'aide de l'algorithme SOM	38
2.8.4	Algorithme de Kohonen	39
2.8.5	Classification ascendante hiérarchique	41
2.8.6	Qualité d'une typologie	41
2.8.7	Le critère de Ward	41
2.8.8	Algorithme CAH	41
2.9	Conclusion	42
3	Présentation de données	43
3.1	Introduction	43
3.2	Historique de la télédétection	43
3.2.1	Les capteurs	44
3.2.2	Quantités physiques mesurées en télédétection	44
3.2.3	Flux (radiant flux)	44
3.2.4	Intensité énergétique (radiant intensity)	45
3.2.5	Éclairement (irradiance)	45
3.2.6	Réflecteurs spéculaires et lambertiens (Specular and Lambertian reflectors)	45
3.2.7	Reflectance et albédo (réflectance and albedo)	45
3.2.8	Luminance (radiance)	45
3.2.9	Réflectance télédéteectée Rrs (remote sensing reflectance)	46
3.2.10	Luminance normalisée de l'eau $L_{WN}(\lambda)$ (normalised water-leaving radiance)	46
3.3	Couleur de l'eau et télédétection	46
3.3.1	Pourquoi s'intéresser au phytoplancton ?	46
3.3.2	Le rôle du phytoplancton dans le cycle du carbone	47
3.3.3	Régulation de la quantité de phytoplancton dans les océans	47
3.3.4	La couleur des océans	48
3.3.5	Les données de couleur de l'océan : SeaWiFS	49
3.3.6	Les données de température de la surface de la mer : Sea Surface Temperature SST	51
3.4	Conclusion	52
4	Analyse des situations océaniques à l'aide de cartes de Kohonen	53
4.1	Introduction	53
4.2	Upwelling Sénégal-Mauritanien	53
4.3	Bases de données utilisées	55
4.3.1	Les températures de surface SST	55
4.3.2	Calcul de la climatologie	55
4.3.3	Indice de température (ISST)	57

4.3.4	Les données de couleur de l'océan	57
4.3.5	Méthodologie	57
4.3.6	Détermination des classes de SOM2	59
4.4	Variabilité saisonnière de l'upwelling	63
4.5	Classification sur 3 variables : SOM3 (ISST, SST, Chl-a)	68
4.5.1	Variabilité interannuelle de l'upwelling	70
4.5.2	Discussions	72
4.6	Conclusion	73
5	Calcul d'une reflectance océanique normalisée (nLw) ne dépendant que de la chl-a à l'aide d'un MLP	74
5.1	Introduction	74
5.2	Données couleur de l'océan : nLw (normalized water leaving radiance)	74
5.2.1	Statistique classique	75
5.2.2	Développement du MLP	76
5.2.3	Prétraitements de données	76
5.2.4	Base d'apprentissage MLP	79
5.2.5	Apprentissage	79
5.2.6	Détermination de l'architecture optimale	80
5.2.7	Application du réseau MLP appris	80
5.3	Conclusion	82
6	Classification spectrale pour la caractérisation d'espèces phytoplanctoniques	84
6.1	Introduction	84
6.2	Les Données	85
6.3	la méthodologie	85
6.3.1	Résultats de l'apprentissage	87
6.3.2	Détermination des classes de $nLw^*(\lambda)$	87
6.3.3	Analyse spectrale	88
6.3.4	Discussions et Conclusions	92
	Conclusion Générale	94
	Annexe A Articles JSPI	98
	Annexe B Articles RJEES	111
	Annexe C Conférence Internationale	136
	Annexe D Présentation orale sous forme de poster (EGU)	138
	Annexe E Spectres climatologiques	140
	Annexe F Statistique SOM2(Chl-a,ISST)	150

Annexe Liste des Figures	157
Annexe Liste des Tableaux	160

Introduction Générale

L'océan côtier est en général caractérisé par une forte productivité biologique en raison d'un enrichissement en sels nutritifs lié aux activités anthropiques ou bien aux apports naturels des fleuves. Ainsi les zones côtières pourraient contribuer pour une part importante à la séquestration de carbone dans les couches profondes de l'océan. Les eaux côtières ont également un rôle important du point de vue économique. La majorité des pêches au niveau mondial se fait sur le plateau continental, les mers adjacentes et les conditions physiques jouent souvent un rôle majeur dans leur productivité. L'origine de cette productivité est due essentiellement aux alizés qui soufflent parallèlement à la côte, ce qui génère un upwelling côtier, ce dernier apportant une forte contribution en nutriments, enrichissant alors l'écosystème de surface. Les écosystèmes d'upwelling fournissent plus de 40% des captures des pêcheries mondiales alors qu'ils représentent moins de 3% de la surface de l'océan.

L'upwelling Sénégal-Mauritanien fait partie de l'écosystème issu du courant de canaris. L'upwelling Sénégal-Mauritanien est bien documenté ces 20 dernières années en utilisant d'abord les campagnes en mer puis à partir des observations satellitaires (Barton, 1998), (Demarcq and Faure, 2000), (Lathuiliere et al., 2008), (Mittelstaedt, 1991) et enfin dernièrement en s'appuyant sur les modèles numériques (Lathuiliere et al., 2008). Dans cette région, des vents alizés persistants entraînent la formation d'un intense upwelling côtier, c'est-à-dire la remontée/résurgence d'eaux froides et riches en nutriments, qui fournissent des conditions adéquates pour le développement d'une production primaire élevée et d'abondants écosystèmes marins.

Cette productivité biologique est vitale pour la communauté des pêcheurs et elle connaît une variabilité saisonnière. La zone est soumise comme partout ailleurs aux effets de changement climatiques et surtout à la surexploitation de flottes étrangères qui pillent, littéralement la zone au détriment de la population locale dont le rendement de la pêche s'affaiblit considérablement et progressivement au cours de ces dernières années. Les phénomènes de remontées d'eaux peuvent être localisés par leurs résultats : une mer froide (Température de la surface de la mer faible : SST) et riche en phytoplancton (mis en évidence par la concentration en chlorophylle-a de surface). Ces données peuvent être fournies par des satellites d'observation de la terre. Les mesures de ces données sont effectuées sur plusieurs bandes spectrales de la lumière, cette information spectrale peut être utilisée pour distinguer par ailleurs les différentes espèces de phytoplancton. Ces espèces présentent une

signature spectrale différente en raison des propriétés d'absorption des pigments et de leurs propriétés de diffusion qui sont spécifiques à chaque espèce. Récemment, (Alvain et al., 2005) ont pu caractériser les groupes de phytoplancton dans les eaux du cas-1 à partir de données satellitaires (capteur SeaWiFS) et ils ont pu identifier quatre groupes liés à d'espèces dominantes de phytoplancton.

Depuis plusieurs années, ces satellites fournissent à la communauté scientifique des données de la télédétection de plus en plus nombreuses et de plus en plus riches. La télédétection permet par ailleurs d'observer la surface de la mer sur de très grandes étendues et de façon quasi instantanée.

Les utilisations scientifiques des données d'observation de la terre depuis l'espace sont en plein essor, comme le montre le très vaste champ d'applications que ces techniques ont ouvert ou ont contribué à développer, en permettant l'accès aux mesures de nombreux paramètres géophysiques (vent, température de la surface de la mer, hauteur dynamique, houles, marées, paramètres atmosphériques, bilans radiatifs, concentration en chlorophylle de l'océan). Ces données servent essentiellement à la compréhension de divers phénomènes (physiques, biogéochimiques, ...) des océans et de l'atmosphère. La plupart des phénomènes abordés (upwelling, ...) dans cette thèse font intervenir de grands volumes de données mesurées par les capteurs embarqués sur des satellites mesurant plusieurs variables (le temps, l'espace). Les progrès informatiques permettent de nos jours le traitement des données volumineuses. Ces progrès rendent possible aussi la représentation graphique et dynamique des données d'origines diverses ainsi que leur analyse spatialisée. Ainsi on pourra représenter et étudier de façon spatialisée les liens entre les différentes composantes de l'écosystème : facteurs physiques et biogéochimiques.

Durant ce travail de thèse, nous avons fait recours à des méthodes d'analyse statistique multivariée pour réduire le volume de données d'une part et caractériser les résultats obtenus d'autre part. Les deux types de techniques, utiles à la caractérisation de tels volumes de données sont les méthodes de décomposition (ou de projection) et les méthodes de classification neuronale. Les méthodes de décomposition (ou projection) sont particulièrement bien adaptées à l'exploration de grands tableaux numériques : leur avantage est de fournir un nouveau jeu de données plus compact. Mais elles ne suffisent pas toujours à fournir une vue satisfaisante de l'ensemble des données. Non seulement les visualisations ne véhiculent qu'une partie de l'information, mais elles sont parfois elles-mêmes trop complexes pour être interprétées facilement.

Dans ces circonstances, les techniques de classification neuronale peuvent compléter et nuancer les résultats de la décomposition. La complémentarité entre la décomposition et la classification neuronale concerne la compréhension de la structure des données et celle des aides pratiques dans la phase d'interprétation des résultats.

Cette thèse a pour objectif principal de répondre à la question suivante : la classification neuronale de données satellitaires pour la caractérisation de l'upwelling Sénégal-Mauritanien peut-elle apporter une meilleure compréhension des processus physiques et biologiques liés au phénomène d'upwelling et de consolider les connaissances d'ordre générales, climatiques et biogéochimiques ? Quelles sont les courants de surfaces associés à l'upwelling ? Quelles sont les variabilités saisonnières et inter-annuelles associées aux phénomènes de l'upwelling ?

Pour répondre à ces questions, cette thèse propose une méthode originale permettant en particulier une approche pertinente basée sur la classification neuronale des paramètres géophysiques afin de caractériser les zones d'upwelling. Cette approche neuronale permet de résumer l'ensemble de données en un ensemble réduit représentatifs (référents) et l'interprétation de ces référents obtenus doit pouvoir faire apparaître des zones distinctes significatives des différents états de l'océan dans la zone Sénégal-Mauritanienne. Nous utiliserons donc les réseaux de neurones artificiels en particulier : les cartes auto organisatrices et les perceptrons multicouches. Les réseaux de neurones ont déjà montré leur utilité dans de nombreux problèmes de télédétection appliquée à l'océanographie.

Le plan de ce travail est le suivant : nous débutons naturellement par une introduction générale suivie de la description dans le chapitre 2 de la physique des phénomènes de l'upwelling Sénégal-Mauritanien et ses processus physiques. Dans le chapitre 3 la méthode des réseaux de neurones est présentée en détail. Le chapitre 4 sera consacré à la présentation des données. Les chapitres 5, 6 et 7 seront dédiés à l'application des réseaux de neurones et enfin nous clôturerons ce travail de thèse par la conclusion générale.

Chapitre 1

Systemes d'Upwelling de Bord Est

1.1 Introduction

Le Systeme de Courant de Canaries (Canary Current System, CCS) est l'un de Systeme d'Upwelling de Bord Est (Eastern Boundary Upwelling System, EBUS) (Lachkar and Gruber, 2012) de l'océan Atlantique Nord. Les EBUS sont situés dans les deux hémisphères le long des côtes Ouest des continents (bord Est des Océans) bordant les deux principaux bassins océaniques, à savoir l'océan Pacifique et l'océan Atlantique. Ils sont visibles depuis l'espace car la production primaire élevée qui caractérise ces systemes permet d'observer leurs variabilités physiques et biogéochimiques avec la chlorophylle-a, comme l'illustre la Figure 1.1.

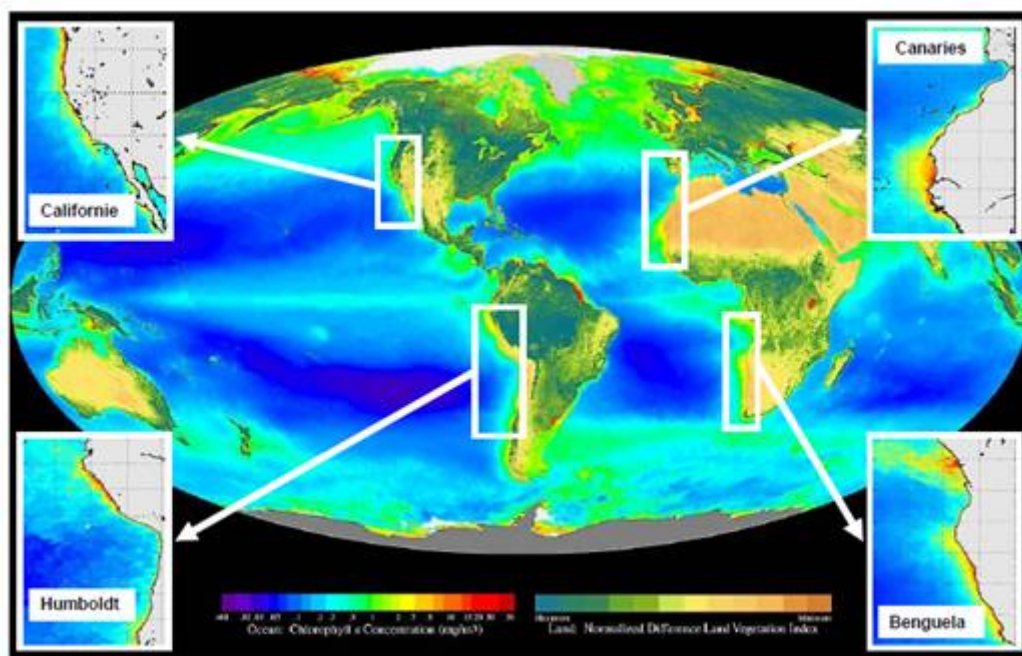


Figure 1.1 – Concentration moyenne en chlorophylle-a en surface ($mg.m^{-3}$) sur la période 09/1997-08/2000 à partir des données satellite SeaWiFs (NASA Goddard Space Center Flight). Les quatre principaux EBUS sont indiqués : le Courant de Californie, le Courant de Humboldt, le Courant des Canaries et le Courant de Benguela. Les cartes regionales sont mises à la disposition par l'IRD/AOOS

Il est généralement admis qu'il existe quatre EBUS principaux :

- **Le Système de Courant des Canaries**, présent le long des côtes Sénégal-mauritanien et jusqu'au nord du Maroc, avec une extension septentrionale le long du Portugal et du nord-ouest de l'Espagne appelé système d'upwelling ibérique occidental ;
- **Le Système de Courant du Benguela**, s'étendant de la pointe sud-ouest de l'Afrique du Sud jusqu'en Angola ;
- **Le Système de Courant de Californie**, s'écoulant le long de la côte ouest de la Basse-Californie (Mexique) et de la côte ouest des États-Unis (Californie, Oregon, Washington) ;
- **Le Système de Courant de Humboldt**, qui caractérise presque toute la côte ouest d'Amérique du Sud, du Chili central vers 40° Sud au nord du Pérou près de l'équateur.

Les emplacements géographiques de ces systèmes d'upwelling sont déterminés par la position des anticyclones des moyennes latitudes, situés dans les bassins Nord et Sud du Pacifique et de l'Atlantique. En effet, ces systèmes de hautes pressions ont des branches orientales caractérisées par des alizés à basse altitude dirigés vers l'équateur. Près de la côte, ces vents ont souvent tendance à avoir une direction parallèle à la côte, induisant un transport d'Ekman des eaux de surface vers le large (Ekman, 1905) qui sont alors remplacées par des eaux plus profondes remontant à la surface, eaux qui sont plus froides et riches en éléments nutritifs ce qui favorisent les blooms phytoplanctoniques. Le paragraphe qui suit décrit particulièrement les caractéristiques du Système de Courant de Canaries (CCS) en terme de : topographie, bathymétrie, climat et hydrologie.

1.2 Topographie et Bathymétrie de la zone d'étude

La zone sur la quelle nous avons focalisé notre étude fait partie du système du CCS. Elle est comprise, en latitude, entre 9°N et 21°N (soit du cap Blanc en Mauritanie au Sud de la Guinée, Figure 1.2). La côte Mauritanienne est un vaste ensemble qui s'étend entre 16°04' et 20°36' de latitude Nord sur près de 720km (Sanyo, 2002). Elle est rocheuse au nord du cap Blanc, sableuse et rectiligne au sud du cap Timilis. Le plateau continental au sud du cap Blanc est large de 40 à 60 miles (1 mile =1.6 km) puis se rétrécit jusqu'au Cap Timiris où l'isobathe 200m ne se trouve plus qu'à 10 milles de la côte. Depuis le sud du cap Timiris jusqu'à la latitude du Saint-Louis, la largeur est en moyenne de 30milles. Les fonds sont généralement réguliers et la pente vers le large est douce pour la plus grande partie du plateau. Un décrochement est cependant perceptible vers 50m où on trouve ça et là des affleurements rocheux longitudinaux ; en face du banc d'Arguin ainsi que devant le cap Timiris, des roches affleurent entre 25m et 50m de profondeur (Maigret, 1972). De nombreuses fosses entaillent profondément le plateau continental et arrivent au contact du banc où les fonds de 10m voisinent alors avec ceux de 300m à 400m. Au Sénégal, la Presqu'île du Cap-Vert sépare le domaine maritime sénégalais en deux zones aux caractéristiques topographiques bien distinctes (Roy, 1992). Au Nord de la presqu'île, le plateau continental est étroit et orienté *N – NE*. Au sud, le plateau continental s'élargit, le talus est situé entre 10 et 30 milles des côtes et est orienté *N* puis *NW – N*. Au niveau de la presqu'île du Vap-Vert, point le plus occidental de l'Afrique, la largeur du plateau est réduite à quelques miles. En effet, situé à 27 miles au large de Saint-Louis, l' isobathe 200m se rapproche doucement de la côte vers le sud. A la hauteur de la pointe des Almadies, la largeur du plateau n'est plus que de 5 miles. La fosse de kayar au nord de Dakar

entaille le plateau et le traverse sur toute sa largeur. Le plateau est étroit à partir de cet endroit jusqu'au Cap-Vert. L' isobathe 100m n'est qu'à 5km ou 6km de la côte. L' isobathe 50m est à quelques centaines de mètres de la plage à kayar ($14^{\circ}55'N, 17^{\circ}7'O$) et reste ensuite à 1km à 3km de la côte jusqu'à la pointe des Almadies. Sur la côte Sud existe deux falaises sous-marines dont l'une a un sommet qui se situe à une profondeur de 35m à 45m et l'autre moins longue a son bord supérieur vers 70m de profondeur avec une dénivellation d'une dizaine de mètres.



Figure 1.2 – Situation géographique de la zone d'étude : bathymétrie (en isoligne, en m), plateau continental (en vert)

Le littoral guinéen est un immense complexe deltaïque constitué par un ensemble de plaines côtières pouvant atteindre 30km de large et limitées vers l'intérieur par des falaises de plateaux gréseux qui constituent les premiers contreforts du Fouta Djallon. Décrit pour la première fois par (Postel, 1955) le plateau continental de la Guinée, de par son étendue est considéré comme la plus grande surface continentale submergée de l'Afrique Atlantique. La distance du littoral à l'isobathe 200m, qui marque la limite de ce plateau continental, augmente régulièrement du sud vers le nord : elle est d'environ 87 milles dans la partie sud et de 104 milles dans la partie nord. Cette irrégularité de la topographie et celle de la bathymétrie sont d'une importance particulière pour la circulation à la côte.

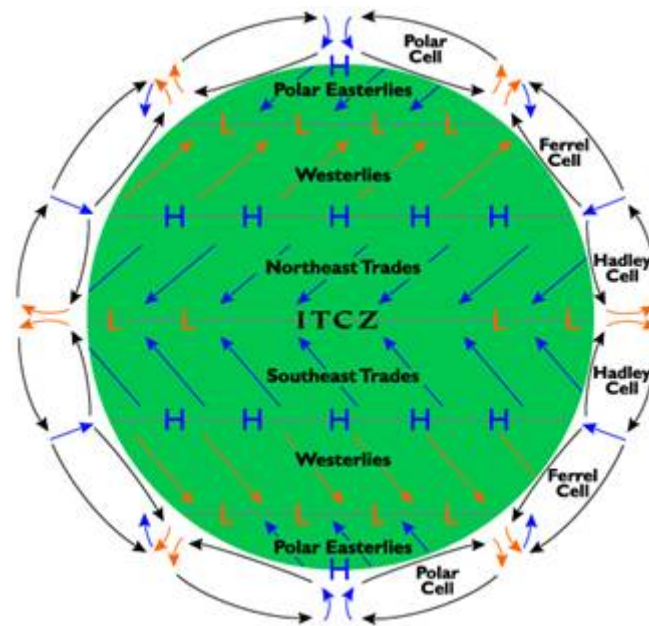


Figure 1.3 – Circulation atmosphérique méridienne montrant les cellules de Hadley et les alizés (Trades)

Cette circulation atmosphérique est modulée par le déplacement saisonnier l'ITCZ. En effet l'ITCZ n'est pas symétrique par rapport à l'équateur dû probablement à la différence de répartition des continents dans l'hémisphère Nord et Sud. Deux points sont importants pour comprendre la relation entre les systèmes de vents et l'ITCZ. Tout d'abord, il s'agit d'une zone de convergence, c'est-à-dire que, à proximité de cette ITCZ, les vents vont du nord vers le sud dans la partie nord, et inversement dans la partie sud. Le deuxième point est que cette ITCZ n'est pas en position fixe, mais suit un rythme de déplacement saisonnier. L'ITCZ est en position nord durant l'été boréal (Figure 1.4a) générant la mousson Africaine et en position sud durant l'été austral (Figure 1.4b). Le balancement des saisons d'un hémisphère à l'autre s'accompagne de la migration de l'ITCZ et donc du changement des régimes de vents. Cette variabilité saisonnière liée au déplacement du maximum d'ensoleillement. Au niveau de la côte Mauritanienne, la position la plus au nord de l'ITCZ monte jusqu'aux environs du Cap Blanc (Figure 1.5a et 1.5c-d), vers $21^{\circ}N$, qui est la limite nord de la zone qui nous intéresse. Or sous l'action de la force de Coriolis, les mouvements de fluides dans l'hémisphère nord sont déviés sur leur droite (et sur leur gauche dans l'hémisphère sud). Au nord de l'ITCZ, le mouvement des vents convergents, d'orientation nord - sud, est alors dévié vers l'ouest, donnant ainsi naissance aux alizés (Figure 1.4). Dans cette zone d'étude, les vents sont donc principalement orientés depuis le nord-est vers le sud-ouest quelle que soit la saison. D'un point de vue hydrologique, une première conséquence est l'existence du courant (appelé improprement courant de Canaries) des Canaries (c'est un courant prolongement du courant du Portugal), longeant les côtes du Maroc et du Sahara Occidental depuis le nord vers le sud. Il faut noter que l'intensité des vents et leur orientation vont être fonction de leur position par rapport à l'ITCZ (en l'absence d'éventuelle influence orographique). Ainsi, lorsque l'ITCZ va migrer vers le sud, un point quelconque de la côte va subir la migration des conditions climatiques. Toujours en considérant cette position nord de l'ITCZ pendant l'été boréal, les vents de la partie sud ont un comportement différent (Figure 1.4a et 1.5c). En effet, étant donné qu'ils convergent vers l'ITCZ, les vents sont orientés du sud vers le nord alors qu'ils sont dans l'hémisphère nord. Ils sont donc déviés

vers l'est et forment le phénomène de mousson Africaine du sud-ouest apportant des vents chargés d'humidité sur le continent africain. Au sud de l'ITCZ, le courant côtier est alors dirigé du sud vers le nord et rencontre le courant des Canaries (de sens opposé) aux environs du Cap Blanc (Figure 1.5c et 1.5). . Au moment où l'ITCZ est en position sud, c'est-à-dire en été austral (Janvier, Février), la

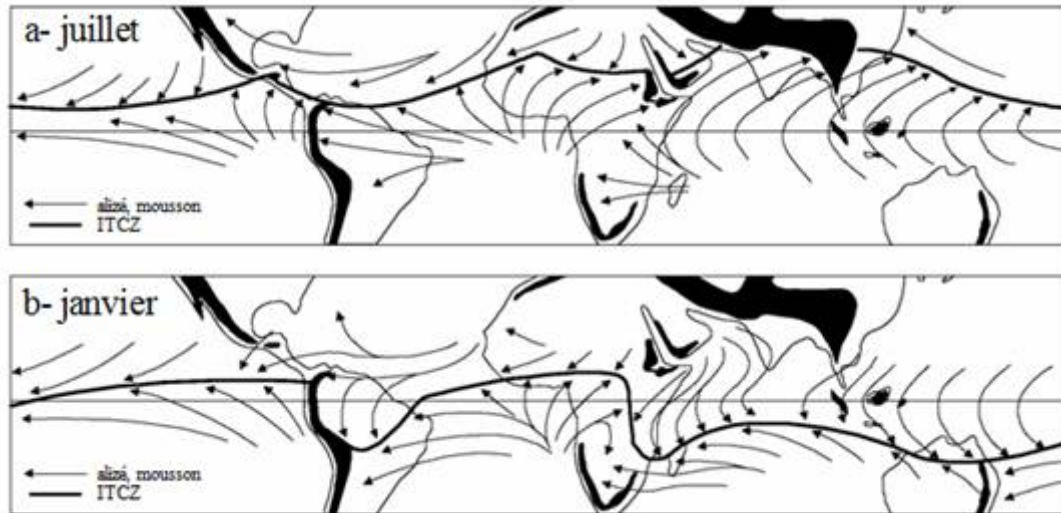


Figure 1.4 – Circulation tropicale du vent en surface (a) en juillet (b) en janvier. (d'après Leroux, 1996)

mousson est beaucoup moins intense car l'incursion des vents sud dans l'hémisphère nord est moins profonde (Figure 1.4b et 1.5a-b) Sur la côte, les vents de secteur nord poussent le courant des Canaries vers le sud au delà du Cap Blanc (Figure 1.5). Enfin, notons qu'étant sur le bord Est du bassin Nord Atlantique, il existe aussi une composante géostrophique au courant des Canaries, issue de la grande gyre Nord – Atlantique (Giraud, 2001). Voyons donc maintenant la relation qui existe entre le régime de vents et le phénomène d'upwelling côtier.

1.3 Dynamique de l'upwelling côtier

1.3.1 Transport et pompage d'Ekman

Les mécanismes physiques qui permettent d'expliquer la présence d'upwelling côtier dans les régions de Bord Est sont connus depuis longtemps (Marie, 2007).

1.3.2 Aspect physique

Selon la théorie d'Ekman (Ekman, 1905), des vents soufflant à basse altitude au-dessus de l'océan forcent une circulation océanique, qui est déviée à droite dans l'hémisphère nord et à gauche dans l'hémisphère sud en raison de la force de Coriolis. A la surface, le mouvement résultant est dirigé à 45° par rapport à la direction du vent dans le sens des aiguilles d'une montre (resp. dans le sens inverse des aiguilles d'une montre) dans l'hémisphère nord (resp. sud). Cette théorie a été élaborée à partir d'observations effectuées par le physicien norvégien Nansen. L'océanographe, lors d'une campagne en

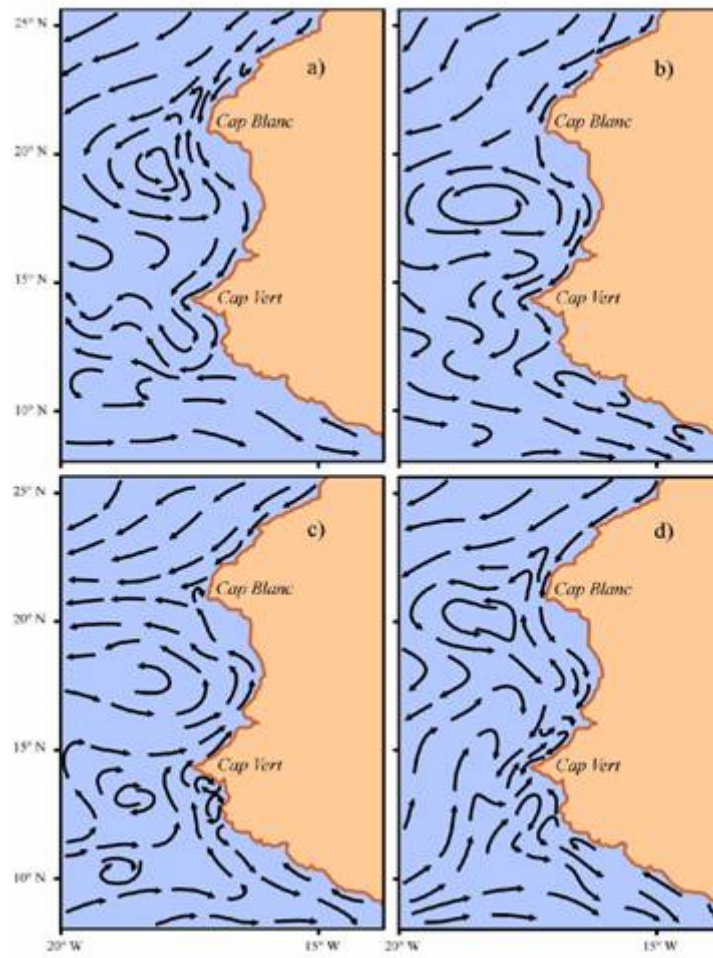


Figure 1.5 – Schéma de circulation de surface pour la côte Nord-Ouest Africaine (a) en hiver boréal, (b) au printemps, (c) en été et (d) en automne (Mittelstaedt, 1991)

Arctique, remarque que les icebergs dérivent avec un angle variant de 20 à 40° par rapport à la direction du vent. Ekman, son étudiant propose alors une explication à ces observations reposant sur l'équilibre entre tension turbulente et force de Coriolis. On a au sein du fluide :

$$\begin{cases} fv + A_z \frac{\partial^2 u}{\partial z^2} = 0 \\ fu - A_z \frac{\partial^2 v}{\partial z^2} = 0 \end{cases} \quad (1.1)$$

où f est le paramètre de Coriolis, u et v sont les vitesses zonale et méridienne, A_z est le coefficient de viscosité turbulente verticale. Ekman suppose que la tension verticale est due à l'influence du vent. Celui-ci étant stationnaire, la surface de l'océan plane et la densité de l'eau ρ constante, on a en surface :

$$\begin{cases} \tau_x = \rho A_z \frac{\partial u}{\partial z} \\ \tau_y = \rho A_z \frac{\partial v}{\partial z} \end{cases} \quad (1.2)$$

l'équilibre entre force de Coriolis et tension vent se traduit par les équations suivantes en surface :

$$\begin{cases} f\rho v + \frac{\partial \tau_x}{\partial z} = 0 \\ -f\rho u + \frac{\partial \tau_y}{\partial z} = 0 \end{cases} \quad (1.3)$$

où τ_x et τ_y sont les tensions de vent zonal et méridien. Ekman intègre ensuite ces équations de la surface (où la tension turbulente est connue et est égale à la tension de vent) jusqu'à la profondeur $-H$ où la tension turbulente est nulle. Il obtient ainsi les transports zonaux T_x et méridionaux T_y par unité de temps et de distance (en m^2/s) :

$$\begin{cases} T_x = \frac{\tau_0^y}{f\rho} \\ T_y = -\frac{\tau_0^x}{f\rho} \end{cases} \quad (1.4)$$

où T_x et T_y sont les tensions de vent zonal et méridien en surface. Le transport net est donc dirigé à ± 90 degrés de la direction du vent suivant l'hémisphère. Si on se place à une latitude moyenne ($f = 10^{-4}s^{-1}$), pour une valeur typique de la tension de $0.1N/m^2$, correspondant à un vent de l'ordre de $8ms^{-1}$ le transport d'Ekman est alors de l'ordre de $1m^2/s$, soit $0.1Sv/deg$ de latitude (pour un vent méridien).

1.3.3 Pompage

Les variations spatiales du vent et du transport d'Ekman induisent des zones de convergence et de divergence des eaux de surface. La vitesse verticale à la base de la couche d'Ekman de profondeur $-H$ (W_{EK} , appelée pompage d'Ekman) résultant de ces convergences et divergences est obtenue à partir de l'équation de continuité appliquée au transport d'Ekman en supposant le paramètre de Coriolis constant :

$$\begin{cases} W_{EK} = \int_{-H}^0 (\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y}) dz = \\ \frac{1}{\rho f} [\frac{\partial}{\partial x} (\int_{-H}^0 \frac{\partial \tau_y}{\partial y} dz) - \frac{\partial}{\partial y} (\int_{-H}^0 \frac{\partial \tau_x}{\partial z} dz)] = \\ \frac{1}{\rho f} (\frac{\partial \tau_0^y}{\partial x} - \frac{\partial \tau_0^x}{\partial y}) \end{cases} \quad (1.5)$$

Si la vitesse est positive (resp. négative), on a un upwelling (resp. un downwelling). Cette expression peut être améliorée en tenant compte de la variation longitudinale du paramètre de Coriolis. Proche de l'équateur, on peut alors écrire : $f = \beta y$ où y est la distance à l'équateur. Le pompage a alors pour expression (Halpern, 2002) :

$$W_{EK} = \frac{\vec{\nabla} \wedge \vec{\tau}_0}{\rho f} + \frac{\beta \tau_x}{\rho f^2} \quad (1.6)$$

1.3.4 Application aux bords Est des bassins océaniques

La théorie d'Ekman appliquée au bord Est des bassins océaniques donne une relation entre la tension de vent parallèle à la côte et un transport d'eau vers le large. En effet, la composante parallèle à la côte et dirigée vers l'équateur de la tension de vent produit un transport vers l'Ouest perpendiculaire à la côte et donc un transport vers le large :

$$T_{EK} = \frac{\tau_{alongshore}}{\rho f} \quad (1.7)$$

$\tau_{alongshore}$ est la composante parallèle à la côte de la tension du vent. Par ailleurs, le long des bords Est des bassins océaniques, l'intensité du vent diminue quand on s'approche de la côte, définissant ainsi une région où le rotationnel de la tension du vent est cyclonique (Bakun et al., 1991). Ce rotationnel de vent est à l'origine d'un pompage positif (ci-après). La zone côtière est alors une zone de divergence

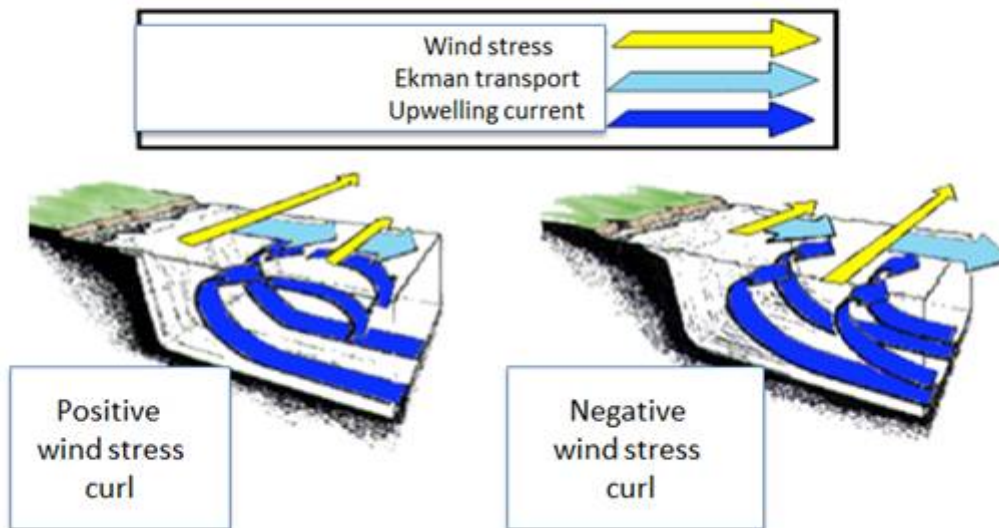


Figure 1.6 – Schéma de l'upwelling et de sa dépendance du rotationnel du vent au voisinage de la côte.

des eaux de surface. L'eau en profondeur, plus froide, remonte en surface afin de remplacer les eaux de surface divergentes : c'est l'upwelling côtier. Cette remontée d'eau s'accompagne d'une remontée des isothermes côtières et de la thermocline. En réalité, le vent ne souffle pas de manière continue mais se manifeste par des épisodes de forts vents (coups de vent) suivis par des épisodes de relaxation. On sait maintenant que ce régime de vent a une influence directe sur les upwellings côtiers. Eux aussi se caractérisent par des périodes d'intenses résurgences suivies de périodes de relaxation formant ainsi des cycles d'upwelling. Durant un cycle (durée de 3 à 7 jours), les vents parallèles à la côte s'intensifient et emportent les eaux froides vers le large sous forme de « plumes » ou filaments puis ces eaux reviennent à la côte durant les périodes de relaxation. Pendant les périodes de fort vent, le front d'upwelling semble se situer parallèlement à la côte, advecté par un transport d'Ekman uniforme. Quand le vent s'affaiblit, le front devient plus compliqué et semble s'intensifier. Ces lieux où se concentrent l'activité des upwellings sont appelés « centres d'upwelling ».

1.3.5 Aspect biologique

Le phénomène de remontée des eaux profondes et froides à la côte s'accompagne de l'apport des nutriments dans la zone euphotique. Cette disponibilité en nutriments initie le développement du réseau trophique, à commencer par un fort développement phytoplanctonique. Ceci est clairement visible sur des photos satellites de la couleur de la mer, donnant des informations sur la concentration en chlorophylle (Figure 1.7). Ces concentrations peuvent alors atteindre des valeurs jusqu'à 3.5

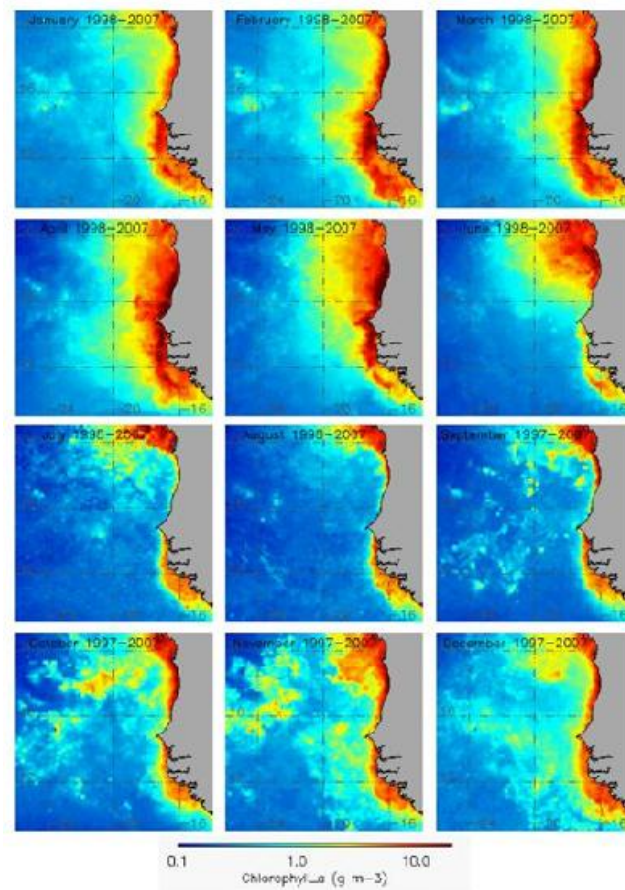


Figure 1.7 – Images satellite SEAWIFS donnant la concentration en chlorophylle des eaux de surface. Données mensuelles (climatologiques) pour 1997 à 2007

mgChl-a.m^{-3} [(Gabric et al., 1993); (Morel, 1988); (Morel and Maritorena, 2001)]. Nous sommes alors en conditions eutrophes, où la production primaire est la plus élevée, jusqu'à $2\text{gC.m}^{-2}.\text{j}^{-1}$ [(Morel, 1988);(Morel and Maritorena, 2001)]. Les masses d'eaux froides et riches en nutriments et phytoplancton se présentent sous forme de longs filaments et méandres transportés vers le large (Barton, 1998). Lors de ce transport, le développement du phytoplancton, ainsi que celui du zooplancton et des niveaux trophiques supérieurs, conduit à l'exportation de matière organique vers le fond de la colonne d'eau sous forme de détrit (pelotes fécales, organismes morts, neige marine) [(Gabric et al., 1993);(Fischer et al., 1996)]. Ces détrit, lors de leur chute, vont subir une reminéralisation, ce qui a pour conséquence de régénérer le stock de nutriments. Pour une concentration initiale en nitrates des eaux de subsurface à environ 10mmolN.m^{-3} (Babin et al., 2003) , cette masse d'eau peut atteindre une concentration allant jusqu'à 15 à 18mmolN.m^{-3} au point de résurgence de l'upwelling (Jacques and Tréguer, 1986), ce qui représente un taux de recyclage de 50 à 80 %.

1.3.6 Variabilité et saisonnalité de l'upwelling

D'un point de vue physique, la saisonnalité de l'upwelling sénégal-mauritanien est liée par le déplacement de l'ITCZ et des régimes de vents qui l'accompagne. Ce changement de position du régime de vent par rapport à la côte, et donc par rapport à la topographie de la plate-forme, va jouer sur la position

et l'intensité des résurgences. Ceci peut se quantifier par l'index d'upwelling, qui est la différence de température des eaux de surface entre la côte et le large (Mittelstaedt, 1991). L'upwelling se caractérisant par la résurgence des eaux froides à la côte, d'une manière simpliste, plus l'index d'upwelling est négatif, plus l'upwelling est intense. Ainsi, cette saisonnalité est clairement visible sur la Figure 1.8 où est

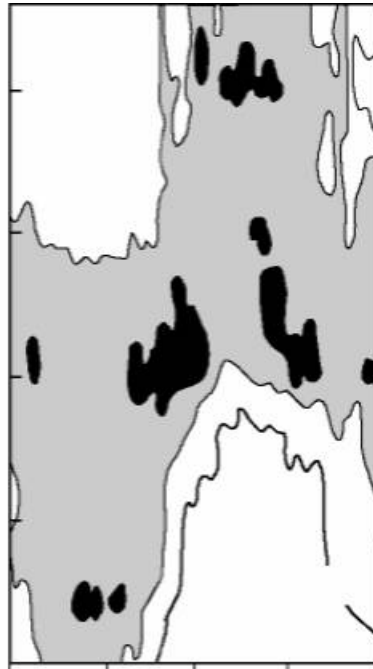


Figure 1.8 – Carte d'index d'upwelling, $10^{\circ}N$ et $35^{\circ}N$ pour la période 1969-1976 (Speth and Detlefsen, 1982)

présenté l'index d'upwelling moyen entre $10^{\circ}N$ et $35^{\circ}N$ sur la côte Nord-Ouest Africaine pour la période 1969-1976 (Speth and Detlefsen, 1982). On peut alors remarquer le déplacement saisonnier du maximum d'upwelling, dont les positions extrêmes correspondent aux positions de l'ITCZ en été et en hivers. Toutefois, si la migration de la position des eaux froides à la côte est nette, la zone du maximum d'index d'upwelling, en deçà de -6° ; n'est pas continue. La résurgence des eaux froides se fait donc de manière plus intense aux environs de $12^{\circ}N$ en mars, entre $19^{\circ}N$ et $25^{\circ}N$ pour le printemps et l'automne (mai - juillet et septembre - octobre) et vers $30^{\circ}N$ pour juillet - août. Il peut s'agir ici de l'interaction plus ou moins favorable entre les différents facteurs physiques nécessaires au développement de l'upwelling : l'orientation du vent de façon locale par rapport au trait de côte, l'intensité du vent et l'extension plus ou moins large de la plate forme continentale.. A ce mécanisme général de la résurgence côtière, sous l'action des régimes de vents moyens, et évoluant selon les saisons le long de la côte, il faut ajouter une autre variation de plus courte échelle. En effet, les alizés décrits précédemment et qui sont à l'origine du forçage de l'upwelling ne sont bien sûr pas constants et réguliers à l'échelle journalière. On observe alors une variabilité dans l'intensité et l'orientation du vent en chaque point de côte, formant ainsi des pulses. Ceci semble important pour la transmission d'énergie depuis l'atmosphère vers l'océan (Roy and Cury, 1988). Mais il s'agit alors de conditions non plus générales, mais spécifiques de chaque année.

1.4 Conclusion

Au cours de ce deuxième chapitre, nous avons présenté les Système d'Upwelling de Bord Est (Eastern Boundary Upwelling System, EBUS), en particulier le Système de Courant de Canaries (Canary Current System, CCS) de l'Océan Atlantique Nord. Le CCS est perceptible depuis l'espace puisque la production primaire élevée qui caractérise ce système permet d'observer les variabilités physiques et biogéochimiques avec la chlorophylle-a qui est un pigment de phytoplancton indicateur de l'état de productivité de l'océan.

Nous avons par ailleurs présenté les aspects topographiques, bathymétriques de la zone d'étude. Les aspects géographiques, climatiques ainsi que hydrologiques y étaient soulignés. Il a été également présenté les processus physiques (Ekman, 1905) qui sont à la base de la productivité des écosystèmes marins.

Le chapitre suivant va s'attaquer à la présentation des méthodes neuronales qui sont au cœur de ce travail de thèse et qui vont permettre de comprendre les processus physiques exposés dans cette partie.

Chapitre 2

Présentation des réseaux de neurones

2.1 Introduction

Depuis une vingtaine d'année, on observe une augmentation importante du nombre de satellites dédiés à l'observation de la Terre. Ces nombreux satellites fournissent d'importantes quantités de données permettant une amélioration de la connaissance de notre environnement qu'il s'agisse de l'atmosphère, du sol ou de l'océan et ainsi, une meilleur compréhension des mécanismes qui régissent le climat. Un problème qui se pose à l'heure actuelle est la définition et la mise en place de méthodes numériques performantes pour le traitement de ces masses de données. Ce problème est abordé par des méthodes statistiques afin de restituer les paramètres d'environnement à partir des observations de télédétection. Les réseaux de neurones entrent dans ce cadre. Les réseaux tels que nous les utiliserons sont à classer parmi les techniques de modélisation dans le sens très général du terme. Les techniques neuronales que nous allons présenter dans ce chapitre ne couvrent qu'un domaine restreint des applications potentielles des réseaux de neurones. Les grands domaines d'application des réseaux de neurones comprennent, la modélisation de processus dynamique non linéaire (par exemple : prédictions financières, prédiction de consommation ou encore détection d'anomalies de fonctionnement), la reconnaissance de formes (par exemple : lecture des codes postaux, lecture des montants des chèques), la commande de processus (pilotage « neuronal » automatique de véhicule), la classification. Un aperçu des réseaux de neurones et de leurs applications peut-être trouvé dans (Bishop, 1995).

2.2 Historique

Les premiers travaux sur les modèles neuronaux remontent aux années 1940 (McCulloch and Pitts, 1943). Dans la dernière décennie, l'intérêt pour les réseaux de neurones a été rehaussé par l'évolution de l'informatique d'une part et d'autre part par la découverte de nouvelles techniques d'apprentissage levant les limitations des premiers modèles. Nous verrons, en particulier, l'algorithme de rétropropagation du gradient pour les perceptions multicouches qui s'est construit en utilisant les méthodes de descente du gradient (Rumelhart et al., 1986).

2.3 Définition

Les réseaux de neurones artificiels sont des réseaux fortement connectés de processeurs élémentaires fonctionnant en parallèle. Chaque processeur élémentaire calcule une sortie unique sur la base des informations qu'il reçoit. Un neurone se caractérise par trois concepts : son état, ses connexions avec d'autres neurones et sa fonction d'activation. Le neurone formel est décrit en détail dans le paragraphe 2.6.1. Un réseau de neurones est constitué généralement de m entrées représentées par $x = \{x_1, \dots, x_n\}$ et p sorties notées $y = \{y_1, \dots, y_p\} = y(x, W)$.

2.4 Apprentissage

Dans cette partie, on décrit succinctement l'apprentissage de réseaux de neurones. Le processus d'apprentissage d'un réseau de neurones consiste à déterminer les valeurs de ses poids synaptiques afin qu'il soit capable de traiter un certain type de données. On se sert donc d'un ensemble de N données statiquement représentatif du problème pour spécialiser le réseau, c'est *l'ensemble d'apprentissage*. Lorsqu'on cherche à apprendre à un réseau de neurones le lien fonctionnel qui existe entre deux ensembles de données, l'apprentissage est supervisé (modélisation de processus ou de fonction de transfert, classification). L'ensemble contient deux sous-ensembles, N données d'entrées $\{x^n\}$, et N données désirés correspondantes $\{d^n\}$. En revanche, si on cherche à analyser les structures d'un seul ensemble de données cohérentes entre elles $\{x^n\}$, l'apprentissage est *non-supervisé* car on ne donne aucune indication *a priori* au réseau de neurones (classification automatique, reconnaissance de forme, quantification vectorielle, analyse en composante principales). Les apprentissages effectués au cours de ce travail sont basé sur un algorithme itératif dont la règle principale (appelée règle de Delta) est de la forme :

$$w_{ij}(t+1) = w_{ij}(t) + \Delta w_{ij}(t) \quad (2.1)$$

où $w_{ij}(t+1)$ est le poids qui relie le neurone j au neurone i à l'itération $t+1$. Cette valeur dépend de la valeur du poids à l'itération précédente $w_{ij}(t)$, et par conséquent ce type d'algorithme nécessite une initialisation des poids du réseau. Cette initialisation est naturellement liée au codage des données et à la nature du réseau utilisé. La modification des poids $\Delta w_{ij}(t)$ peut-être établie sur différents critères, pour les réseaux utilisés durant ce travail, elle est basé sur une *fonction de coût* R qui fourni un mécanisme de contrôle de l'apprentissage (par exemple, la fonction de coût des moindres carrés). La fonction de coût reflète l'erreur que fait le réseau par rapport au but à atteindre, son expression analytique est une fonction des poids W du réseau. L'apprentissage consiste à trouver la valeur des poids W qui *minimise* la fonction de coût. L'expression de $\Delta w_{ij}(t)$ dépendra de celle R . Lorsque l'apprentissage progresse, la fonction de coût diminue et de l'algorithme converge. On peut surveiller les progrès d'un réseau au cours de son apprentissage grâce à plusieurs critères de performances calculés sur les N données dans leurs unités physiques.

$$E^{RMS} = \sqrt{\frac{1}{N \times p} \sum_{i=1}^N \sum_{k=1}^p \{e_{ik}\}^2} \quad (2.2)$$

où p est le nombre de neurones pris en compte et où e_{ik} est un signal erreur (par exemple, pour l'apprentissage supervisé, la $i^{\text{ième}}$ donnée de l'ensemble d'apprentissage généré par l'erreur $e_{ik} = y_{ik} - d_{ik}$ à la $i^{\text{ième}}$ sortie du réseau.

1. L'erreur relative moyenne, utilisée en apprentissage supervisé, et qui permet de s'affranchir de l'ordre de grandeur des données où p est le nombre de sorties du réseau.

$$E^{Rel} = \frac{1}{N \times p} \sum_{i=1}^N \sum_{k=1}^p \sqrt{\left\{ \frac{y_{ik} - d_{ik}}{d_{ik}} \right\}^2} \quad (2.3)$$

2. Le coefficient de corrélation linéaire, utilisé en régression, qui donne une mesure de la qualité d'un estimateur de fonction :

$$r^2 = \frac{1}{N} \sum_{i=1}^N \frac{(y_i - \bar{y}) \times (d_i - \bar{d})}{\sigma_y \times \sigma_d} \quad (2.4)$$

σ_y et σ_d sont les écarts types des données simulées y et des données désirées d . Le coefficient r^2 n'atteint 100% que lorsque le réseau est capable de restituer exactement les données désirées, ce qui n'est pas toujours possible, notamment lorsque les données sont bruitées.

2.5 Généralisation

On appelle phase de généralisation, la phase d'exploitation d'un réseau de neurones. Ce terme vient du fait qu'un réseau en phase d'exploitation va traiter des données qui appartiennent à son domaine d'apprentissage mais qui n'ont pas fait nécessairement partie de son ensemble d'apprentissage. Si les performances précédentes se dégradent en phase d'exploitation, il est possible qu'un sur-apprentissage ait eu lieu. Ce sur-apprentissage peut par exemple apparaître quand on travaille avec des données en nombre insuffisant par rapport à la complexité potentielle du réseau. Un réseau peut avoir suffisamment de souplesse pour *aller trop loin* et apprendre par cœur les données d'apprentissage. Il sera alors très mauvais en phase de généralisation car il n'aura pas appris la relation sous-jacente aux données (Figure 2.1). Le sur-apprentissage peut-être évité en limitant le degré de liberté du réseau par rapport à la taille de l'ensemble d'apprentissage (toujours avoir 10 à 20 fois plus de données que de poids à ajuster) et en arrêtant l'apprentissage à la fin de la convergence (*early stopping*). La complexité du réseau, ainsi que l'arrêt de l'apprentissage se déterminent grâce à la méthode de *cross-validation*. Cette méthode utilise un ensemble de données représentatives du problème qui n'ont pas participé à l'apprentissage (l'ensemble de validation), et qui sert à tester le réseau. Lorsqu'il y a sur-apprentissage, les performances du réseau continuent à s'améliorer sur l'ensemble d'apprentissage, mais se dégradent sur l'ensemble de validation. La figure 2.2 représente le comportement des erreurs en apprentissage et en validation au cours d'un apprentissage. L'itération t_0 de l'algorithme d'apprentissage correspond au minimum de l'erreur sur l'ensemble de validation, il convient donc d'arrêter l'apprentissage à ce point. La même figure 2.2 est obtenue quand on trace l'erreur RMS en fin d'apprentissage en fonction du nombre de poids du réseau. Le minimum de la courbe de validation permet alors de décider de la complexité du réseau. Les performances réelles du réseau se calculent sur un troisième ensemble de test. Celui-ci sert à vérifier que la modélisation s'est faite en toute généralité, et qu'elle ne dépend pas de l'ensemble de

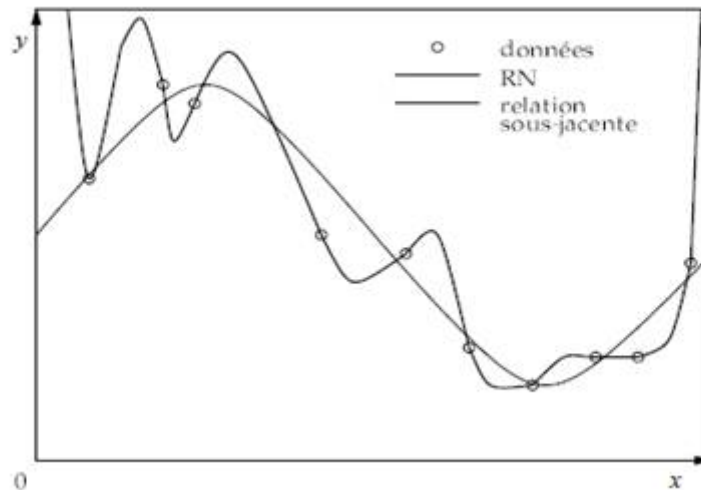


Figure 2.1 – Exemple de sur-apprentissage d'un réseau approximant une fonction $y = G(x)$ sur un ensemble de données réelles générées par cette loi.

validation. C'est sur l'ensemble de test que les résultats obtenus durant les diverses expériences seront évalués. Les performances en apprentissage et en validation ne seront en général pas précisées.

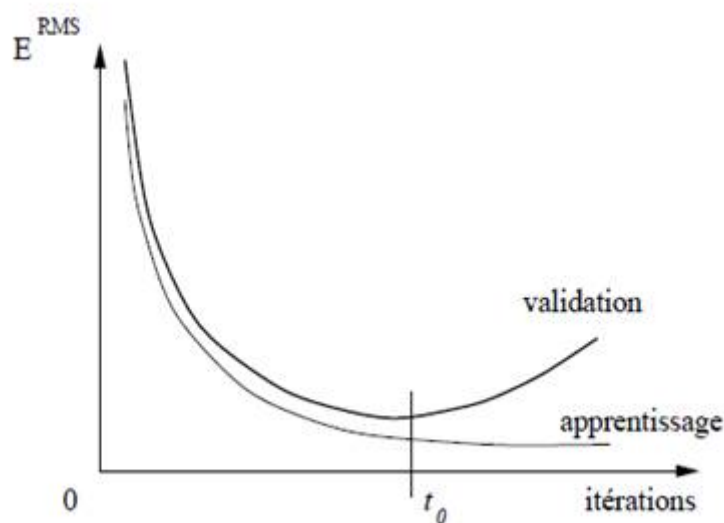


Figure 2.2 – Comportement des erreurs en apprentissage et en validation au cours d'un apprentissage

2.6 Les Perceptrons Multicouches

Dans cette partie, nous allons décrire en détail les formalismes mathématiques et algorithmiques des apprentissages des perceptrons multicouches. On abordera les apprentissages supervisés et non-supervisés de réseaux mono et multicouches en exposant de manière non exhaustive les différents types de règles d'apprentissages.

2.6.1 Neurone artificiel

Un neurone formel ou artificiel est un opérateur mathématique très simple. Un neurone possède des entrées qui peuvent être les sorties d'autres neurones, ou des entrées de signaux extérieures, et une sortie. La valeur de la sortie résulte du calcul de la somme des entrées, pondérées par des coefficients (dits poids de connexions ou poids synaptiques) et du calcul d'une fonction non linéaire (dite fonction d'activation) de cette somme pondérée. L'état du neurone, appelé aussi activité, est définie comme la somme pondérée de ses entrées. L'information est ainsi transmise de manière unidirectionnelle. Un neurone se caractérise par trois concepts : son état, ses connexions avec d'autres neurones et sa fonction d'activation. L'élément de base d'un réseau de neurone modélise une fonction $f(y)$ d'une sommation pondérée y de $n + 1$ signaux x_0, \dots, x_n qui lui parviennent (Figure 2.3) (Borme et al., 2007).

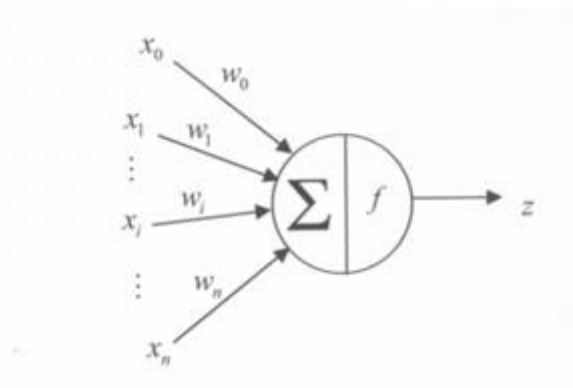


Figure 2.3 – neurone artificiel

$$\begin{cases} y = \sum_{i=0}^n w_i x_i \\ z = f(y) \end{cases} \quad (2.5)$$

Les coefficients de pondérations sont $w_i, i = 0, \dots, n$, appelées également les poids synaptiques. Si w_i est positif, l'entrée x_i est excitatrice alors que si w_i est négatif, elle est inhibitrice.

Codage : la dynamique des fonctions de transition détermine le codage des données traitées par le réseau. En effet, il est rare que l'on puisse utiliser des données numériques (géophysiques ou biologiques etc...) dans leurs propres unités sans craindre de sortir des domaines utilisés des fonctions de transition. De plus, il n'est pas rare qu'un réseau de neurone ait des données d'entrées de différentes natures donc de différents ordres de grandeur, alors qu'il est nécessaire que chaque variable ait le même poids numérique. De même, il est courant que les sorties du réseau aient aussi différents ordres de grandeur, ce qui est numériquement difficile à gérer. Aussi le réseau de neurone est en général la partie *calculateur* d'un système de traitement comprenant une phase de codage, le réseau de neurones lui même (RN), et une phase de décodage figure 2.4. Aussi notre boîte de codage est simple : pour ramener un maximum de données entre -1 et 1 et éviter les saturations, les ensembles de données se verront centrés-réduits (élément par élément du vecteur d'entrée) de la façon suivante : soit $\{x_i, i = 1 \dots N\}$ un ensemble de données :

$$\hat{x}_{ik} = \frac{2}{3} \times \frac{x_{ik} - \bar{x}_k}{\sigma_{x_k}} \quad (2.6)$$

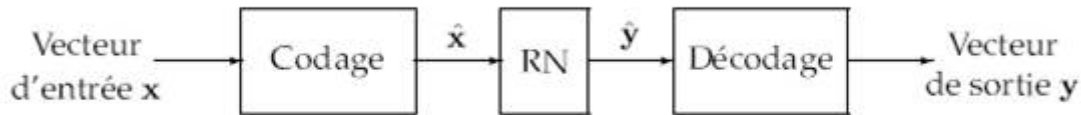


Figure 2.4 – Traitement de données par réseau de neurones

où x_{ik} est la valeur de la $k^{\text{ième}}$ dimension de la $i^{\text{ième}}$ donnée, \bar{x}_k est la moyenne de la $k^{\text{ième}}$ dimension sur les N données et σ_{x_k} son écart-type. Le facteur $\frac{2}{3}$ permet de ramener environ 80% de données entre -1 et 1 . Pour ramener toutes les données entre ces deux bornes, on peut utiliser généralement la formule suivante :

$$\hat{x}_{ik} = 1 - 2 \times \frac{x_{ik} - \min(x_k)}{\max(x_k) - \min(x_k)} \quad (2.7)$$

Le décodage inverse éventuellement ces formules. Il est important que les boîtes de codage et décodage sont fixés par le modélisateur et relève de connaissance à *priori* sur le problème.

2.7 Le Perceptron

2.7.1 Architecture et mise en équation

Le perceptron comporte une couche d'entrée constituée de n neurones élémentaires dont la fonction d'activation est linéaire et une couche de sortie constituée d'un ou plusieurs neurones dont la fonction d'activation de sortie est en général du type linéaire ou sigmoïde (Figure 2.5) (Borme et al., 2007). Notons w_{ik} le poids reliant le $j^{\text{ième}}$ neurone de la couche d'entrée au $i^{\text{ième}}$ neurone de la couche de sortie, il vient :

$$\left\{ \begin{array}{l} y_i = \sum_{j=1}^n w_{ij}x_j + w_{i0}x_0 \\ z_i = f(y_i) \text{ ou encore} \\ y_i = w_i^T x + w_0x_0 \\ \text{avec } w_i^T = [w_{1i}, w_{2i}, \dots, w_{1n}] \text{ et } x^T = [x_1, x_2, \dots, x_n] \end{array} \right. \quad (2.8)$$

Ici, x_0 correspond à l'entrée inhibitrice constante.

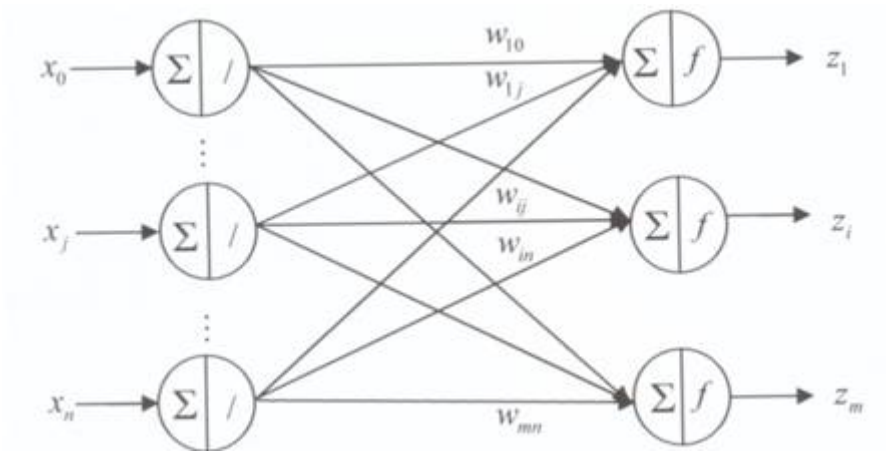


Figure 2.5 – Perceptron

Il est possible de décrire, le perceptron sous forme matricielle présentée sur la Figure 2.6 :

$$\left\{ \begin{array}{l} y = W_i^T x + w_0 x_0 \\ y = [y_1, y_2, \dots, y_m]^T \\ z = [z_1, z_2, \dots, z_m]^T \\ w_i^T = [w_{i0}, w_{i1}, \dots, w_{in}] \\ W = [w_1^T, \dots, w_m^T] = w_{ij} \\ W = \begin{bmatrix} w_1^T \\ \vdots \\ w_m^T \end{bmatrix} \end{array} \right. \quad (2.9)$$

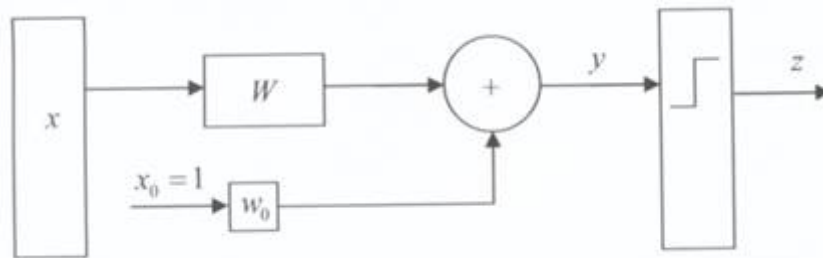


Figure 2.6 – Représentation matricielle du perceptron

2.7.2 Fonctionnement

Pour comprendre le fonctionnement du perceptron, prenons l'exemple du réseau de la Figure 2.7 avec 2 neurones dans la couche d'entrée et 5 neurones dans la couche de sortie.

$$\left. \begin{array}{l} w_i^T = [w_{i1}, w_{i2}] \\ w_0 = 0 \\ y_i = \sum_{j=1}^2 w_{ij} x_j = w_i^T x \end{array} \right\} \forall i = 1, 2, \dots, 5 \quad (2.10)$$

Si les vecteurs poids ont la même module :

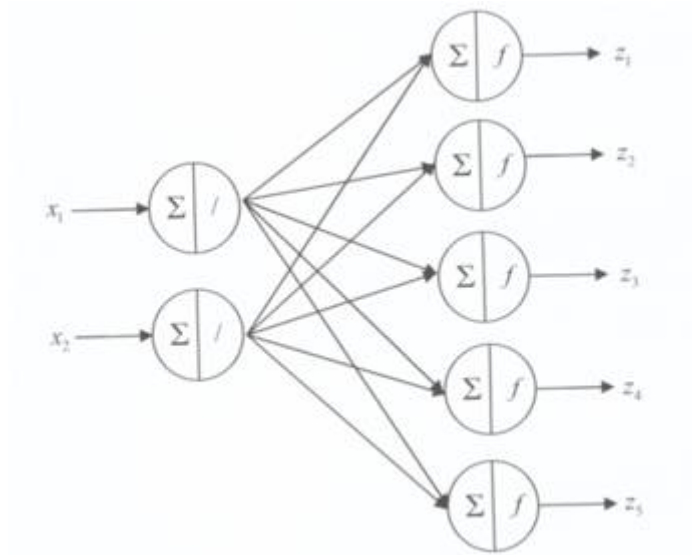


Figure 2.7 – Exemple de perceptron

$$w_i^T w_i = 1, \forall i = 1, 2, \dots, 5 \quad (2.11)$$

y_i représente le produit scalaire du w_i^T par x qui correspond dans ce cas, à la projection de x sur w_i . Le produit scalaire est maximal pour le vecteur w_i qui est la plus proche de l'entrée x . En associant la valeur 1 à la sortie z_{i_M} du neurone i_M pour lequel y_i prend la plus grande valeur : $y_i = y_{i_M}$ et la valeur 0 aux autres, il s'effectue une classification des positions des vecteurs d'entrées selon leur appartenance à l'un des cinq zones définies à partir des vecteurs de poids w_i Figure 2.8 [Borne, 2007]. Avec l'ajout

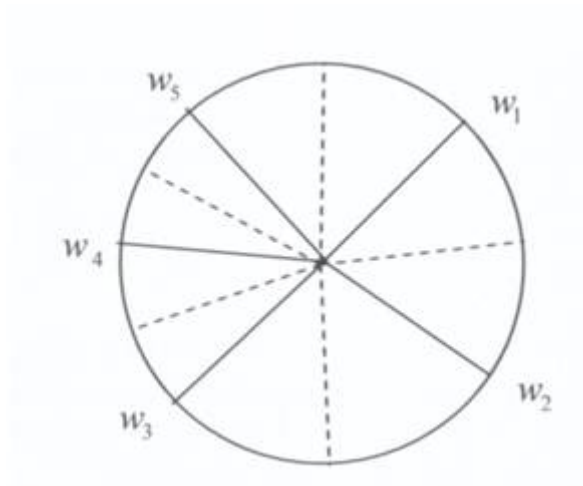


Figure 2.8 – Classification à partir du perceptron

de la fonction d'apprentissage supervisé ou non, il est possible d'apprendre à ce réseau à faire de la classification.

2.7.3 Réseau à couches

2.7.4 Réseau monocouche, présentation

Dans le paragraphe précédent, a déjà été présenté le perceptron qui comporte $n + 1$ entrées et m neurones de sortie, w_{ij} représente le poids caractérisant la liaison de la $j^{\text{ième}}$ entrée d'entrée au $i^{\text{ième}}$ neurone de sortie et w_{i0} représente le poids associé à l'entrée généralement inhibitrice par l'entrée $x_{i0} = 1$ équation 2.5. L'utilisation d'un automate linéaire à seuil pour la couche de sortie dite de décision, permet de réaliser pour chaque neurones i une partition des vecteurs d'entrée en deux classes dont la frontière est définie par l'hyperplan de dimension $n - 1$ et d'équation :

$$\sum_{j=1}^n w_{ij}x_j + w_{i0} = 0 \quad (2.12)$$

Le perceptron permet donc, ainsi qu'il a été souligné par Minisky et Papert de séparer que des exemples linéairement séparables. Par exemple, il ne peut réaliser le OU exclusif (ou XOR). Les réseaux multicouches permettent de pallier à cette limitation.

2.7.5 Réseaux multicouches

Dans ce cas, le réseau comporte en général au moins trois couches : une couche d'entrée, une ou plusieurs couche(s) cachée(s) et une couche de sortie. L'information circule de l'entrée vers la sortie à travers les couche(s) cachée(s) (Borme et al., 2007). Les notations sont les suivantes : $z_i^{(l)}$: sortie du $i^{\text{ième}}$ neurone de la couche à n_l neurones,

$$z^{(l)} = [z_1^{(l)}, z_2^{(l)}, \dots, z_{n_l}^{(l)}],$$

$$y^{(l)} = [y_1^{(l)}, y_2^{(l)}, \dots, y_{n_l}^{(l)}],$$

$w_{ij}^{(l)}$: les poids liant $j^{\text{ième}}$ de la couche $l - 1$ au $i^{\text{ième}}$ de la couche l

$f_i^{(l)}(\cdot)$: fonction d'activation du $i^{\text{ième}}$ de la couche l ,

$$z_i^{(l)} = x_i,$$

$$x = [x_1, x_2, \dots, x_n]^T$$

On a :

$$\begin{cases} y_i^{(l+1)} = \sum_{j=1}^{n_l} w_{ij}^{(l+1)} z_j^{(l)} + w_{i0}^{(l+1)} \\ z_i^{(l)} = f_i^{(l)}(y_i^{(l)}) \end{cases} \quad (2.13)$$

Les techniques d'apprentissage de ce type de réseau sont développées plus loin.

En notant : $W^{(l)} = w_{ij}^{(l)}$, $j = 1, \dots, n_{l-1}$; $i = 1, \dots, n_l$

$$w_{ij}^{(l)} = [w_{i0}^{(l)}, w_{i1}^{(l)}, \dots, w_{in_{l-1}}^{(l)}],$$

$$f^{(l)}(y^{(l)}) = [f_1^{(l)}(y_1^{(l)}), f_2^{(l)}(y_2^{(l)}), \dots, f_{n_l}^{(l)}(y_{n_l}^{(l)})],$$

On arrive à la représentation matricielle :

$$\begin{cases} z_i^{(l)} = f_i^{(l)}(y_i^{(l)}) \\ y^{(l+1)} = W^{(l+1)}z^{(l)} + w_0^{(l+1)} \end{cases} \quad (2.14)$$

2.7.6 Apprentissage

2.7.7 Différents types d'apprentissage

IL existe essentiellement deux types d'apprentissage, l'apprentissage non-supervisé et l'apprentissage supervisé.

2.7.8 Apprentissage non-supervisé

Dans ce cas, des exemples ou *prototypes* ou *patrons* sont présentés au réseau qu'on laisse s'auto-organiser au moyen des lois locales qui régissent l'évolution des poids synaptiques. Ce mode d'apprentissage est aussi appelé *apprentissage par compétition*.

2.7.9 Apprentissage supervisé

Dans ce type d'apprentissage, on cherche à imposer au réseau un fonctionnement donné en forçant à partir des entrées qui lui sont présentées, les sorties du réseau à prendre des valeurs données en modifiant les poids synaptiques. Il se comporte alors comme un filtre dont les paramètres de transfert sont ajustés à partir des couples entrée/sortie présentés. L'adaptation des paramètres du réseau s'effectue à partir d'un algorithme d'optimisation, l'initialisation des poids synaptiques étant souvent aléatoire.

2.7.10 Apprentissage supervisé du réseau monocouche

2.7.11 principe

Le but est de réaliser la fonction choisie :

$$s = g(x), g : \mathfrak{R}^n \rightarrow \mathfrak{R}^m$$

Notons $x \in \mathfrak{R}^n$, l'entrée du réseau et $z \in \mathfrak{R}^m$ sa sortie.

$$x = [x_1, x_2, \dots, x_n]^T,$$

$$y = [y_1, y_2, \dots, y_m]^T,$$

$$z = [z_1, z_2, \dots, z_m]^T$$

En l'absence d'entrée inhibitrice, il vient :

$$\begin{cases} y_i = w_i^T = \sum_j w_{ij} x_j \\ \text{et} \\ z_i = f(y_i) \end{cases} \quad (2.15)$$

2.7.12 Méthode de gradient totale

A la valeur $s(p)$ de l'entrée, correspond la sortie $z(p)$ du réseau et $s(p) = g(x(p))$. Soit un ensemble de P patrons, chacun étant défini par une paire $(x(p), s(p))$, avec $p = 1, \dots, P$.

Notons $z(p)$ la sortie du réseau lorsque le $p^{\text{ième}}$ patron est présenté. L'erreur sur l'ensemble des valeurs des patrons s'écrit :

$$E_{\text{totale}} = \sum_{p=1}^P \sum_{i=1}^m (z_i(p) - s_i(p))^2 \quad (2.16)$$

La méthode du gradient total consiste, en notant $w_{ij}(t)$ la valeur du poids w_{ij} à la $i^{\text{ème}}$ itération, à l'algorithme :

$$\begin{cases} w_{ij}(t+1) = w_{ij}(t) - \eta(t) \frac{\partial E_{totale}}{\partial w_{ij}} \\ \frac{\partial E_{totale}}{\partial w_{ij}} = \sum_{p=1}^P (z_i(p) - s_i(p)) \frac{\partial z_i(p)}{\partial w_{ij}} \\ \frac{\partial z_i(p)}{\partial w_{ij}} = \frac{\partial f(y_i(p)) \partial y_i(p)}{\partial y_i(p) \partial w_{ij}} \end{cases} \quad (2.17)$$

$\eta(t)$ étant le coefficient d'apprentissage. Il vient ainsi la nouvelle formulation de l'algorithme :

$$w_{ij}(t+1) = w_{ij}(t) - \eta(t) \sum_{p=1}^P (z_i(p) - s_i(p)) \frac{\partial f(y_i(p))}{\partial y_i(p)} x_j \quad (2.18)$$

L'algorithme d'apprentissage utilisant l'erreur totale conduit généralement à des calcul particulièrement complexes. Il est préférable d'utiliser la méthode dite rétropropagation du gradient, qui conduit en suite d'approximation partielle, couche par couche et patron par patron. Il vient pour le $j^{\text{ème}}$ patron présenté à la $i^{\text{ème}}$ itération, la relation :

$$z_i^{(3)}(t) = f\left[\sum_{i=1}^{n_2} w_{ij}^{(3)}(t) f\left(\sum_{k=1}^{n_1} w_{kj}^{(2)}(t) z_k^{(1)}(t) + w_{j0}^{(2)}(t)\right) + w_{i0}^{(3)}\right] \quad \forall i = 1, \dots, n_3 \quad (2.19)$$

et l'erreur à minimiser :

$$E(t) = \frac{1}{2} \sum_{i=1}^{n_3} (z_i^{(3)}(t) - s_i(t))^2 \quad (2.20)$$

Il vient :

$$w_{ij}^{(l)}(t+1) = w_{ij}^{(l)}(t) - \eta(t) \frac{\partial E_{totale}}{\partial w_{ij}^{(l)}(t)} \quad (2.21)$$

$$\begin{cases} \text{Pour : } & i = 1, 2, \dots, n_l; \\ & j = 1, 2, \dots, n_l \\ \text{et } & l = 2, 3 \end{cases}$$

1. Optimisation des $w_{ij}^{(3)}$

Pour l'erreur à minimiser :

$$\frac{\partial E(t)}{\partial w_{ij}^{(3)}} = (z_i^{(3)}(t) - s_i(t)) \frac{\partial z_i^{(3)}(t)}{\partial w_{ij}^{(3)}}$$

comme :

$$z_i^{(3)}(t) = f(y_i^{(3)}(t)),$$

il en résulte :

$$\begin{aligned} \frac{\partial z_i^{(3)}(t)}{\partial w_{ij}^{(3)}} &= \frac{\partial f(y_i^{(3)})}{\partial w_{ij}^{(3)}} \\ &= \frac{\partial f(y_i^{(3)}) \partial y_i^{(3)}}{\partial y_i^{(3)} \partial w_{ij}^{(3)}} \end{aligned} \quad (2.22)$$

soit :

$$\frac{\partial z_i^{(3)}(t)}{\partial w_{ij}^{(3)}} = \frac{\partial f(y_i^{(3)})}{\partial w_{ij}^{(3)}} z_i^{(3)} \quad (2.23)$$

Dans le cas d'une fonction d'activation de la forme : $f(v) = \frac{1}{1+e^{-\frac{v}{\tau}}}$

Il vient :

$$\frac{\partial f(y_i^{(3)})}{\partial y_i^{(3)}} = \frac{1}{\tau} [1 - f(y_i^{(3)})] \quad (2.24)$$

L'algorithme d'apprentissage admet alors l'expression :

$$w_{ij}^{(l)}(t+1) = w_{ij}^{(l)}(t) - \eta(t) \delta_i^{(3)}(t) z_i^{(3)}(t) \quad (2.25)$$

$$\left\{ \begin{array}{l} \text{Pour : } i = 1, 2, \dots, n_3; \\ j = 1, 2, \dots, n_2 \end{array} \right.$$

$$\delta_i^{(3)}(t) = (z_i^{(3)}(t) - s_i(t)) \frac{\partial f(y_i^{(3)})}{\partial y_i^{(3)}} \quad (2.26)$$

 2. Optimisation des $w_{jk}^{(2)}$

L'algorithme d'apprentissage prend la forme :

$$w_{jk}^{(2)}(t+1) = w_{jk}^{(2)}(t) - \eta(t) \frac{\partial E_{totale}}{\partial w_{jk}^{(2)}} \quad (2.27)$$

avec :

$$\frac{\partial E_{totale}}{\partial w_{jk}^{(2)}} = \sum_{i=0}^{n_3} (z_i^{(3)}(t) - s_i(t)) \frac{\partial f(y_i^{(3)}(t))}{\partial y_i^{(3)}} \frac{\partial y_i^{(3)}(t)}{\partial z_j^{(3)}} \frac{\partial f(y_i^{(2)}(t))}{\partial y_j^{(2)}} \frac{\partial y_j^{(2)}(t)}{\partial w_{jk}^{(2)}} \quad (2.28)$$

Comme :

$$\left\{ \begin{array}{l} \frac{\partial y_i^{(3)}(t)}{\partial z_j^{(3)}} = w_{jk}^{(3)} \\ \text{et} \\ \frac{\partial y_j^{(2)}(t)}{\partial w_{jk}^{(2)}} = z_k^{(1)} \end{array} \right.$$

Il vient :

$$w_{jk}^{(2)}(t+1) = w_{jk}^{(2)}(t) - \eta(t) \delta_j^{(2)}(t) z_k^{(1)}(t) \quad (2.29)$$

$$\left\{ \begin{array}{l} \text{Pour } : j = 1, 2, \dots, n_2; \\ k = 1, 2, \dots, n_1 \end{array} \right.$$

avec :

$$\delta_i^{(2)}(t) = \sum_{i=1}^{n_3} \delta_i^{(3)}(t) w_{ij}^{(3)}(t) \frac{\partial f(y_j^{(2)}(t))}{\partial y_j^{(2)}} \quad (2.30)$$

2.7.13 Accélération de la convergence

La méthode du gradient avec un coefficient d'apprentissage η faible conduit à une convergence lente alors qu'un coefficient élevé risque de produire des oscillations. Afin de remédier à ce problème tout en conservant un coefficient η élevé, l'algorithme suivant peut être utilisé :

$$w^{(l+1)}(t+1) = w^{(l)}(t) - \eta(t) \frac{\partial E_{totale}}{\partial w^{(l)}} + \gamma(w^{(l)}(t) - w^{(l-1)}(t-1)) \quad (2.31)$$

avec : $\gamma \in [0, 1]$,
par exemple $\gamma = 0.8$.

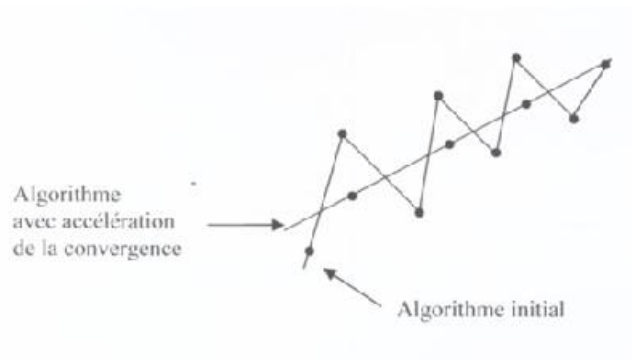


Figure 2.9 – Accélération de la convergence

2.8 Cartes topologiques auto organisatrices de Kohonen

L'objectif de cette partie est de rappeler le principe général des réseaux de neurones et de présenter les cartes topologiques auto-organisatrices. Une présentation complète de ces modèles peut être trouvée dans (Kohonen, 2001) et (Dreyfus et al., 2002). On rappelle qu'un réseau de neurones artificiels est un ensemble de processeurs élémentaires, les neurones, reliés entre eux et qui échangent des informations par ces liaisons. L'interaction entre ses différents éléments donne au réseau son comportement global. On appelle « état » du neurone la valeur de sortie de ce neurone. Un neurone calcule son état à partir des informations qui lui parviennent de l'extérieur du réseau ou bien de l'état des neurones auxquels il est connecté. Les réseaux de neurones constituent des modèles paramétrés dont les paramètres sont estimés au cours d'une phase dite « d'apprentissage » à partir d'un ensemble d'observations statistiquement représentatives du problème à traiter (base d'apprentissage). Suivant le type d'apprentissage, on distingue deux grandes familles de réseaux de neurones :

- l'apprentissage est « supervisé » lorsque l'on cherche à apprendre au réseau le lien qui existe entre deux ensembles de données (On veut régresser une variable en fonction de variables explicatives),
- l'apprentissage est « non supervisé » lorsqu'il s'agit d'analyser les structures d'un seul ensemble de données cohérentes entre elles, en ne donnant aucune indication a priori sur les structures recherchées.

La partie qui suit s'intéresse aux cartes de Kohonen qui font partie des méthodes non supervisées.

2.8.1 Principe de la modélisation

L'ensemble A de dimension n représente l'espace des observations possibles. Les données dont on dispose (N vecteurs réels de dimension n , représentatifs du problème à traiter) sont contenues dans un sous-ensemble $A = \{z_i; i = 1, \dots, N\}$ de D ($A \subset D$) qui constitue notre base d'apprentissage. La méthode utilisée, dite « partitionnelle » vise à résumer l'information contenue dans D , à partir de A , sous la forme d'un ensemble $W = \{w_c; c = 1, \dots, p\}$ de p vecteurs de D . Par la suite, ces vecteurs seront appelés les référents. Sachant que $p \ll n$, ce procédé permet donc de réduire considérablement la taille du problème. Chaque observation de A est affectée à son référent w_c au moyen d'une « fonction d'affectation » χ . C'est la connaissance de cette fonction qui permet de définir la partition $P = P_c; c = 1, \dots, p$ de D en p sous-ensembles. Chaque référent w_c est alors un représentant de l'ensemble P_c dont il résume l'ensemble des observations (Figure 2.10).

2.8.2 Architecture des cartes topologiques auto organisatrices de Kohonen

Le modèle utilisé dans le cadre de cette étude consiste à projeter de façon non linéaire notre ensemble de données de grande dimension sur un espace discret de faible dimension (2 dans notre cas), la carte (notée C). Une carte de Kohonen est un réseau de neurones à deux couches dont l'architecture peut se

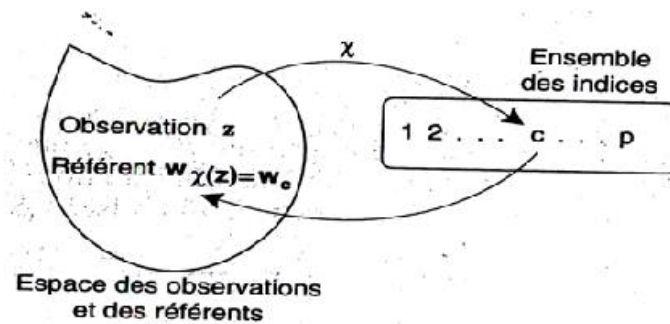


Figure 2.10 – Principe général de la modélisation : une observation z est associée à un indice c choisi parmi p l'aide de χ ; cet indice permet de définir le référent w_c (Dreyfus et al., 2002).

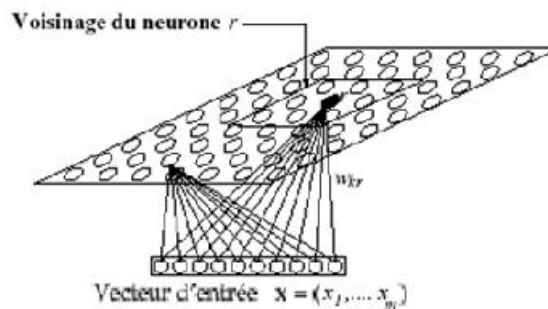


Figure 2.11 – Architecture d'une carte topologique en $2 - D$. Le réseau est constitué de deux couches : une couche d'entrée qui sert à la présentation des observations et une couche d'adaptation (pour laquelle il faut définir un voisinage, figure 2.12) formée d'un treillis régulier à 2 dimensions dont chaque nœud est occupé par un neurone, qui est lui-même connecté à tous les éléments de la couche d'entrée. Chaque neurone c est affecté d'un référent w_c (Dreyfus et al., 2002).

définir de la manière suivante :

- La couche d'entrée, servant à la présentation des données, prend la valeur du vecteur d'observation z_i ,
- La couche d'adaptation est formée d'un treillis régulier à 2 dimensions dont chaque nœud est occupé par un neurone, qui est lui-même connecté à tous les éléments de la couche d'entrée (Figure 2.11).

Entre deux neurones est définie une distance discrète δ comme étant la longueur du plus court chemin entre ces neurones sur la carte C . Pour chaque neurone c , cette distance permet de définir le voisinage d'ordre d , qui représente l'ensemble des neurones dont la distance au neurone c est inférieure ou égale à d (Figure 2.12). Cette notion de voisinage définit, avec la forme de la carte, ce qu'on appelle la « topologie » de la carte. Le terme « topologique » est lié à la particularité de ce type de cartes. Elles prennent en compte la continuité des données dans $D : C$ est à dire que deux neurones voisins dans C (distance discrète) doivent représenter deux ensembles d'observations voisines dans D (distance

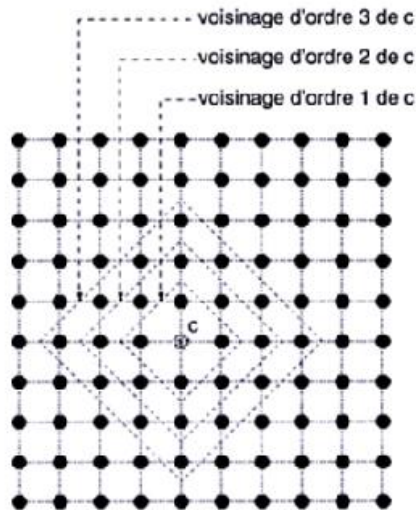


Figure 2.12 – Représentation de la topologie discrète d’une carte à deux dimensions constituée de $10 * 10$ neurones ; chaque point de la figure représente un neurone c . La distance δ entre deux neurones est définie sur le maillage. Elle permet de définir le voisinage d’ordre d (ici $d = 1, 2$ ou 3), qui représente l’ensemble des neurones dont la distance au neurone c est inférieure ou égale à d (Dreyfus et al., 2002).

euclidienne) (Figure 2.15). Quant au terme « auto-organisatrice », il fait référence à la technique dite de l’apprentissage non supervisé évoquée précédemment. En fin d’apprentissage, chaque neurone de la carte représente un sous-ensemble P_c de données, et est donc associé au référent w_c correspondant (Figure 2.15). Le nombre d’observations captées par un neurone à l’issue de l’apprentissage est appelé « cardinalité ».

2.8.3 Apprentissage à l’aide de l’algorithme SOM

La construction des cartes est rendue possible grâce aux algorithmes partitionnels, qui permettent la détermination de w et δ en minimisant une « fonction de coût ». Ils font appel au formalisme des nuées dynamiques, qui procède par itérations successives. Dans notre étude, nous avons utilisé l’algorithme Self Organizing Map (SOM). L’apprentissage se présente alors de la manière suivante :

- **Initialisation** : à $t = 0$, les référents w_c sont initialisés aléatoirement entre les valeurs minimales et maximales des observations,
- **Compétition** : à l’itération t , chaque neurone calcule sa distance à partir de l’observation $z(t)$ présentée en entrée. A l’issue de cette étape, le neurone c dont le référent w_c est le plus proche de l’observation $z(t)$, au sens de la distance euclidienne, est déclaré « gagnant ». L’observation est alors affectée au sous-ensemble P_c associé à ce neurone.
- **Adaptation** : les référents dans le voisinage du neurone ayant capté l’observation $z(t)$ vont être modifiés pour se rapprocher de celui-ci. Globalement, plus un neurone est proche du neurone gagnant et plus il sera influencé. A mesure que l’apprentissage progresse, la taille du voisinage du neurone gagnant décroît. Autrement dit, au cours du temps de moins en moins de neurones seront

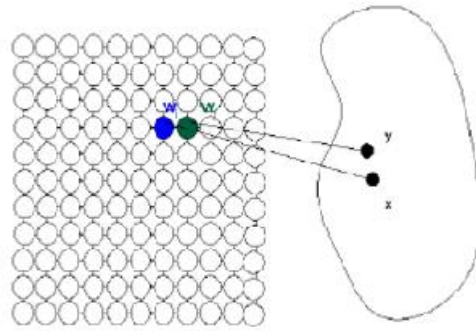


Figure 2.13 – Conservation de la topologie des données : deux neurones voisins dans C (distance discrète) doivent représenter deux ensembles d’observations voisines dans D (distance euclidienne) (Lachkar and Gruber, 2012).

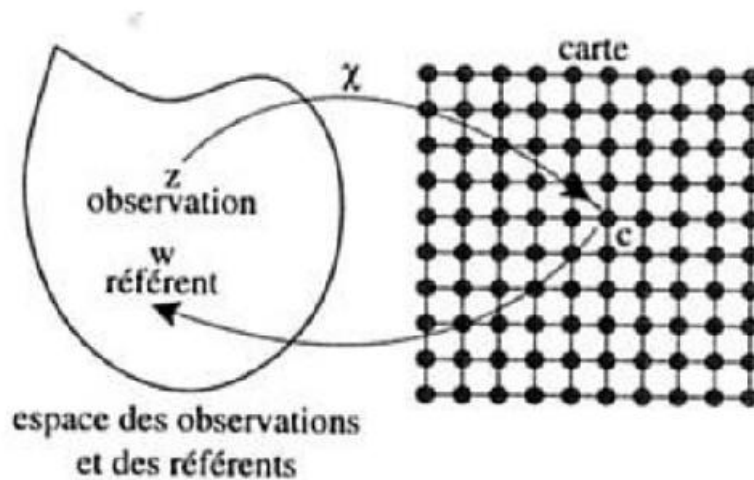


Figure 2.14 – Principe général de la modélisation par carte auto-organisatrice : une observation z est associée à un indice c choisi sur la carte (parmi p neurones) à l’aide de la fonction χ ; cet indice permet de définir le référent w_c (Dreyfus et al., 2002).

influencés par le neurone gagnant. Cette décroissance est régie par un paramètre communément appelé « Température » que nous faisons diminuer progressivement. Un autre paramètre décroît également lors de cette phase. Il s’agit du « pas d’apprentissage » qui permet de faire converger le référent de chaque neurone vers une valeur stable. Cette valeur représente une observation « moyenne » résumant le sous-ensemble P_c des observations qui lui sont affectées (Figure 2.15).

2.8.4 Algorithme de Kohonen

On peut résumer un cycle de l’algorithme SOM de la manière suivante : Cet algorithme se fait en deux étapes :

1. Etape d’initialisation :

- Fixer la structure et la taille de la carte ; Choix aléatoire des vecteurs référents ;

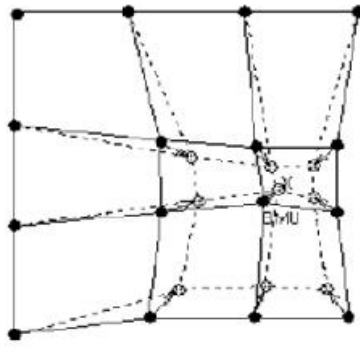


Figure 2.15 – Adaptation du neurone gagnant et de son voisinage : plus un neurone est proche du neurone gagnant et plus il sera influencé (Lachkar and Gruber, 2012).

- Fixer le nombre d'itération $Niter$;
- Fixer les valeurs T_{max} et T_{min} (température initiale et finale) ;
- Prendre $t = 0$; initialisation à 0 du nombre d'itération ;

2. Etape itérative :

- Choix aléatoire d'une observation z ;
- Calcul de la nouvelle valeur de T : $T = T_{max} * \left(\frac{T_{min}}{T_{max}}\right)^{\frac{t}{Niter-1}}$;
- Phase d'affectation $c \quad \chi(z) = \arg \min |z - W_r|^2$
- Phase de minimisation ;
 Pour tout neurone r de la carte :
 $W_r(t) = W_r(t-1) + \alpha(t)K^T(\delta(c,r))(W_r(t-1) - z(t))$ où
 $\delta(c,r)$ est la distance calculée entre les neurones c et r
 $\alpha(t)$: pas d'apprentissage fonction du nombre d'itération

3. Condition d'arrêt

Répéter l'étape itérative en faisant décroître T jusqu'à atteindre $t = Niter$. Un bon minimum de la fonction de coût s'obtient en présentant chacune des observations à un grand nombre de fois.

2.8.5 Classification ascendante hiérarchique

La méthode de classification hiérarchique ascendante (CAH) a pour but la construction d'une typologie ou encore partition des individus en classes telles que les individus appartenant à une même classe sont proches, alors que les individus appartenant à des classes différentes sont éloignés.

2.8.6 Qualité d'une typologie

Considérons une typologie d'un ensemble d'individus en k classes d'effectifs respectifs n_1, \dots, n_k . Notons G_1, \dots, G_k la typologie correspondante du nuage de points associé $N = x_1, \dots, x_n$ et g_1, \dots, g_n les centres de gravité de ces classes. L'inertie totale du nuage N se décompose de la manière suivante :

$$I(N, g) = \sum_{i=1}^k \left(\frac{n_i}{n}\right) d^2(g_i, g) + \sum_{i=1}^k \frac{n_i}{n} I(G_i, g_i) \quad (2.32)$$

Le premier terme de droite s'appelle l'inertie inter-classes et mesure la manière dont les classes s'éloignent les unes des autres. Ce terme est noté $I(G_1, \dots, G_k)$ et représente l'inertie expliquée par la typologie. Le deuxième terme de droite s'appelle l'inertie intra-classes et mesure l'homogénéité des classes. On mesure la qualité de la typologie à l'aide du rapport inertie inter-classes sur inertie totale.

2.8.7 Le critère de Ward

Lorsque dans la typologie G_1, \dots, G_k on remplace deux classes G_i et G_j par leur réunion $G_i \cup G_j$, il y a diminution de l'inertie inter-classes. Cette diminution : $D(G_i, G_j) = I(G_1, \dots, G_i, \dots, G_j, \dots, G_k) - I(G_1, \dots, G_i \cup G_j, \dots, G_k)$ peut se calculer et vaut :

$$D(G_i, G_j) = \frac{n_i n_j}{n(n_i + n_j)} d^2(g_i, g_j) \quad (2.33)$$

Ce critère est utilisé pour mesurer la distance entre les deux classes G_i et G_j et s'appelle le critère d'agrégation de Ward.

2.8.8 Algorithme CAH

L'algorithme de classification ascendante hiérarchique est itératif. À l'étape courante, on part d'une partition de l'ensemble des individus en k classes G_1, \dots, G_k et on groupe les deux classes G_i, G_j minimisant le critère de Ward $D(G_i, G_j)$. À l'étape initiale, chaque individu forme une classe et l'inertie totale est alors égale à l'inertie inter-classes. À l'étape finale, il n'y a plus qu'une seule classe et l'inertie inter-classes est donc nulle. La somme des pertes d'inertie inter-classes des différentes étapes est donc égale à l'inertie totale. À chaque étape, on calcule un indice obtenu en divisant la perte d'inertie inter-classes par l'inertie totale. On choisit la typologie obtenue à l'étape correspondant à une augmentation brutale de l'indice.

2.9 Conclusion

Dans ce chapitre, nous avons décrit d'abord l'historique des réseaux de neurones, ensuite les concepts de bases ont été mis en évidence en passant par le neurone formel, l'apprentissage et la généralisation. Par ailleurs, cette partie a fait l'objet de la mise en équation mathématique des concepts d'apprentissages et les algorithmes correspondants. Les cartes topologiques de Kohonen ont été présentées ainsi son algorithme sans oublier la classification ascendante hiérarchique CAH.

Toutes ces théories expliquées ci-dessus seront appliquées sur des données issues des différents capteurs (SeaWiFS, AVHRR, ...) dans la suite de cette thèse. Ces applications devraient permettre la résolution de problèmes hautement non-linéaires tels que ceux rencontrés dans le domaine des sciences atmosphériques.

Chapitre 3

Présentation de données

3.1 Introduction

Ce chapitre a pour objet de passer en revue les différents jeux de données utilisés au cours de ce travail de thèse, tout en décrivant succinctement le rôle capital de la télédétection de satellites ainsi les grandeurs physiques mesurées.

3.2 Historique de la télédétection

La télédétection se définit comme une technique qui permet, à l'aide d'un capteur, "**d'observer**" et d'enregistrer le rayonnement électromagnétique, émis ou réfléchi, par une cible quelconque sans contact direct avec celle-ci. Le traitement et l'analyse des informations véhiculées par le rayonnement enregistré permettent d'accéder à certaines propriétés de cette cible : géométriques (position, forme et dimensions), optiques (réflexion, transmission, absorption, etc.) et physico-chimiques (température, teneur en eau, chlorophylle foliaire, matière organique du sol, . . .), etc.

L'imagerie satellitaire s'est grandement développée au cours de la seconde moitié du XXe siècle et les plateformes spatiales se sont multipliées et spécialisées afin de répondre aux besoins accrus de la communauté scientifique en matière d'observation de la planète. En effet, la télédétection est un outil d'inventaire (en occupation des sols par exemple), d'analyse (en météorologie par exemple) et d'aide à la prévision (en agriculture notamment). Ses enjeux sont non seulement scientifiques mais également économiques et stratégiques. C'est pourquoi tant de nations se sont lancées dans cette conquête spatiale afin d'avoir les meilleurs outils d'observation possibles. Le premier satellite de télédétection d'observation de la terre est le satellite météorologique TIROS-1 qui a été lancé en 1960. Il avait à son bord un radiomètre visible qui lui permettait de voir les nuages et un radiomètre proche infra-rouge qui lui permettait de mesurer la température de surface de la terre et des océans. Les satellites de la série TIROS exploités par la NOAA et la plupart des services météorologiques mondiaux sont des satellites polaires. Ils ont à bord des radiomètres infra-rouge performants permettant de mesurer avec une précision du 1/10ème de degré la température de surface des océans. Le premier satellite imageur mesurant la surface de la terre dans plusieurs bandes visibles a été lancé en 1972. Il s'agit de LANDSAT-1, lancé par

la NASA. A partir de là, de nombreux programmes spatiaux ont été réalisés afin d'affiner la qualité des capteurs et des images obtenues.

L'océanographie a grandement bénéficiée de ces méthodes de télédétection. En effet, l'imagerie spatiale permet d'obtenir des informations concernant une large zone (tout un océan) en très peu de temps, en comparaison avec le temps mis pour faire une campagne en mer. Pour répondre à ce cahier des charges, de nouveaux satellites ont dû être lancés. Les premiers satellites de télédétection à vocation purement océanographique furent CZCS (premier satellite dédié à la mesure de la couleur de l'océan) puis SeaSAT (premier satellite altimétrique) tous deux lancés en 1978. Malheureusement la durée de vie de SeaSAT fut très courte (108 jours). Il a été suivi 12 ans plus tard par les satellites altimétriques Topex/Poséidon en 1992, ERS1/2 puis la série JASON. CZCS dont les mesures se sont arrêtées en 1986, a été suivi 10 ans plus tard des capteurs satellitaires SeaWiFS en 1997 MODIS en 1999 de bien meilleure résolution spectrale et spatiale. Actuellement, il existe plusieurs capteurs détectant la couleur des océans (Ocean Color Sensor) qui sont en activités et d'autres sont venus s'ajouter comme : OCTS-China (Ocean Color and Temperature Scanner), GLI (Global Imager, Japon) et POLDER-2 (POLarization and Directionnality of the Earth's Reflectances, France), tous les trois en 2002 et VIIRS (Visible Infrared Imager/Radiometer Suite, USA) en 2005.

3.2.1 Les capteurs

Les capteurs à bord de satellites ou aéroportés, appelés radiomètres (quelques bandes spectrales) ou spectrophotomètres (multi bandes), enregistrent des luminances, c'est-à-dire des intensités de lumière. Ces intensités sont exprimées en watts par unité de longueur d'onde, par unité angulaire et par unité de surface. La luminance et la réflectance sont les deux principales unités utilisées en télédétection. Le paragraphe suivant présente les grandeurs physiques utilisées en télédétection.

3.2.2 Quantités physiques mesurées en télédétection

En télédétection, les quantités mesurées sont le rayonnement réfléchi ou émis. Énergie rayonnée par une source dans tout l'espace environnant : E : Unité en Joule(J).

3.2.3 Flux (radiant flux)

Énergie rayonnée par unité de temps :

$$\left\{ \begin{array}{l} \Phi = \frac{dE}{dt} \\ \text{Unité en Watt(W)} \end{array} \right. \quad (3.1)$$

3.2.4 Intensité énergétique (radiant intensity)

Flux émis dans un angle solide selon une direction donnée :

$$\left\{ \begin{array}{l} I = \frac{d\Phi}{d\Omega} \\ \text{Unité : W.sr}^{-1} \text{ (stéradian, angle solide)} \end{array} \right. \quad (3.2)$$

3.2.5 Éclairement (irradiance)

Flux reçu par unité de surface :

$$\left\{ \begin{array}{l} E = \frac{d\Phi}{dS} \\ \text{Unité : W.m}^{-2} \text{ (stéradian, angle solide)} \end{array} \right. \quad (3.3)$$

3.2.6 Réflecteurs spéculaires et lambertiens (Specular and Lambertian reflectors)

Un objet peut être un réflecteur spéculaire, c'est-à-dire qu'il renvoie toute la lumière reçue, mais dans la direction opposée à l'éclairement, c'est le cas du miroir. Un objet peut aussi être un réflecteur lambertien ; dans ce cas, il renvoie la lumière de façon égale dans toutes les directions. C'est presque le cas d'une feuille de papier blanc qui renvoie la lumière de façon à peu près égale dans toutes les directions. Un objet peut également être un réflecteur complexe, comme la surface de l'eau qui renvoie davantage de lumière dans la direction opposée à l'éclairement. On est ébloui par la réflexion du soleil sur l'eau. Cette réflexion dépend aussi des ondulations de surface (clapotis, vagues, etc.).

3.2.7 Reflectance et albédo (réflectance and albedo)

La réflectance est le taux d'énergie (lumière ici) renvoyé par l'objet. Ainsi, la neige a une réflectance proche de 90 %, c'est-à-dire qu'elle renvoie 90 % de la lumière qu'elle reçoit. Par contre, un papier noir aura une réflectance proche de 0 %. Cette réflectance qui dépend de la longueur d'onde, est aussi appelée albédo, La réflectance diffuse de l'eau s'écrit :

$$R(\lambda) = \frac{E_u(\lambda_0-)}{E_d(\lambda_0-)} \quad (3.4)$$

où $E_u(\lambda_0-)$ est l'éclairement remontant (upwelling irradiance) à la longueur d'onde λ juste sous la surface de l'eau et $E_d(\lambda_0-)$ est l'éclairement descendant (downwelling irradiance) à la longueur d'onde λ juste sous la surface de l'eau. Ce facteur de réflectance varie de 0 à 1 ou de 0 à 100 %.

3.2.8 Luminance (radiance)

Intensité énergétique émise d'un élément de surface dS considéré dont la normale est inclinée de θ par rapport à Ox (Bonn and Rochon, 1992) :

$$L(\theta) = \frac{1}{\Delta A \cos(\theta)} \quad (3.5)$$

où ΔA est l'unité de surface (m^2), $\cos(\theta)$ est la projection de la surface et l'unité est $W.m^2.sr^{-1}$. Si la luminance est exprimée en fonction de la longueur d'onde, elle est dénommée luminance spectrale $L(\theta, \lambda)$ et les unités sont en $W.m^2.sr^{-1}.\mu m^{-1}$.

3.2.9 Réflectance télédéctée Rrs (remote sensing reflectance)

La réflectance télédéctée est de plus en plus utilisée pour les mesures de terrain (Mobley, 1999).

$$\begin{cases} Rrs(\lambda) = \frac{L_w(\lambda)}{E_d(0^+, \lambda)} \\ \text{Unite : } sr^{-1} \end{cases} \quad (3.6)$$

où L_w est la luminance émergente mesurée juste au-dessus de l'eau et E_d l'éclairement descendant mesuré également juste au-dessus de l'eau. Deux grandeurs supplémentaires sont utilisées pour les données satellitaires.

3.2.10 Luminance normalisée de l'eau $L_{WN}(\lambda)$ (normalised water-leaving radiance)

Cette définition a été introduite récemment pour les mesures effectuées au moyen de capteurs satellitaires. Cette luminance normalisée est définie suivant la relation [(Gordon et al., 1988); (Yang and Gordon, 1997)]

$$\begin{cases} L_{WN}(\lambda) = \frac{L_w(\lambda)}{E_0(\lambda, \theta_0)} \\ \text{avec : } L_w(\lambda) = \frac{E_0(\lambda)}{Rrs(\lambda)} \\ \text{Unite : } \mu W.cm^{-2}.nm^{-1}.sr^{-1} \end{cases} \quad (3.7)$$

Dans cette équation, $E_0(\lambda, \theta_0)$ est l'éclairement descendant juste au dessus de la surface de l'eau avec un angle solaire θ_0 . Les données spatiales concernant le domaine marin, sont souvent fournies sous la forme $L_{WN}(\lambda)$. C'est le cas des données de niveau 2 fournies par la NASA pour les capteurs SeaWiFS et MODIS et par l'ESA pour les données POLDER et pour les données MERIS. Dans le cas où tient compte de l'effet directionnel, on obtient la réflectance normalisée (normalised water leaving reflectance) ρ_{WN} qui a pour forme :

$$\rho_{WN}(\lambda) = \frac{\pi L_w(\lambda)}{E_0(\lambda) \cos(\theta)} \quad (3.8)$$

3.3 Couleur de l'eau et télédétection

3.3.1 Pourquoi s'intéresser au phytoplancton ?

Grâce à la détection de la concentration de la chlorophylle, il est possible d'estimer la productivité primaire, c'est-à-dire la quantité de carbone intégré par les plantes sous forme de matière organique lors de la photosynthèse chlorophyllienne. De cette manière, le phytoplancton joue un rôle important dans le cycle du carbone. En effet, à l'échelle mondiale, le phytoplancton produit au moins la moitié de

l'oxygène que nous respirons. Plus de 99.9 % du **CO₂** incorporé par les organismes vivants tout au long des temps géologiques sont enfouis dans les sédiments marins. De plus, le phytoplancton est la base de toute la chaîne alimentaire marine ; des changements dans sa répartition et/ou son abondance peuvent avoir un impact négatif sur tout l'écosystème marin. La croissance du phytoplancton nécessite de la lumière solaire, de l'eau, du **CO₂** et des nutriments. C'est ce dernier élément qui est le facteur limitant la croissance du phytoplancton.

3.3.2 Le rôle du phytoplancton dans le cycle du carbone

La concentration de **CO₂** dans l'atmosphère et celle dans les océans sont en équilibre. Lors de la photosynthèse, le phytoplancton utilise le **CO₂** contenu dans les couches de surface océaniques ce qui diminue sa concentration dans ces couches et libère de l'oxygène en tant que sous-produit. Cette diminution a pour conséquence de permettre aux océans l'absorption de **CO₂** additionnel provenant de l'atmosphère. De plus, lors de leur mort, les organismes phytoplanctoniques tombent au fond des océans et sont petit à petit recouverts par d'autres organismes morts formant ainsi les sédiments marins. En piégeant le carbone de cette manière, les océans jouent le rôle d'un important puits de carbone.

3.3.3 Régulation de la quantité de phytoplancton dans les océans

Ainsi que nous l'avons vu plus haut, le facteur limitant la croissance du phytoplancton est l'abondance de nutriments (nitrates, phosphates, silicates, ...). Ceux-ci se trouvent en grande quantité dans les eaux froides et profondes des océans qui remontent à la surface dans les zones d'upwelling ou dans les zones d'eaux froides peu stables qui peuvent être le siège de mouvements de mélanges verticaux importants favorisant l'apport de nutriments stockés dans les eaux de sub-surface et les eaux profondes. De cette manière, la quantité de phytoplancton présent dans les océans est étroitement liée à la température des eaux de surface. Il existe cependant d'autres formes d'apport de nutriment dans les eaux océaniques. Les fleuves et rivières se jetant à la mer après avoir charrié de la matière en suspension lessivée tout au long de leur parcours terrestre en font partie. De manière générale, toutes les arrivées d'eaux douces à la mer, qu'elles soient à l'air libre ou souterraines, sont chargées de nutriments de part leurs parcours terrestres. Sont comprises dans ce lot les eaux de surface ayant lessivé les fertilisants agricoles, les eaux usées, les jus de décharges, ... en bref, toutes les eaux polluées. Un des moyens de discriminer entre ces différentes sources de nutriments possible est de corrélérer la concentration de chlorophylle avec la température de surface des océans. Si les zones de forte concentration correspondent avec des zones de basses températures et si elles sont variables au cours des saisons, il y a de fortes chances pour que cet apport de nutriments soit naturel. Si, au contraire, les zones de fortes concentrations ne varient pas au fil des saisons et ne sont pas corrélées avec la température de surface des océans, il se peut que les nutriments soient d'origine anthropique. Les apports provenant de cours d'eaux sont eux-mêmes considérés comme polluants. En effet, leurs eaux ont drainé tout un bassin versant de fertilisants et d'eaux usées rejetées dans les rivières. Dans le cas du phosphate, par exemple, dont la concentration naturelle est très faible et peu soluble, une fois parvenu à l'embouchure des fleuves, va couler vers le fond et sédimenter. Au contraire, le phosphate d'origine anthropique est hautement soluble et va de cette manière intervenir dans la croissance du phytoplancton. C'est ainsi que l'on comprend mieux pourquoi

la détection de la concentration de chlorophylle dans les océans (ainsi que l'estimation de la production primaire) est importante. Elle permet d'indiquer les zones de forte production biologique, de cerner les limites entre des eaux de forte et de faible productivité, d'indiquer les régions où la concentration en nutriments est élevée, de visualiser le transport des sédiments et les interactions entre les courants (le phytoplancton étant un organisme se laissant passivement porter par les courants, il est considéré comme un marqueur passif). D'autres applications à partir des données issues de la télédétection marine sont la recherche sur les propriétés optiques des océans, la surveillance de la pollution et de l'eutrophisation côtière, la variabilité saisonnière et annuelle, la variabilité des courants, le transport sédimentaire à partir des deltas, l'influence régionale de phénomènes de type El Nino et des moussons, les recherches pour la pêche, la corrélation des phénomènes océaniques et des maladies infectieuses, la remise en suspension et le transport des sédiments par les tempêtes, la surveillance des coraux, ...

3.3.4 La couleur des océans

La couleur des océans est déterminée par les interactions de la lumière solaire incidente avec les substances et les particules présentes dans l'eau. La matière en suspension est en grande partie composée de phytoplancton ainsi que de particules inorganiques, de substances dissoutes,... Le phytoplancton est un organisme photosynthétique contenant de la chlorophylle, un pigment absorbant la lumière dans les longueurs d'ondes rouge et bleue et, ce qui explique sa couleur verte. Lorsque la lumière solaire illumine la surface des océans, elle est sujette à plusieurs effets optiques, tels que la réflexion et l'absorption. Une partie est directement réfléchi (= Sun Glint), mais la majeure partie pénètre sous la surface de l'océan et interfère avec les molécules d'eau et la matière en suspension rencontrées. L'absorption de l'eau élimine en premier lieu les longueurs d'ondes les plus longues (rouge) mais permet la transmission des longueurs d'ondes les plus courtes (bleu), ce qui explique la couleur généralement bleue des océans.

Cette couleur est toutefois altérée par la présence de matières en suspension organiques, telles que le phytoplancton vivant, de matières en suspension minérale, telles que les sédiments et les détritiques, et de matière organique dissoute. Les zones deltaïques constituent un exemple d'eaux chargées de matières en suspension variées dont la couleur apparaît plutôt brune.

L'intensité du rayonnement émis par la surface des océans résultant de ces interactions (water-leaving radiance) est mesurée par des radiomètres dans certaines longueurs d'onde. Ces mesures peuvent ensuite être mises en relation avec les divers constituants de la colonne d'eau interférant avec la lumière visible, tels que la chlorophylle. La concentration de chlorophylle peut par la suite être utilisée pour estimer la production primaire qui est la quantité de matière organique, et donc de carbone, fixée lors de la photosynthèse chlorophyllienne. L'utilisation de capteurs placés en orbite autour de la Terre permet une vue d'ensemble globale qu'il serait impossible à obtenir à partir d'un bateau ou d'une installation terrestre. De plus, le passage régulier des satellites au-dessus d'une zone donnée permet une surveillance constante et donc le suivi autant spatial que temporel d'un phénomène donné. Les buts fondamentaux de la télédétection dans le domaine de la couleur des océans sont d'une part de mesurer la concentration de la chlorophylle et d'autre part de déterminer avec précision la productivité primaire du phytoplancton.

La manière principale de déterminer la concentration de la chlorophylle est de mesurer la quantité de lumière absorbée pour une longueur d'onde particulière, 443 nm. Des algorithmes analytiques

développés par les chercheurs en optique océanographique permettent de convertir ces mesures en concentration de chlorophylle a. Dans la pratique on utilise le rapport des reflectances $L_w(443)/L_w(555)$, l'absorption à $555nm$ due à la chlorophylle étant quasi nulle, la reflectance à cette longueur d'onde est donc indépendante de sa concentration. Dans les derniers algorithmes de type **OC4V4** on utilise le Max $L_w(443,490,510)/L_w(555)$. De plus le fait de prendre un rapport de réflectance atténuée

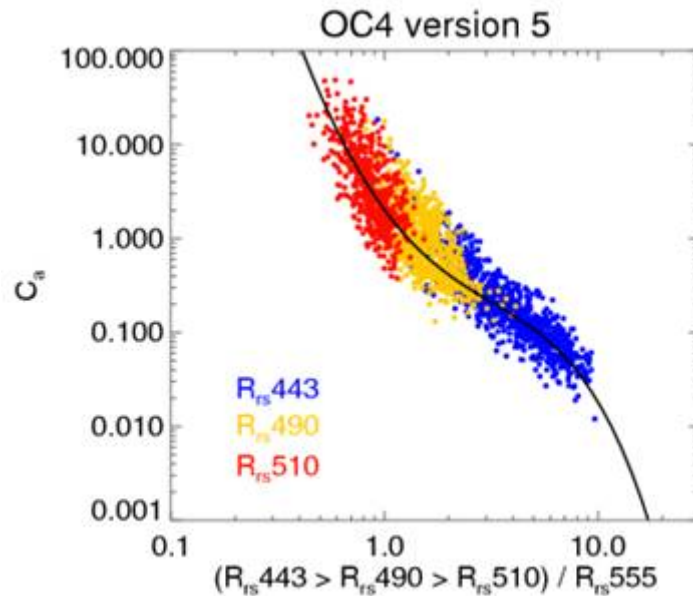


Figure 3.1 – Principe de l'algorithme OC4V4. En abscisse on a le rapport de réflectance marine (en log10,), en ordonnée la concentration en Chl-a en log10)

systématiquement l'effet de l'atmosphère, cet effet étant voisin pour deux longueurs d'onde proches. Il existe d'autres molécules de chlorophylle, appelés b, c, d, et e, mais la chlorophylle-a prédomine et est la plus importante.

3.3.5 Les données de couleur de l'océan : SeaWiFS

Le choix du capteur s'est porté sur SeaWiFS pour plusieurs raisons. D'une part, ce capteur a été tout particulièrement développé (sur la base de son prédécesseur CZCS) pour fournir des données océanographiques et d'autre part, l'accessibilité et la gratuité des données, ainsi que la possibilité de les traiter soi-même à l'aide d'un logiciel spécialement conçu à cet effet par la NASA, SeaDAS, le rendent très attrayant. De plus, SeaDAS est également gratuitement téléchargeable à partir d'internet. Ainsi, un des buts de ce travail est aussi de se familiariser et d'explorer les outils offerts par ce logiciel. La résolution spatiale de SeaWiFS est de 1km et la fréquence de ses passages au-dessus d'une zone donnée est quotidienne. Ce capteur permet ainsi un bon suivi de l'évolution des paramètres traités à une relativement grande échelle. Les bandes de SeaWiFS et leur utilité sont reportées dans le tableau 3.1.

Tableau 3.1 – Les bandes de SeaWiFS et leurs utilités.

Bande	Longueur d'onde centrale (nm)	Usage
1	412 (violet)	Matière organique dissoute (incl. Gelbstoff)
2	443 (bleu)	Absorption de la chlorophylle
3	490 (bleu-vert)	Absorption des pigments (cas 2), K 490
4	510 (bleu-vert)	Absorption de la chlorophylle
5	555 (vert)	Pigments, propriétés optiques, sédiments
6	670 (rouge)	Corrections atmosphériques
7	765 (proche infra-rouge)	Corr. atmosphériques, radiance des aérosols
8	865 (proche infra-rouge)	Corr. atmosphériques, radiance des aérosols

Notes :

- Les bandes 1 à 6 sont larges de 20 nm alors que les bandes 7 et 8 ont une largeur de 40 nm.
- Le Gelbstoff (substances jaunes) décrit la matière organique amorphe de grand poids moléculaire. Il absorbe fortement dans la région bleue du spectre.
- Les termes Cas 1 et Cas 2 se réfèrent à un type d'eau défini par ses caractéristiques optiques. Les eaux du Cas 1 sont les eaux claires des mers ouvertes alors que le Cas 2 caractérise généralement les zones côtières où les eaux sont plus turbides et sujettes à une plus grande productivité.
- K 490 est le coefficient d'atténuation diffuse à 490 nm, qui correspond à maximum d'absorption des pigments photo-synthétiques.

Les données de SeaWiFS sont accessibles sous la forme de plusieurs produits divisés en différents niveaux. Les fichiers de données de niveau 1A (Level 1A LAC (Local Area Coverage)) contiennent les données brutes de radiance pour chacune des 8 bandes du capteur, des données de calibration et de navigation ainsi que des données sur la télémétrie des instrument et de la plate-forme satellitaire. La résolution est de 1.13 km au nadir. Elles sont disponible au travers des stations au sol HRPT (High Resolution Picture Transmission). Les données de niveaux 1A GAC (Level 1A GAC (Global Area Coverage)) sont constituées d'un ré échantillonnage des données 1A LAC prenant en compte une donnée sur 4. La résolution spatiale se voit alors diminuée à 4.5 km au nadir. Les données de niveau 2 (Level 2 GAC) sont un produit dérivé des fichier L1A contenant des valeurs géophysiques. les produits principaux sont :

- la radiance normalisée quittant l'eau (normalized water-leaving radiances) pour les bandes 412, 443, 490, 510, 555 et 670,
- la concentration de chlorophylle-a,

- le coefficient d'atténuation diffuse K490,
- le coefficient d'Angström à 510-865 nm (Angstrom coefficient),
- l'épsilon de correction aérosol à 765 et 865 nm,
- l'épaisseur optique des aérosols à 865 nm.

Le niveau 3 (level 3 binned data) consiste en des moyennes issues des valeurs géophysiques de niveau 2. Ces moyennes peuvent être soit journalières, hebdomadaires, mensuelles ou annuelles.

Dans le cadre de ce travail, ce sont les données de niveau L1A LAG qui nous intéressent à partir desquelles nous avons effectué le traitement sur les images sélectionnées à l'aide de SeaDAS afin d'obtenir les paramètres choisis.

3.3.6 Les données de température de la surface de la mer : Sea Surface Temperature SST

Les données satellites de SST (Sea Surface Température) ont été utilisées pour l'étude de la dynamique de l'upwelling. Les données de SST AVHRR Pathfinder et MCSST (Multi-Channel High resolution Radiometer SST), diffusées par le JPL (Jet Propulsion Laboratory, <http://podaac.jpl.nasa.gov>) ont une résolution spatiale d'environ 4 km et 0.176°C (soit environ 18 km), respectivement, et une résolution temporelle hebdomadaire ou mensuelle. Le produit Pathfinder utilisé est le produit du niveau 3 (level 3) de la chaîne de traitement du JPL et les infra-rouges utilisés pour la mesure ne traversent pas les nuages : les données contiennent donc de nombreux trous. Par ailleurs, du fait de l'impact des gaz (vapeur d'eau, CO₂, azote, ..) atmosphériques sur les mesures, des corrections doivent être apportées aux données brutes ; l'erreur associée aux données traitées est au final de 0.3°C à 0.7°C. Le produit MCSST, lui, a été élaboré par interpolation optimale à partir des données AVHRR, il n'y a plus de données manquantes, en revanche l'erreur associée est un peu plus grande : estimée à 0.5 – 0.7°C. Le produit MCSST utilisé recouvre la période 1997-2007.

3.4 Conclusion

Dans cette partie, nous avons brossé l'historique de la télédétection satellitaire. Les capteurs abords de ces satellites fournissent des données pour la compréhension de processus physiques et biogéochimiques de l'océan. Les données couleur de l'océan sont fournies par le capteur SeaWiifs disponible sur le site de la NASA. Plusieurs types de données existent, dans ce travail de thèse, nous nous sommes intéressés aux données de niveau L2 et L3. Les données de la température de la surface de mer (SST) sont fournies AVHRR Pathfinder et c'est le produit du niveau 3 (L3) qui ont été utilisées. Toutes ces données vont servir dans les méthodes statistiques tout au long de cette thèse.

Chapitre 4

Analyse des situations océaniques à l'aide de cartes de Kohonen

4.1 Introduction

Cette partie présente l'application des réseaux de neurones exposés dans les chapitres précédents à l'étude de l'upwelling Sénégal-Mauritanien. Pour cela nous allons utiliser des méthodes neuronales en particulier les cartes topologiques de Kohonen suivies de la classification hiérarchique ascendante afin de caractériser l'évolution spatio-temporelle du phénomène étudié. On utilise des données issues de capteurs satellitaires mesurant la température de surface de la mer et la couleur de la mer ceci à plusieurs longueurs d'ondes. On décrit dans cette partie le phénomène de l'upwelling Sénégal-Mauritanien, en suite la méthode développée pour calculer une moyenne particulière appelée «**climatologie**». Cette étude permettra de mettre en évidence des zones bien définies par des indices spécifiques. Ces indices que nous avons identifiés puis caractérisés, peuvent être considérés comme des indicateurs statistiques permettant d'étudier la variabilité du phénomène étudié.

4.2 Upwelling Sénégal-Mauritanien

Les upwellings côtiers sont des zones océaniques très productives du point de vue biologiques. Plus de 80 % de la productivité océanique se passe dans les zones d'upwelling. Les mécanismes régissant les upwellings ont été abondamment décrits dans la littérature scientifique [(O'Brien and Hurlburt, 1972); (Allen, 1973)]. La signature de l'upwelling est bien marquée sur les observations satellitaires de température de surface (SST) (Figure 4.1) et de couleur de l'océan (Figure 4.2) marquant la teneur en phytoplancton, ce qui permet de suivre leur variabilité spatio-temporelle. Les profils longitudinaux de SST et de Chl-a montrent la forte signature de la zone d'upwelling caractérisée surtout durant les mois d'hiver par des températures froides et des concentrations élevées en Chl-a (Figure 4.3). On remarque que le minimum de température n'est pas situé à la côte mais à quelques dizaines de kilomètres au large (visible en Février) comme l'avait déjà signalé (Demarcq and Faure, 2000).

Dans ce qui suit, nous nous proposons de décrire la variabilité saisonnière et interannuelle de l'upwelling Sénégal-Mauritanien à partir des observations satellitaires de SST et de couleur de l'océan.

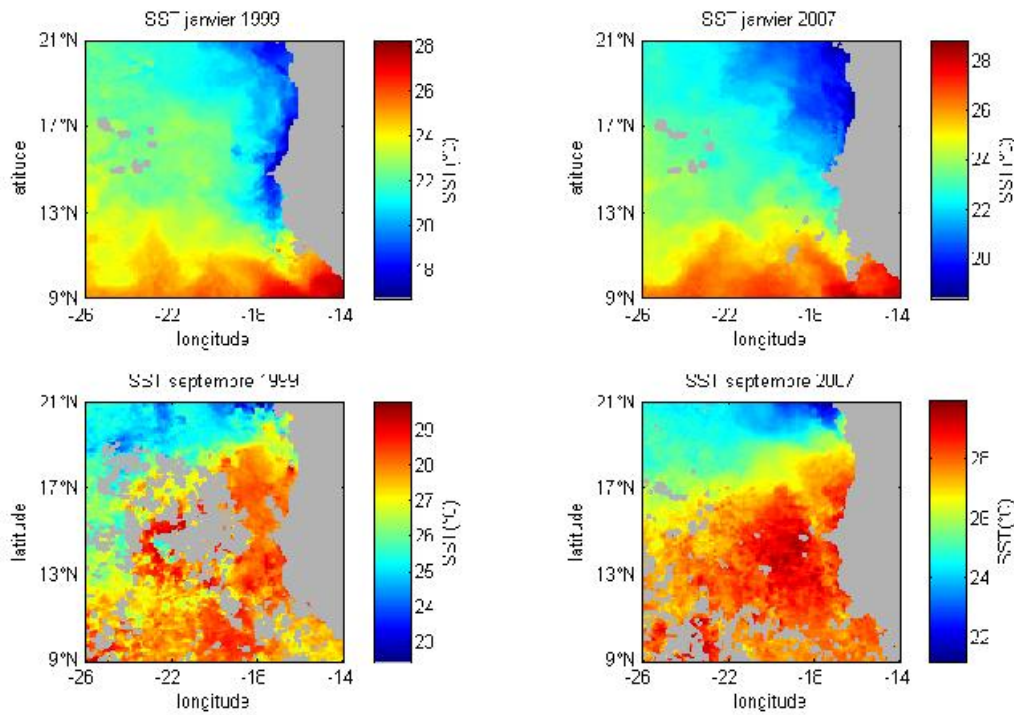


Figure 4.1 – Haut : –SST de Janvier 1999 et de Janvier 2007 ; Bas : –SST de Septembre 1999 et de Septembre 2007

Due à la nécessité de moyenner les images satellitaires sur des périodes suffisamment longues pour filtrer la couverture nuageuse qui est importante pendant la saison pluvieuse d'été (Juillet, Aout, Septembre), nous avons considéré dans cette expérience les images mensuelles uniquement. A cette échelle temporelle, la SST et la concentration en chlorophylle (dénotee Chl-a dans ce qui suit) mesurés par satellite sont des indicateurs de l'activité de l'upwelling, la SST reflétant les caractéristiques dynamiques de l'upwelling, tandis que la Chl-a étant une signature de l'activité biologique et aussi de la dynamique via l'apport en sels nutritifs. L'objectif recherché dans cette partie est d'identifier les différentes situations de l'upwelling Sénégal-Mauritanien, et de caractériser leur variabilité. Pour cela nous avons analysé 10 ans (de 1998 à 2007) de SST mesurées par le capteur AVHRR (Advanced Very High Resolution Radiometry) des satellites NOAA (National Oceanic and Admospheric Admistation) et de *Chl_a* mesurées par le capteur multi-spectral SeaWiFS (Sea-viewing Wide Field-of-view Sensor).

Comme l'upwelling est caractérisé par un gradient de température entre la côte et le large, une idée simple est de le caractériser par un indice de SST (dénotee ISST)[(Demarcq and Faure, 2000)] défini comme la différence à une latitude donnée (puisque la côte est approximativement Nord-Sud) entre la SST moyenne dans la bande 22° - 26°W et la température en un point donné situé entre la côte et ladite bande de latitude. Cette procédure à l'avantage de filtrer les variations saisonnières de température de surface qui peuvent être importante et celles dues à des phénomènes de très grande échelle. Etant donné ses variations importantes et sa distribution statistique (beaucoup de petites valeurs et peu de grandes) Chl-a est caractérisée par son log 10, quantité fréquemment utilisée par les bio-physiciens. Les images que nous analysons s'étendent de 9°N à 21°N et de 14°W à 26°W. Les pixels des bases de données que

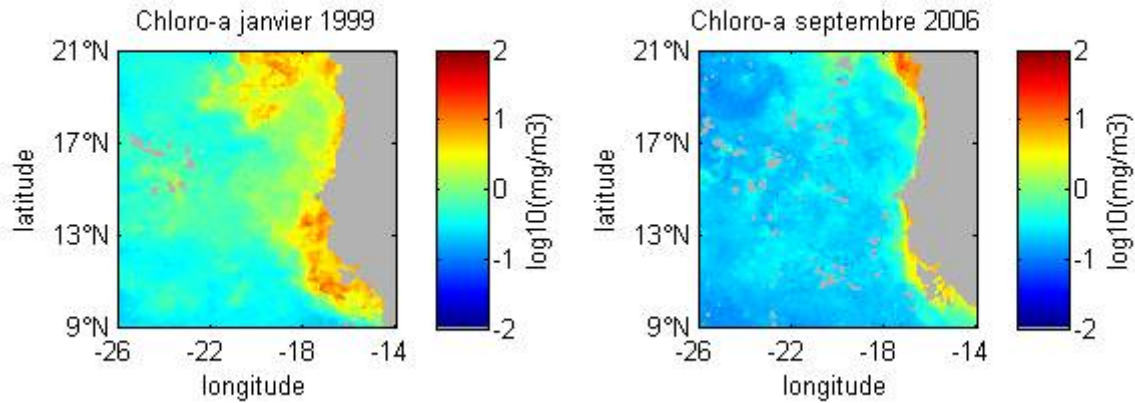


Figure 4.2 – Image de la Chl-a en Janvier 1999 et en Septembre 2006. On remarque une forte variation saisonnière et aussi une forte variation d’une année à autre.

nous utilisons ont une superficie d’environ 11.1x11.1 km de telle sorte que chaque image analysée est composée de 121x121 pixels. Chaque pixel de la zone d’upwelling est donc caractérisé par un vecteur à deux dimensions dont les composantes sont [Log10 (Chl-a) et ISST].

4.3 Bases de données utilisées

4.3.1 Les températures de surface SST

Les données de SST sont issues du capteur *NOAA/AVHRR* et sont récupérées via le site web (<http://las.pfeg.noaa.gov/OceanWatch/>). La résolution spatiale est : 0.1° de latitude et la résolution temporelle est mensuelle. La précision est de l’ordre de 0.1°C. La période analysée va de 1998 à 2007 ; le type de donnée récupérée est codé en ASCII (fichiers .txt). Chaque fichier du mois de donnée est représenté par une matrice de 121x121 donnant les températures SST_{xy} aux Pixels P_{xy} (les indices x et y représentant les coordonnées géographiques). Chaque Pixel P_{xy} correspond à un point géographique bien précis qui est le même pour chaque P_{xy} de chaque matrice. Comme chaque année a 12 mois et on a donc 12 matrices (janvier à décembre). Le nombre total de matrice de données pour ces 10 ans est 12x10. Pour l’apprentissage du réseau de neurone SOM, nous avons utilisé des climatologies mensuelles dont on décrit la procédure de calcul dans le paragraphe ci-après.

4.3.2 Calcul de la climatologie

Pour calculer la climatologie SST du mois de janvier : on considère l’ensemble des mois de janvier de la période (*janvier_1998* à *janvier_2007*, dénotée respectivement $SST_{jan1}, \dots, SST_{jan10}$; chaque matrice est constituée à la fois des pixels valides (ayant une valeur *SST*) et pixels non valides (nuages ou la terre) et chacune ayant une taille de 121x121. On transforme ensuite chaque matrice du mois en une matrice colonne de 41641x1, le résultat de la transformation est représenté par la matrice.

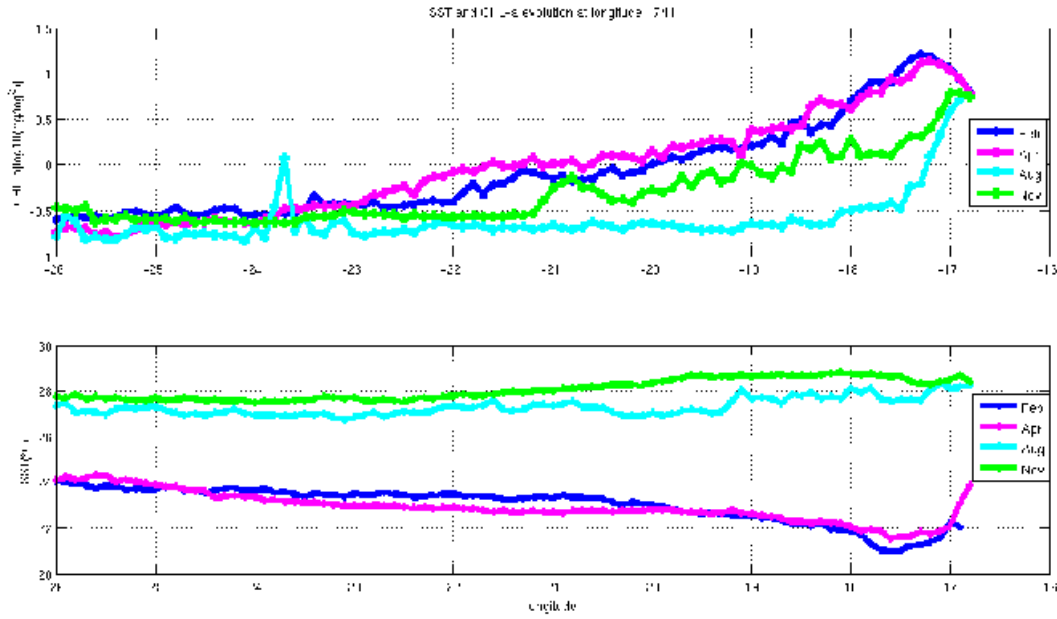


Figure 4.3 – Profils longitudinaux de température (haut) et de chlorophylle (bas) pour les mois de Février, Avril, Aout et Novembre à $17^{\circ}N$. On remarque la zone d'upwelling marquée par des températures froides et une forte concentration en Chl-a près de la côte et la forte variabilité saisonnière.

$$SST_{jan} = \begin{bmatrix} SST_{jan1}(1,1), \dots, SST_{jan10}(1,12) \\ SST_{jan1}(2,1), \dots, SST_{jan10}(2,12) \\ \vdots \quad \dots \quad \vdots \\ SST_{jan1}(14641,1), \dots, SST_{jan10}(14641,12) \end{bmatrix}$$

Le calcul de la moyenne se fait ensuite ligne par ligne de la matrice SST_{jan} ont le résultat est stocké dans la matrice $MSST_{jan}$ équation : 4.1. Les matrices $MSST_{jan}(1), \dots, MSST_{jan}(14641)$ représentent des résultats intermédiaires. Il faut noter que la valeur k dépend de nombre de pixels valides sur chaque ligne de la matrice SST_{jan} .

$$\begin{bmatrix} MSST_{jan}(1) = \frac{1}{k} \sum_{j=1}^k SST_{janj}(1, j) = \frac{1}{k} (SST_{jan1}(1,1) + \dots + SST_{jan10}(1,10)) \\ MSST_{jan}(2) = \frac{1}{k} \sum_{j=1}^k SST_{janj}(2, j) = \frac{1}{k} (SST_{jan1}(2,1) + \dots + SST_{jan10}(2,10)) \\ \vdots \quad \dots \quad \vdots \\ MSST_{jan}(14641) = \frac{1}{k} \sum_{j=1}^k SST_{janj}(14641, j) = \frac{1}{k} (SST_{jan1}(14641,1) + \dots + SST_{jan10}(14641,10)) \end{bmatrix}$$

$$MSST_{jan} = \begin{bmatrix} MSST_{jan}(1) \\ MSST_{jan}(2) \\ \vdots \\ MSST_{jan}(14641) \end{bmatrix} \quad (4.1)$$

En appliquant une transformation spécifique à la matrice colonne de $MSST_{jan}$, on retrouve une nouvelle matrice appelée matrice climato(dénotée $MSST_{xy}$) dont la structure matricielle est identique aux matrices initiales. La procédure est réitérée pour les autres mois (*fevrier_1998* à *fevrier_2007* jusqu'à

decembre_1998 à *decembre_2007*). A la fin, nous obtenons 12 matrices de $MSST_{xy}$ (*climato_janvier* à *climato_decembre*).

4.3.3 Indice de température (ISST)

Puisque l'upwelling se manifeste par un gradient de température entre la côte et le large, une idée simple est de le caractériser par un indice de SST (Sawadogo et al., 2009) (dénomé $ISST$ dans ce qui suit). Cet indice est défini comme la différence, à une latitude donnée, entre la SST moyenne dans la bande $22^\circ - 24^\circ W$ et la température en un point donné situé entre la côte et ladite bande. Il est calculé pour chaque mois de l'année en faisant la différence entre la moyenne de la SST dans la bande (22° et $24^\circ W$) au large (valeur dénotée SST_{Wy}) et la SST_{xy} en un pixel de la côte.

$$ISST_{xy} = SST_{xy} - SST_{Wy} \quad (4.2)$$

Cette procédure a l'avantage de filtrer les variations saisonnières de température de surface qui peuvent être importante et celles dues à des phénomènes de très grande échelle. On calcule de la même façon que précédemment la climatologie mensuelle des $ISST$. A l'issue de cette démarche, on obtient 12 matrices $MISST_{xy}$ représentant la climatologie mensuelle des $ISST$ ($ISST_{janvier}$ à $ISST_{decembre}$).

4.3.4 Les données de couleur de l'océan

Les données chlorophylle qui sont issues du capteur SEAWIFS sont recueillies sur le site web : (<http://reason.gsfc.nasa.gov/OPS/Giovanni/ocean.seawifs.shtml>). On utilise la même procédure que celle décrite pour les SST pour obtenir les climatologies mensuelles de Chl-a. Il est à remarquer de toutes les images de la climatologie sont lissées car elles sont moyennées. Pour résumer, on dispose de 2 types de fichiers de données :

- le fichier F1 composé de SST , $ISST$ et Chl-a qui comprennent chacun $121 \times 121 \times 12 \times 10$ valeurs. Ce sont les fichiers que l'on va analyser. Ils sont constitués de valeurs mensuelles s'étalant sur les 10 ans d'observation,
- le Fichier F2 composé de $MSST$, $MISST$ et MChl-a qui comprennent chacun $121 \times 121 \times 12$ valeurs. Ils sont constitués à partir de la base F1 précédente en faisant la moyenne de tous les mois de Janvier de tous les mois de Février, ceci de façon à avoir approximativement un même nombre de pixels pour chaque mois climato (contrainte pour l'apprentissage de SOM décrit dans le chapitre 2).

Le tableau 4.1 nous donne le résumé des fichiers dont on dispose.

4.3.5 Méthodologie

Etant donné le grand nombre de vecteur de la base de données décrivant l'upwelling et la bi-dimensionnalité de ces vecteurs, nous avons choisi d'utiliser une méthode de classification capable de

Tableau 4.1 – Fichiers de base de données.

Base de données	Taille	Type	Descriptif
F1	3x121x121x12x10	(<i>ISST</i> , <i>Chloro</i>) + <i>SST</i>	Données Originales
F2	3x121x121x12	(<i>MISST</i> , <i>MChloro</i>) + <i>MSST</i>	Climato mensuelle Moyennes Mensuelles sur 10 ans

faire ressortir les principales caractéristiques de l'upwelling. Le but est de trouver un certain nombre de vecteurs caractéristiques correspondant à des situations 'type' résumant les phases principales de l'upwelling. Afin d'avoir une base de donnée peu bruitée et d'en réduire la dimension, SOM a été appris à partir de la climatologie mensuelle F2 selon la procédure décrite dans (Badran et al., 2005). Les vecteurs de la base d'apprentissage sont des vecteurs à deux dimensions dont les composantes sont les moyennes mensuelles *MISST* et *MChl-a* au pixel P_{xy} . Afin de réduire la dimension de la base d'apprentissage, on échantillonne en prenant une ligne sur 3. Pour cette première expérience de la classification, on se servira des données climatologiques calculées précédemment sur la période de 10 ans. La Figure 4.4 présente les 2 variables utilisées lors de l'apprentissage (climatologie de janvier et novembre).

Paramètres et produits SOM2 (*MISST*, *MChl-a*)

Pour chaque expérience, le programme SOM offre un certain nombre d'options permettant de fixer différents paramètres de l'apprentissage :

- taille de la carte : en précisant le nombre des neurones et la structure de la carte. (Exemple : une taille 20x20 pour un treillis régulier, en forme de carré, de 400 neurones). Le voisinage est de type hexagonal par défaut.
- Température : en spécifiant deux valeurs entre lesquelles la fonction température (chapitre 2) décroît exponentiellement.
- seuil de voisinage : permet de se donner une valeur comme rayon maximum de voisinage.

En sortie du modèle, on récupère les produits suivants :

- fichier des poids des neurones après apprentissage,
- fichier de cardinalité des neurones (nombre d'observations captées par chaque neurone)
- fichier des "**neurones gagnants**"(ou BMUs) pour chaque observation.

Les cartes de distribution (« distribution maps » ou « component planes »), issues de la carte de Kohonen, sont représentées à la Figure 4.5. Elles permettent de visualiser la répartition des variables en fonction de leurs caractéristiques. On peut remarquer au passage une bonne organisation des neurones.

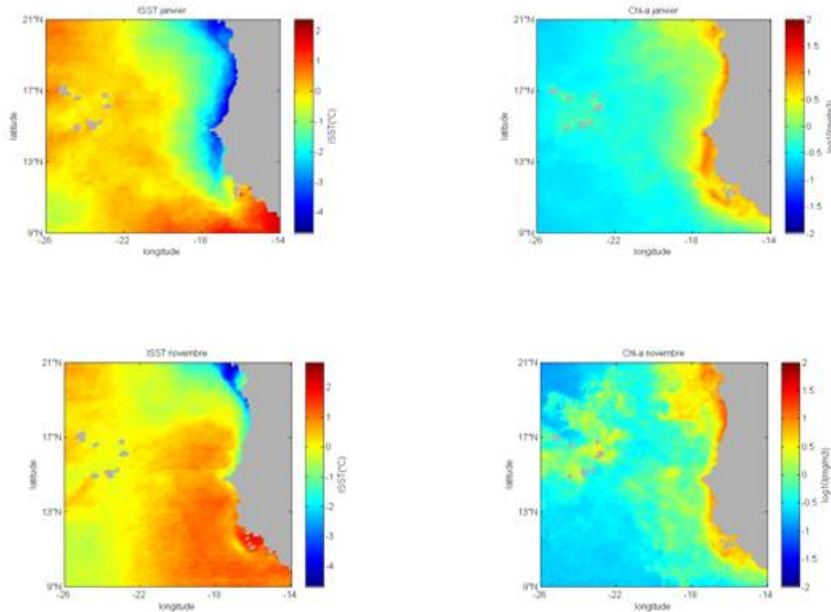


Figure 4.4 – Mois de janvier-novembre (climatologie)

4.3.6 Détermination des classes de SOM2

La U-matrix et la classification descendante hiérarchique de Ward ont ensuite permis de regrouper les cellules en classes (groupes). Le nombre de classes provenant de cette classification est laissé au choix de l'utilisateur : il doit donc être déterminé selon les objectifs de l'application et l'appréciation de l'analyste. Ainsi, nous avons choisi de distinguer six (6) classes car cela constituait un bon compromis entre un nombre de classes restreint et une bonne représentativité des différences de propriétés entre les variables. Les six classes ainsi obtenus sont représentés à la Figure 4.6. On constate sur la Figure 4.6 que la classe 6 est plus restreinte que les autres classes, ne comprenant que 14 neurones qui se distinguent par des valeurs élevées pour chacun des six. L'expertise de l'analyste est également nécessaire à cette étape pour valider ou corriger cette classification. Il est à remarquer que les différents groupes sont statistiquement homogènes (ils comprennent à peu près le même nombre d'éléments de F2) et que les six classes sont contiguës sur le graphe représentant SOM montrant ainsi que la relation topologique sur SOM fonctionne bien. Dans ce qui suit, nous analysons le comportement océanique de chacune des six classes. Chaque neurone de SOM2 a capturé un certain nombre de pixels de F2. Nous remarquons qu'il existe une bonne adéquation entre les classes, la concentration en Chl-a et les températures, montrant ainsi la pertinence de la décomposition en classes. Les chlorophylles les plus fortes et les températures les plus faibles sont captées par la classe 6 (upwelling) tandis que les chlorophylles les plus faibles et les températures les plus fortes (eaux du large) sont captées par les classes 1 et 2 (Figure 4.5).

Projetons les pixels associés à chaque classe sur une carte géographique (Figure 4.9). Nous remarquons que les six classes sont bien définies et correspondent chacune à une zone géographique bien délimitée montrant la cohérence de la classification. Cette cohérence géographique des pixels des classes qui est forte pour les données annuelles est aussi importante lorsque l'on projette les données mensuelles de chaque classe sur la carte géographique (Figure 4.9) ce qui nous permet d'analyser le

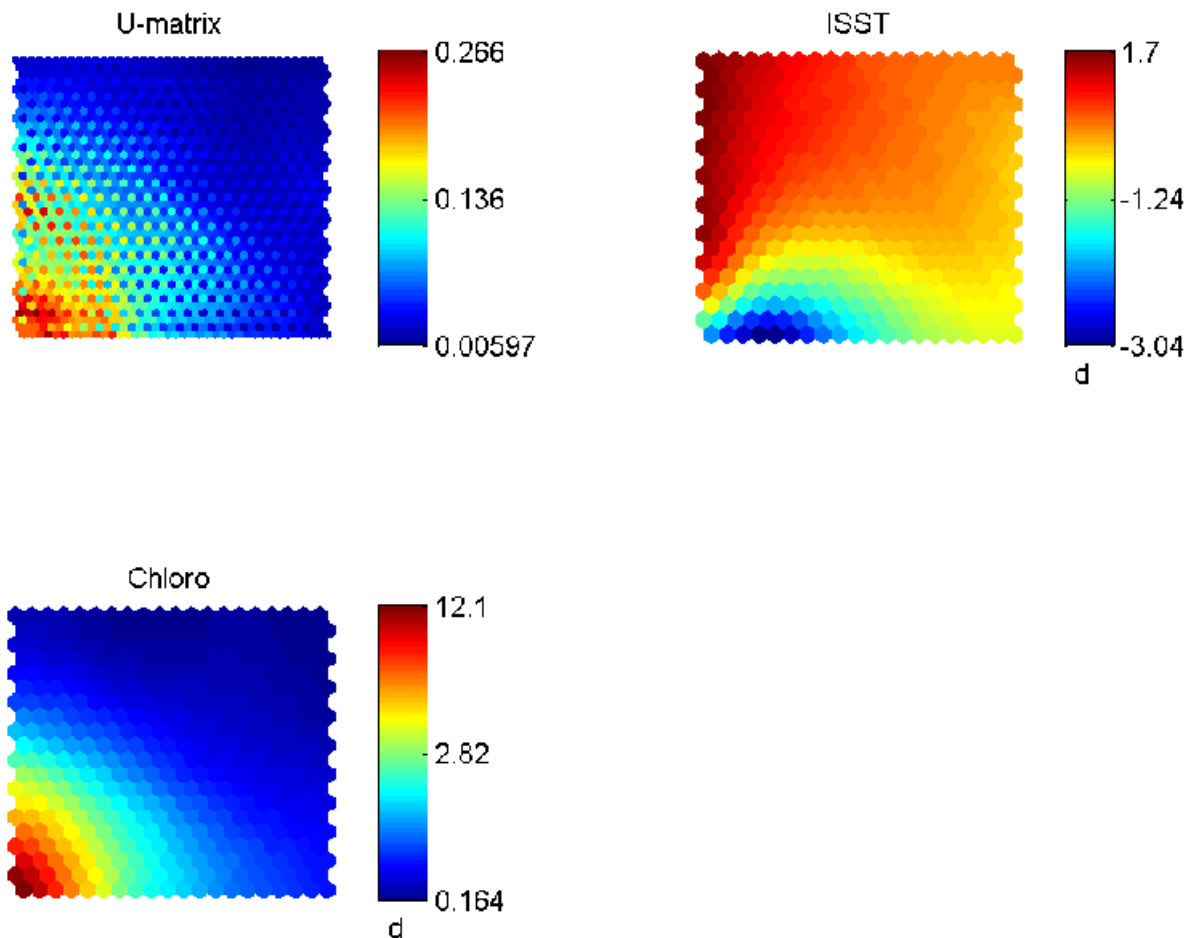


Figure 4.5 – Cartes de distribution issues de la carte auto-organisatrice de Kohonen. Les cellules bleues foncées représentent des valeurs faibles, alors que les cellules rouges foncées (couleur chocolat) représentent des valeurs élevées. L'échelle d indique les valeurs transformées de l'unité de variables.

comportement saisonnier de l'upwelling. On remarque que la variabilité mensuelle des classes 4, 5 et 6 est très importante. Décrivons d'abord les caractéristiques des six classes.

- La classe 1 (bleu foncée) correspond à des eaux où la concentration en Chl-a est très faible et la ISST voisine de zéro. Elle est associée aux eaux du large. Elle occupe une large partie de la carte SOM et est géographiquement très étendue. Les classes suivantes sont influencées par l'upwelling à des degrés divers. Elles occupent chacune une petite partie de la carte SOM2. En allant du large vers la côte on a :
- La classe 2 (bleu pale) correspondent à des eaux dont la concentration en Chl-a est un peu plus forte et dont l'ISST est faiblement négatives. Les eaux de la classe 2 jouxtent celles de la classe 1. La classe 3 (jaune pale) et la classe 4 correspondent respectivement à des eaux dont la concentration en Chl-a va en croissant et dont les ISST deviennent respectivement un peu plus négatives. La classe 5 est atypique. Elle est caractérisée par des eaux riches en chlorophylle et de ISST et de SST chaude. On la trouve surtout le plateau de la petite côte et sur le banc d'Arguin. Elle présente

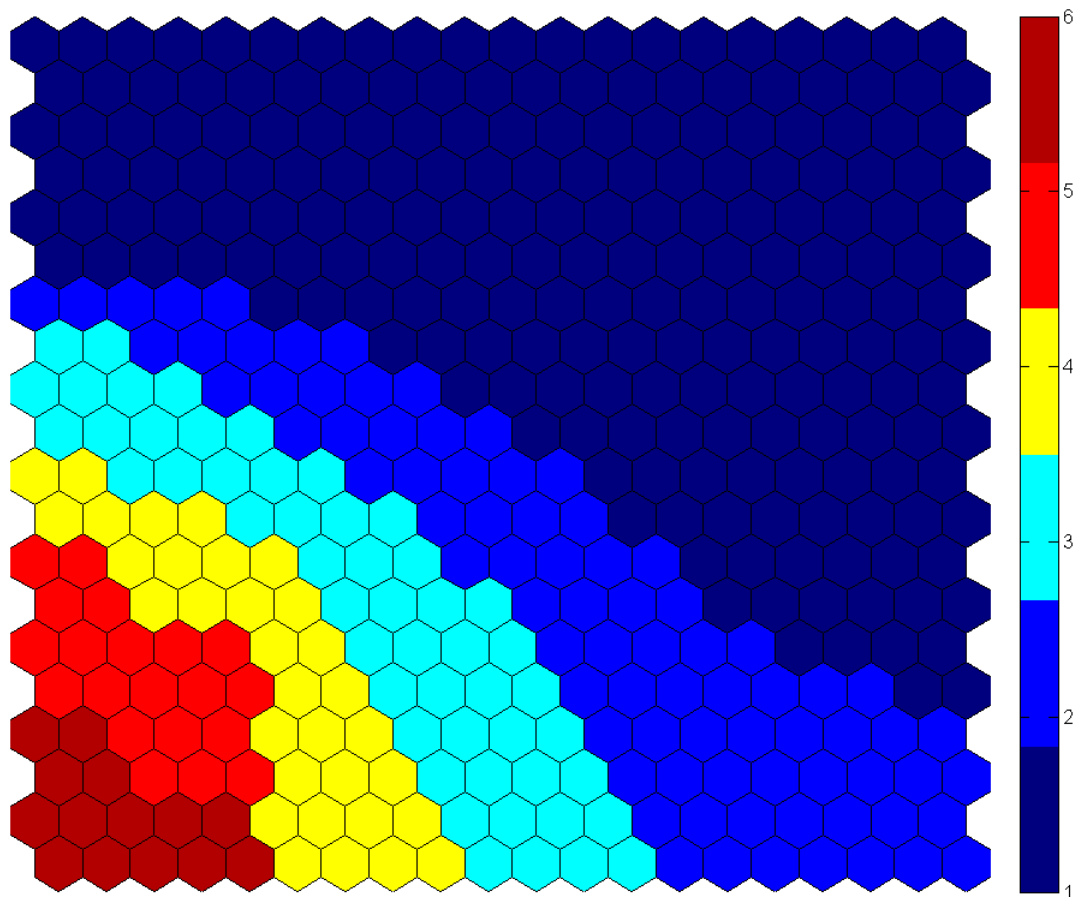


Figure 4.6 – Résultat pour les six classes de SOM2 formés à partir de la carte auto-organisatrice de Kohonen (résultats bruts de l’analyse).

une grande variabilité saisonnière

- La classe 6 correspond à des eaux qui ont des très hautes concentrations en Chl-a et des ISST et SST très froides. Elle forme une bande étroite située le long de la côte et présente une grande variabilité saisonnière. Elle caractérise les eaux de l’upwelling côtier.

Les caractéristiques en ISST et en concentration de Chl-a pour les différentes classes sont données sous forme graphique par les “**box plots**” (Figure 4.10) qui représentent la médiane des valeurs (trait plein dans la boîte) ; le premier quartile (25% des valeurs) et le dernier quartile (75% des valeurs) sont donnés par les bords de la boites. Les moustaches représentent 1.5 de l’espace inter quartile. Les classes sont toujours bien individualisées en Chl-a, quelque soit la saison (Figure 4.10) avec une augmentation de la concentration en Chl-a de la classe 1 à la classe 6 lorsque l’on va du large vers la côte. Le comportement en température est plus compliqué. On observe bien en Janvier une décroissance en ISST de la classe 1 à la classe 6 avec un comportement atypique de la classe 5 qui présente des ISST beaucoup plus chaudes que celles de la classe 4 et la classe 6.

Tableau 4.2 – Statistiques descriptives des caractéristiques (Chl-a) de chacun des six classes.

JANVIER		Maximum	Manimum	Mèdiane	Moyenne	Ecart-type
Chl-a	classe 1	0,193	0,725	0,335	0,361	0,102
	classe 2	0,604	1,253	0,865	0,872	0,155
	classe 3	1,017	2,165	1,439	1,491	0,247
	classe 4	1,690	4,315	2,484	2,629	0,590
	classe 5	3,086	6,993	4,628	4,733	0,783
	classe 6	3,569	16,050	6,756	7,027	2,221

Tableau 4.3 – Statistiques descriptives des caractéristiques (ISST) de chacun des six classes.

JANVIER		Maximum	Manimum	Mèdiane	Moyenne	Ecart-type
ISST	classe 1	-1,185	1,547	0,015	0,036	0,332
	classe 2	-1,461	1,326	-0,784	-0,670	0,488
	classe 3	-3,227	1,794	-1,439	-1,177	0,925
	classe 4	-3,860	1,804	-2,282	-1,915	1,283
	classe 5	-2,214	1,988	-0,148	-0,184	1,060
	classe 6	-4,705	1,939	-3,147	-2,948	0,969

Tableau 4.4 – Statistiques descriptives des caractéristiques (Chl-a) de chacun des six classes.

AOUT		Maximum	Manimum	Mèdiane	Moyenne	Ecart-type
Chl-a	classe 1	0,113	0,572	0,211	0,226	0,065
	classe 2	0,216	1,15	0,439	0,501	0,184
	classe 3	0,755	2,086	1,245	1,302	0,299
	classe 4	1,552	4,459	2,548	2,642	0,608
	classe 5	3,094	7,180	4,557	4,778	0,969
	classe 6	4,378	16,962	8,160	8,219	1,879

Tableau 4.5 – Statistiques descriptives des caractéristiques (ISST) de chacun des six classes.

AOUT		Maximum	Manimum	Mèdiane	Moyenne	Ecart-type
ISST	classe 1	-0,904	2,562	0,115	0,264	0,487
	classe 2	-1,093	2,415	-0,036	0,118	0,596
	classe 3	-0,864	2,505	0,010	0,195	0,748
	classe 4	-1,339	2,518	0,177	0,274	0,819
	classe 5	-0,266	3,146	0,762	0,833	0,639
	classe 6	-2,759	2,823	-0,115	0,434	1,253

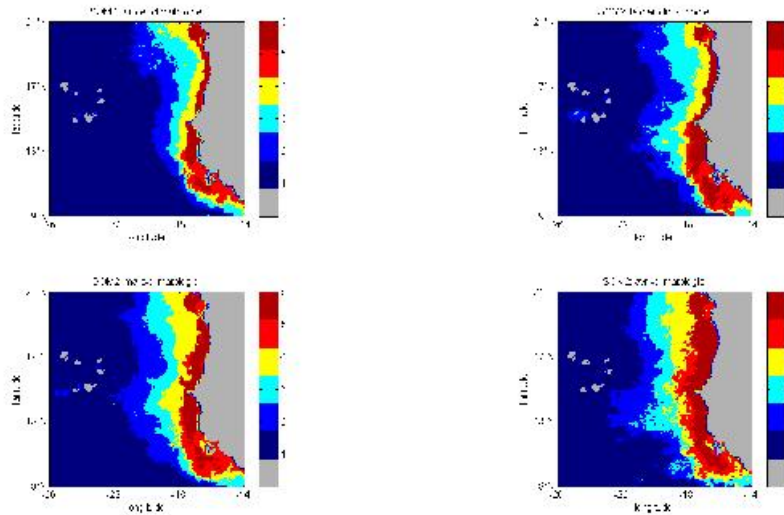


Figure 4.7 – projection sur la carte géographique (climatologie :Janvier-Avril)

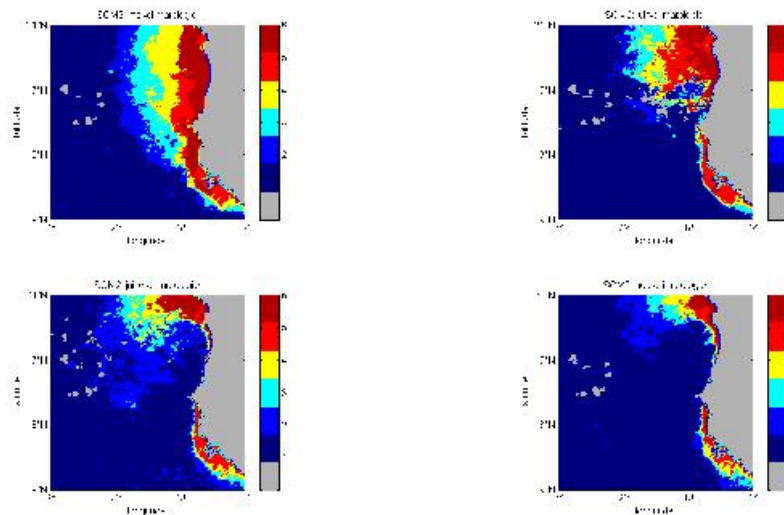


Figure 4.8 – projection sur la carte géographique (climatologie :Mai-Aout)

4.4 Variabilité saisonnière de l'upwelling

Les figures 4.11 et 4.12 montrent une variabilité saisonnière bien marquée du comportement des traceurs océaniques associés à l'upwelling. Cette variabilité est visible sur les SST et ISST médianes des différentes classes (Figure 4.11), beaucoup moins sur la concentration en Chl-a (Figure 4.12) qui reste quasi constante. Cette variabilité saisonnière bien marquée des SST et ISST des différentes classes pose la question de la pertinence de la décomposition en classes. En effet les caractéristiques hydrographiques des différentes classes présentent une certaine variabilité en ce qui concerne l'ISST et aussi de la SST. La question est alors de comprendre la signification d'une classe en terme d'hydrologie : la classe 1 est associée à des eaux dont l'ISST (Figure 4.11) est voisine de zéro et dont la teneur en Chlorophylle est très basse (Figure 4.12). La classe 1 regroupe les eaux du large, ceci quelque soit la saison. Les eaux de la classe 1 sont éloignées de la côte pendant la saison d'upwelling (Décembre à Avril) et viennent la rejoindre pendant les mois d'été (Juillet, Aout, Septembre). La valeur médiane de la SST de la classe

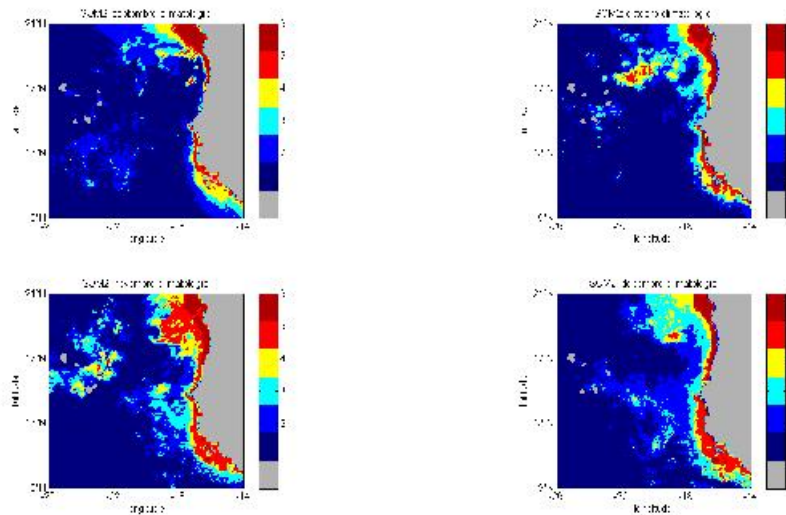


Figure 4.9 – projection sur la carte géographique (climatologie : Septembre-Décembre)

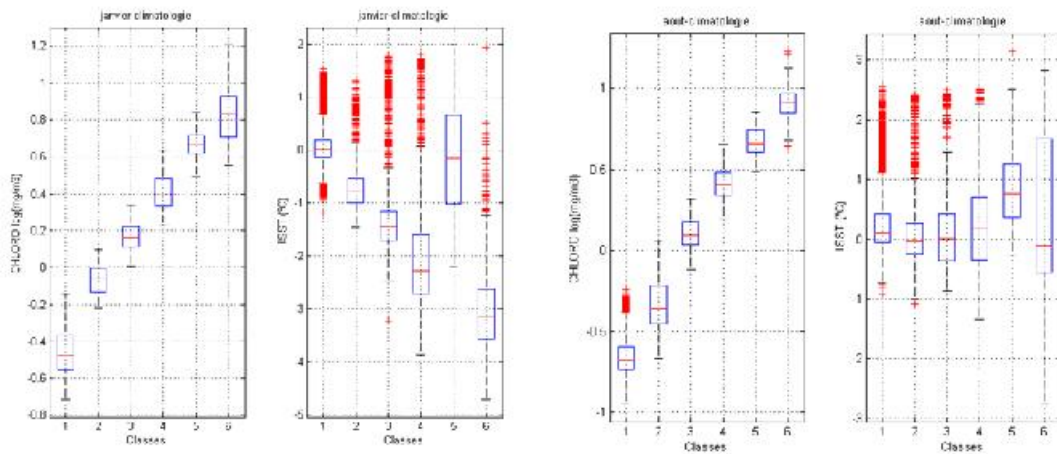


Figure 4.10 – Box-plot représentant les statistiques descriptives des caractéristiques géophysiques de chacun de classes six. Les valeurs sont réparties de manière logarithmique sur l'axe des ordonnées pour les concentrations en chlorophylle-a.

1 (Figure 4.6) est minimum en Avril et maximum en Octobre. Pour expliquer cela on peut incriminer l'ensoleillement qui est minimum en Hiver et maximum en été, avec un décalage temporel dû à l'inertie thermique. Cet effet est renforcé par le vent qui est maximum en hiver et donc brasse les eaux de surface jusqu'à la profondeur de la thermocline saisonnière contribuant ainsi au refroidissement des eaux de surface. L'effet inverse se passe à l'été où le vent est faible favorisant la stratification et donc l'apparition d'eaux chaudes en surface. La classe 6 est associée à des eaux dont l'ISST médiane (Figure 4.11) est la plus négative des 6 classes en hiver (-3°C en Janvier) et voisine de zéro en été. La SST médiane est aussi la plus basse en hiver et chaude en été. La Chl-a (Figure 4.12) est la plus élevée des 6 classes et ne présente pas de variations significatives liées à la saison. La classe 6 (noir sur la (Figure 4.9) est confinée à la côte où elle forme une bande étroite (quelques dizaines de kilomètres) épousant le trait de côte. La classe 6 disparaît quasiment au Sud de 20°N en Juillet, Aout, Septembre. Il semble clair que la classe 6 représente les zones côtières où l'effet de l'upwelling est le plus intense. La très forte modulation

Tableau 4.6 – Classification générale en six classes (caractérisation des variables géophysiques)

	Chl-a	ISST
classe 1	Très faibles	Très fort
classe 2	faibles	fort
classe 3	Moyenne	Moyenne
classe 4	Elévé	Faible
classe 5	Elévé/Très Elévé	Faible/Très Faible
classe 6	Très Elévé	Très Faible

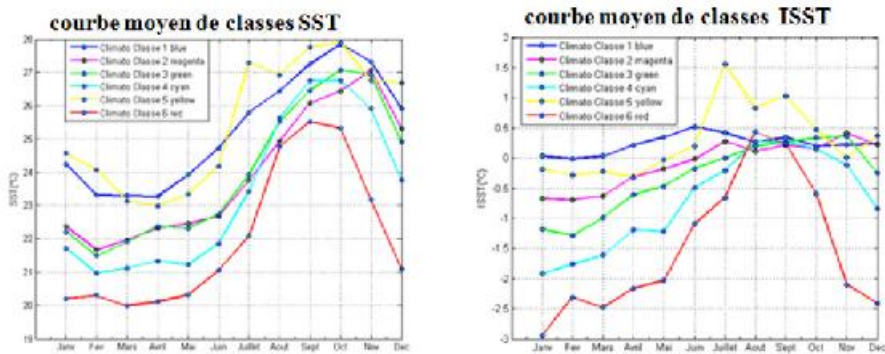


Figure 4.11 – Variation saisonnière des SST et ISST médianes des différentes classes

saisonniers en température pourrait être due à l'ensoleillement d'été associé à l'atténuation des vents.

La classe 5 (rouge foncé sur la Figure 4.9) montre un comportement atypique. Des eaux riches en Chl-a et des ISST (et des SST) élevées par rapport aux classes adjacentes (6 et 4 et 3) comme cela se voit sur la Figure 4.10. Il semble que la classe 5 caractérise, au moins durant les mois d'hiver (Décembre, Janvier, Février, Mars, Avril), les eaux situées sur des plateaux continentaux (plateau continental de la petite côte et banc d'Arguin). Ceci expliquerait les valeurs élevées de la Chl-a (eaux du cas-2) et la température (petits fonds). En Juin et Novembre on remarque des structures de classe 5 entre 18° - 20°N s'étendant vers le large. Ces patterns ainsi que la bande de classe 5 trouvée en mai entre 17 - 19°N posent des problèmes d'interprétation. Une première explication pourrait provenir de la construction de la base de données qui moyenne des informations provenant d'années différentes, ce qui peut biaiser la continuité des patterns spatiaux. Une solution serait d'analyser ces types de structures sur chacune des années de la base de données. Les classes 2, 3, 4 ont des ISST qui diminuent et des concentrations en Chl-a qui augmentent lorsque l'on se déplace de l'ouest (large) vers l'est. Leur structure spatiale suit approximativement celle de la classe 6 laissant supposer que les zones correspondant à ces classes sont influencées par l'upwelling via des structures de méso échelle (filaments, tourbillons, visibles sur Figure 4.10. En particulier, leur signature disparaît au Sud de 20°N en Juillet, Aout, Septembre. Nous avons aussi calculé la variabilité saisonnière l'extension spatiale (nombre de pixels) des classes 5 et 6 (Figure 4.13). On remarque que ces deux classes présentent une forte variabilité saisonnière avec un maximum en mars, avril, mai, un second maximum en octobre novembre et un minimum pendant la saison des pluies (Juillet, Aout). Les variations saisonnières de l'upwelling sont en accord avec celles du vent. On a tracé sur la Figure 4.14, la composante N - S du vent de 1990 à 2007 à Dakar. On remarque la

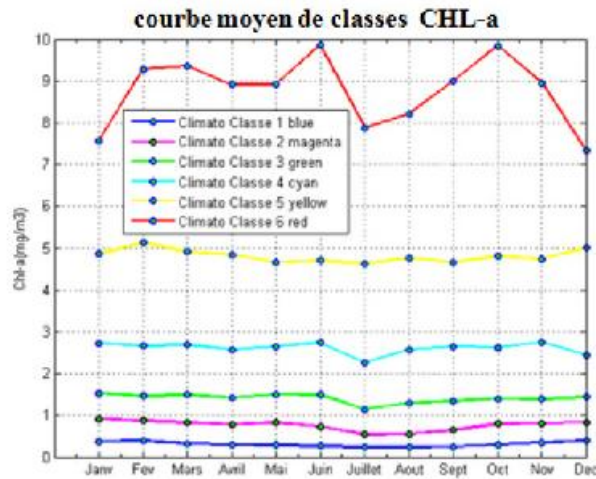


Figure 4.12 – Variation saisonnière des concentrations médianes en Chl-a des différentes classes

forte variabilité saisonnière du vent avec un maximum au mois de mars lorsque l'extension de la classe 6 est maximale et son ISST et SST minimales. La disparition quasi totale de la classe 6 en dessous de $20^{\circ}N$ correspond à l'annulation du vent parallèle à la côte en Juillet-Aout. Cette valeur minimale du vent est associée avec le déplacement vers le Nord de l'ITCZ qui atteint sa latitude maximale au mois d'aout. On ne remarque pas de variabilité interannuelle caractéristique du vent, à part une diminution sensible du pic hivernal en 1997. La variabilité saisonnière de la classe 6 peut s'expliquer selon le scénario suivant : les forts vents d'hiver génèrent un upwelling intense le long de la côte de 21° à $15^{\circ}N$ associé au développement de la classe 6. L'affaiblissement du vent en été (Juin, Juillet, Aout) conduit à la disparition de l'upwelling et donc de la classe 6 de $15^{\circ}N$ jusque $20^{\circ}N$. La classe 6 est alors remplacée par la classe 1 qui s'étend maintenant du large jusqu'à la côte. Ce schéma est-il dû à une dynamique horizontale comme le suggère (Lathuilière et al., 2008), les eaux associées au Contre Courant Equatorial remontant vers le Nord, ou plutôt comme nous le pensons, à une simple dynamique verticale, la disparition de la composante du vent parallèle à la côte et du transport d'Ekman vers l'ouest qui lui est associé favorisant la progression vers l'Est des eaux de la classe 1 très chaudes en cette saison ? Un élément de réponse peut être trouvé en utilisant 3 variables pour faire la classification.

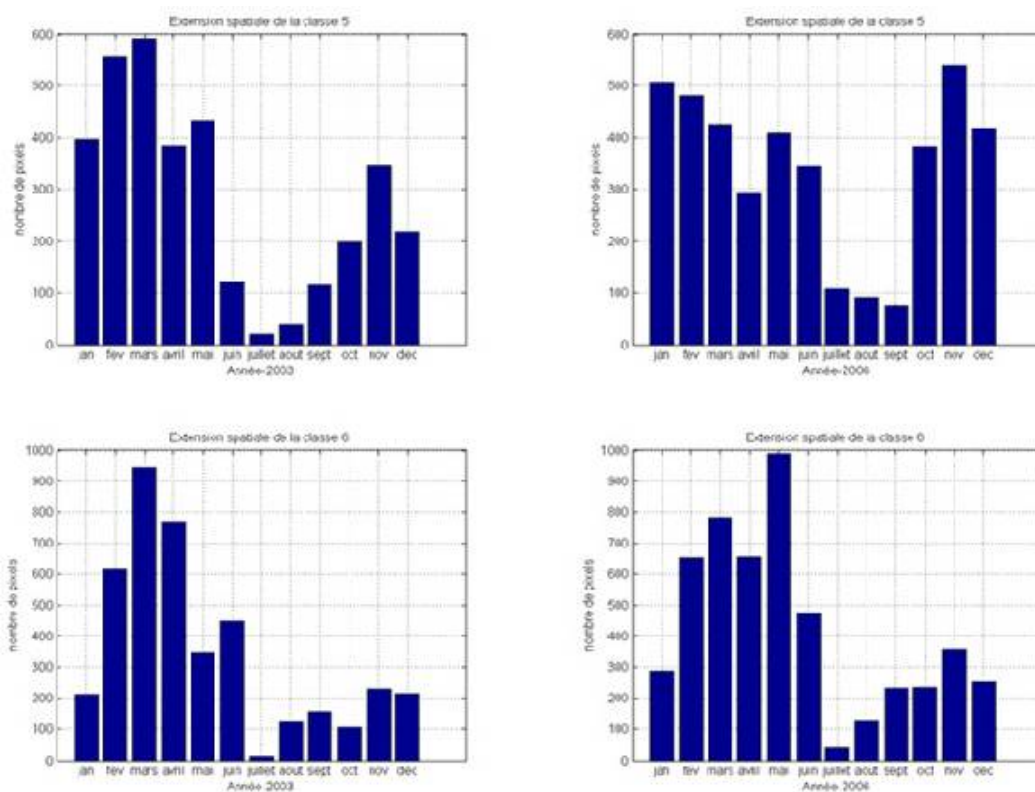


Figure 4.13 – Extension spatiale (nombre de pixels) des classes 5 et 6

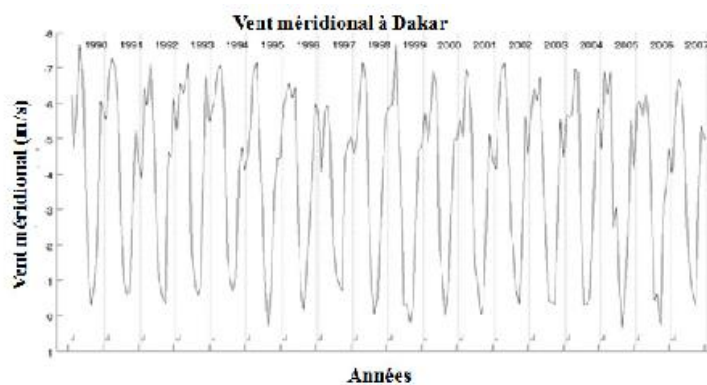


Figure 4.14 – Composante N-S du vent à Dakar. On remarque la forte variabilité saisonnière avec un maximum au mois de mars.

4.5 Classification sur 3 variables : SOM3 (ISST, SST, Chl-a)

Nous avons donc fait une classification de la zone en utilisant maintenant 3 variables qui sont ISST, SST et la concentration en Chl-a. L'ajout de la SST devrait permettre de mieux tenir compte de l'effet saisonnier qui est en grande partie filtré par l'ISST. Comme précédemment, on utilise l'algorithme SOM suivi d'une CAH. On obtient alors une carte topologique de 20x20 divisée en 6 classes qui est présentée sur la figure 4.15 similaire à la carte SOM2 apprise avec 2 variables. Les classes 1, 4, 5,

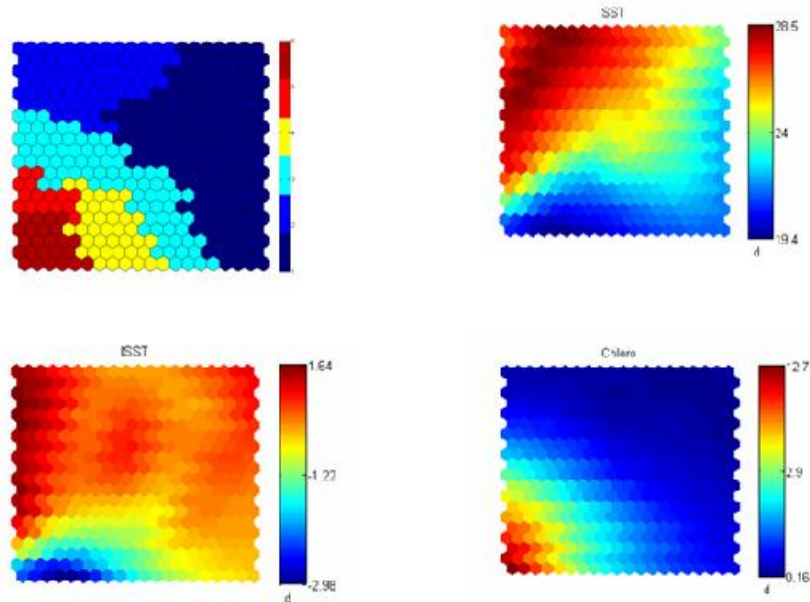


Figure 4.15 – Carte topologique SOM3 apprise avec les 3 variables et décomposée en 6 classes (en haut à droite) et projection des 3 variables (SST, ISST et Chloro-a) sur la carte. On remarque que la topologie de la nouvelle carte SOM dénotée SOM3 est similaire à la

6 sont positionnées de façon analogue sur SOM2 et SOM3. L'inspection des **“box plots”** montrent qu'elles représentent des eaux de même caractéristique. En particulier la classe 6 (en rouge foncé sur la Figure 4.17) présente la même variabilité que dans SOM2 avec une remontée vers $20^{\circ}N$ en Juillet, Aout, Septembre. La classe 5 est maintenant confinée sur le plateau continental de la petite côte sauf en novembre où une grande tache de classe 5 de SOM3 apparaît en novembre au large de $19^{\circ}N$. Les classes 2 et 3 de SOM2 qui sont des classes soumises à l'influence de l'upwelling grâce à des structures de méso-échelle qui engendrent une diffusion horizontale des eaux de l'upwelling se fondent maintenant dans la classe 3 de SOM3. La classe 2 de SOM3 qui n'a pas d'équivalent dans SOM2, provient d'une partition de la classe 1 de SOM2. La classe 2 de SOM3 est associée à des eaux chaudes pauvres en Chl-a situées dans le sud de notre zone d'observation provenant probablement du contre-courant équatorial (Figure 4.17). Elles proviennent d'une partition des eaux de la classe 1 de SOM2 en deux classes. Durant les mois de Juin, Juillet, Aout, Septembre, la classe 2 de SOM3 envahit les côtes Sénégalaises de $13^{\circ}N$ à $17^{\circ}N$ remplaçant ainsi les eaux des classes 6, 4 et 3, ce qui est en accord avec le schéma esquissé par (Lathuiliere et al., 2008). Notons que les eaux de la classe 2 n'envahissent jamais le plateau continental de la petite côte dont les eaux restent toujours bien individualisées comme cela se passe avec les eaux de la classe 1 de SOM2. Les structures des classes sont analogues à celles de SOM3. Le tracé des boxplots

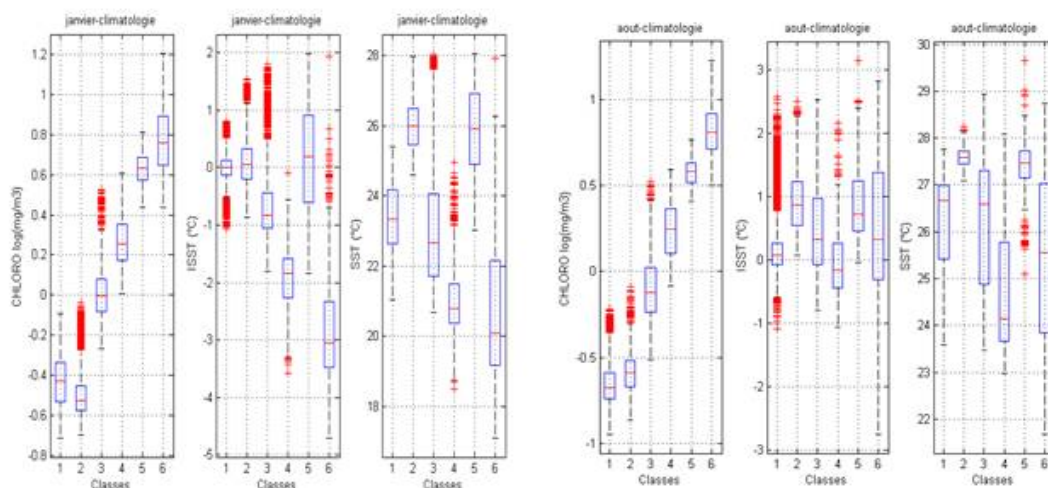
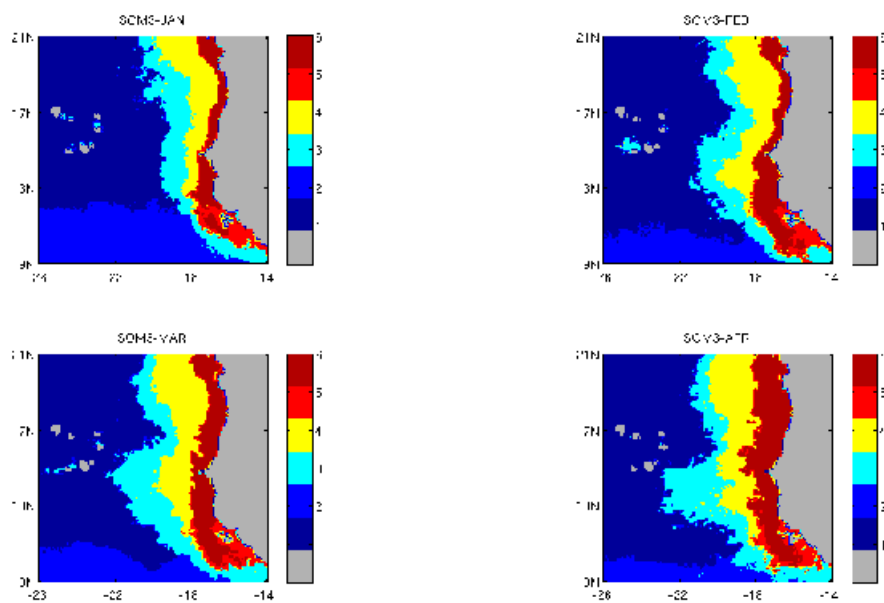


Figure 4.16 – Boxplots pour Janvier et Aout de SOM3. Les structures des classes sont analogues à celles de SOM2

(Figure 4.16) montre toujours le comportement atypique de la classe 5 inhérente à des eaux situées sur le plateau continental de la petite côte.



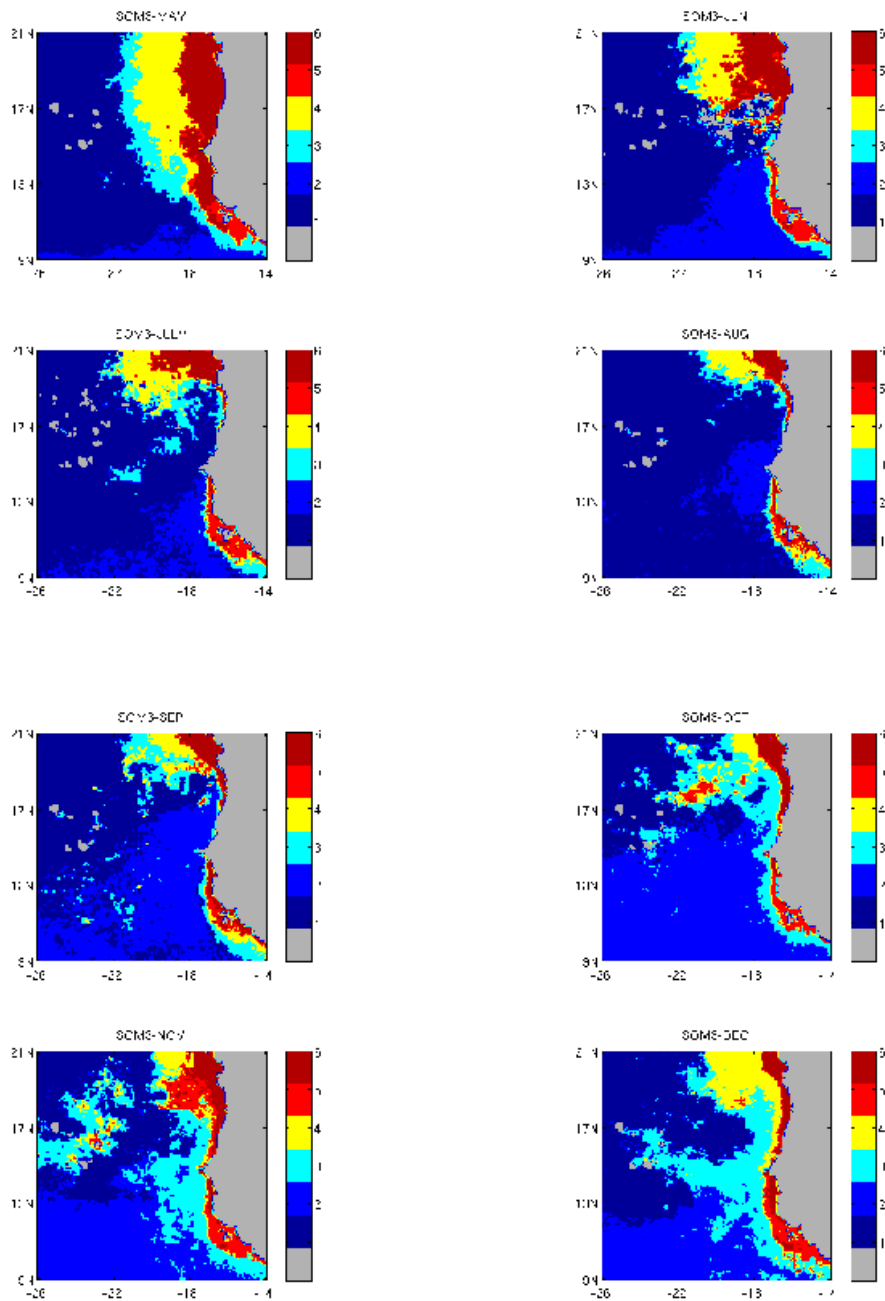


Figure 4.17 – Répartition mensuelle des six classes de SOM3. On observe la quasi disparition de la classe 6 qui est celle de l’upwelling en Juillet Aout Septembre pendant la saison des pluies

4.5.1 Variabilité interannuelle de l’upwelling

Si la variabilité saisonnière est bien marquée et corrélée à celle du vent, la variabilité interannuelle est beaucoup plus difficile à analyser.

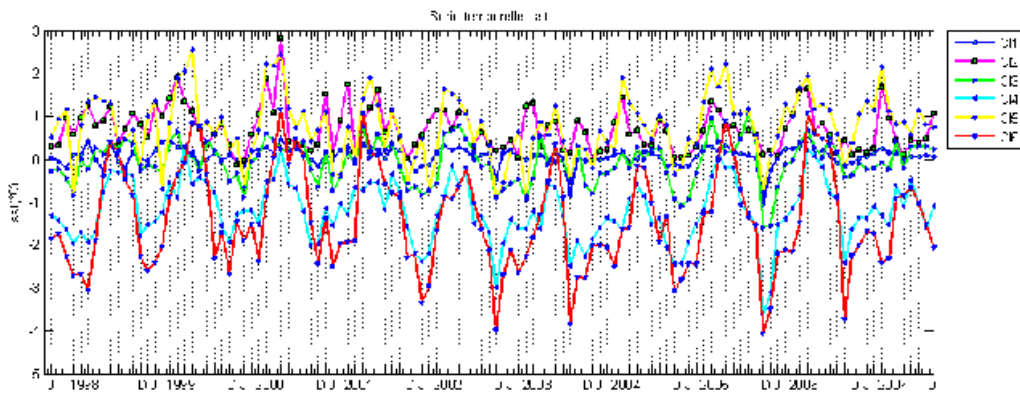


Figure 4.18 – séries temporelles des ISST

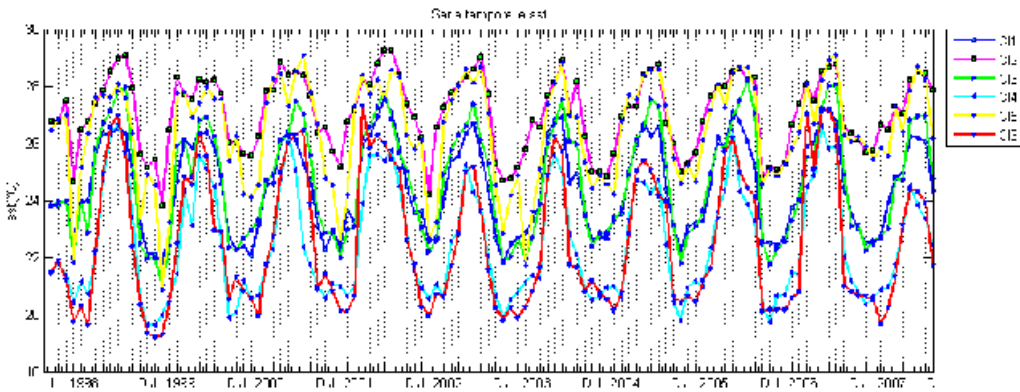


Figure 4.19 – séries temporelles des SST

Nous avons reporté sur les figures 4.18, 4.19 et 4.20 les séries temporelles des ISST, SST et des concentrations en Chl-a pour chaque classe et pour les dix ans des observations. Aucun signal n'apparaît clairement. Nous avons regardé aussi une éventuelle influence de l'ENSO (l'EL-niño Southern Oscillation (oscillation australe)) et de la NAO (Nord Atlantique Oscillation) en particulier un signal éventuel dû à l'EL-niño exceptionnel de 1998. Nous n'avons pas été en mesure de détecter une influence significative de ce phénomène. On remarque que les série temporelle des ISST de la classe 6 de SOM2 et SOM3 sont fortement corrélées ($R=0.991$) ainsi que les séries temporelles de la classe 6 de la Chl-a données par SOM2 et SOM3 ($R=0.879$) montrant la bonne similitude pour les classes d'upwelling des deux décompositions données par SOM2 et SOM3. Les corrélations entre ISST et Chl-a sont très faibles ($R=0.11$) montrant qu'il n'existe pas de liaison simple entre la dynamique (ISST) et la productivité (Chl-a) et que par conséquent la dynamique de l'upwelling est complexe.

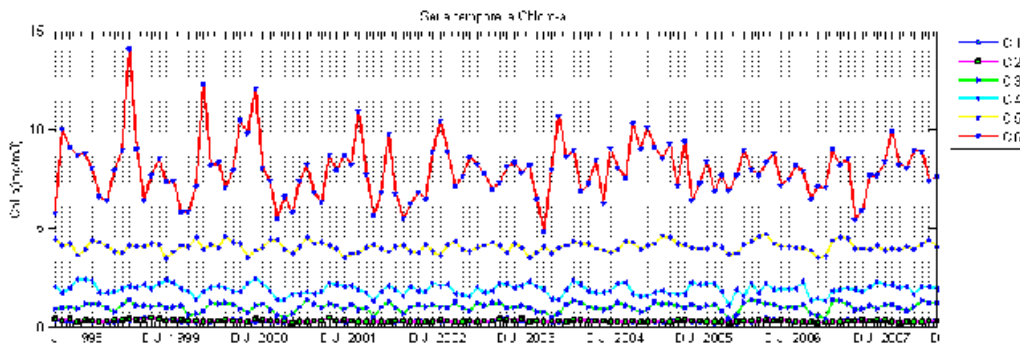


Figure 4.20 – séries temporelles des Chl-a

4.5.2 Discussions

La classification multi-variables par carte topologique SOM des eaux de surface de la région de l'Upwelling Sénégal-Mauritanien entre $9^{\circ}N$ et $21^{\circ}N$, nous a permis de faire une description de l'upwelling en terme de variabilité spatio-temporelle. En utilisant 2 variables (SOM2) ou 3 variables (SOM3) accessibles par mesures satellitaires (ISST, la SST et la concentration en Chl-a), nous avons décomposé les situations hydrologiques de surface en 6 classes. Ces classes forment des ensembles statistiquement cohérents dans le sens où elles représentent des neurones contigus sur les cartes SOM (donc ayant des propriétés statistiques communes) et où elles sont aussi géographiquement continues formant des structures spatiales bien homogènes. On peut les considérer comme des indices multi-variés permettant de caractériser les différentes masses d'eaux. La classe 6 de SOM2 (et de SOM3) est associée aux eaux froides et riches en phytoplancton de l'upwelling côtier. Ces eaux longent la côte en formant une bande de quelques dizaines de kilomètres de large. Elles sont bien définies en hiver (Décembre-Avril) où on les retrouve jusque vers $13^{\circ}N$. Pendant les mois de Juillet, Aout, Septembre, ces eaux disparaissent au sud de $20^{\circ}N$ et sont remplacées par des eaux du large qui en utilisant les résultats de SOM3 laissent à penser qu'elles proviennent du Sud et pourraient être dues à une remontée du Contre-Courant Equatorial provoquée par la remontée vers le Nord de l'ITCZ et la quasi-disparition des alizées. Les classes 4, 3, 2 ont un comportement similaire à celui de la classe 6, en accord avec le fait que les paramètres océaniques de ces classes seraient liés à une diffusion vers le large des paramètres de la classe 6. Ce schéma est conforté par l'analyse des signaux altimétriques. En Janvier, Février, Mars on observe un 'trou' de SLA (Sea Level Anomaly) au niveau de la zone d'upwelling (Figure 4.21) associé à la classe 6 laissant supposé un courant de surface dirigé vers le sud forcé par les alizées en accord avec la théorie (Allen, 1973). En Juillet, Aout, Septembre, on observe une disparition de ce 'trou' de SLA en accord avec la disparition de l'upwelling, tandis que l'on observe un 'trou' de SLA autour de $10^{\circ}N$, $24^{\circ}W$ laissant supposé une circulation cyclonique qui serait associée au Dôme de Guinée et expliquerait la remontée vers le Nord des eaux de la classe 2 venant remplacer la disparition de l'upwelling (Figure 4.21).

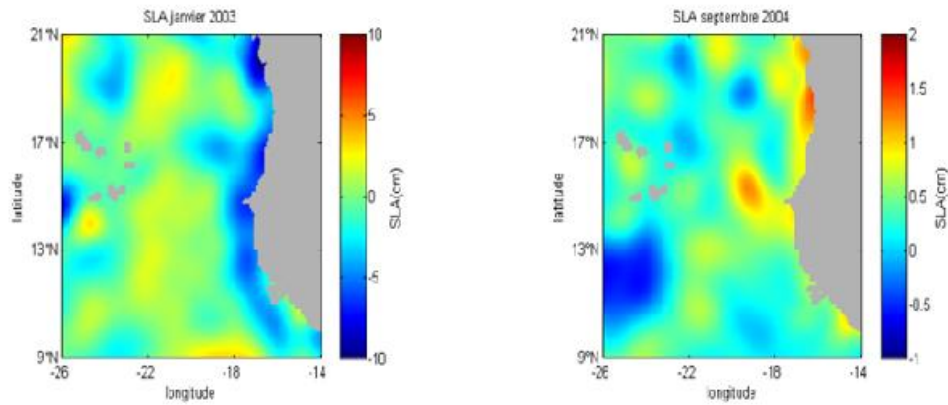


Figure 4.21 – SLA en Janvier 2003 et Septembre 2004

4.6 Conclusion

Cette étude a permis de caractériser l'upwelling côtier Sénégal-Mauritanien et de décrire sa variabilité saisonnière et interannuelle à partir des observations satellitaires de SST et de couleur de l'océan. On peut dire aussi que la décomposition en classe nous a permis de mieux comprendre la dynamique de l'upwelling Sénégal-Mauritanien, en particulier sa variabilité spatiale en fonction des saisons. On a mis en évidence différentes zones aux caractéristiques bien définies dont l'étude permet de mieux saisir le comportement de cet upwelling. La zone étudiée présente une très forte variabilité saisonnière qui est due à la fois au vent et à l'ensoleillement. Le vent est responsable de la variabilité saisonnière de l'upwelling, en particulier de la position des classes 6, 4, 3 et 2. La disparition de ces classes au sud de $20^{\circ}N$ est due à l'arrivée de l'ITCZ et de la disparition des alizées. Il semble d'après cette étude qu'il pourrait exister des eaux du cas-2 associées à la classe 5 principalement et situées sur le plateau continental de la petite côte. Ceci pourrait être analysé en considérant les spectres de réflectance marine ρ_w associés à chaque pixel. La classe 5 est définie tant par SOM2 que par SOM3 par des eaux qui sont chaudes et riches en chlorophylle. Elle est caractéristique des eaux situées sur le plateau continental de la petite côte où on les trouve en permanence. Elles apparaissent parfois au large (quelques dizaines de kilomètres des côtes) entre $18^{\circ} - 20^{\circ}N$. Ces structures ainsi que la bande de classe 5 trouvée en mai entre $17 - 19^{\circ}N$ posent des problèmes d'interprétation. Une première explication pourrait provenir de la construction de la base de données qui moyenne des informations provenant d'années différentes, ce qui peut biaiser la continuité des patterns spatiaux. Une solution serait d'analyser ces types de structures sur chacune des années de la base de données. Des sections de SST et Chl-a à $17^{\circ}N$ en mai 1999 et en mai 2001 où ces situations sont bien contrastées devrait permettre de préciser l'origine de ces structures. Une étude intéressante serait aussi d'analyser les spectres des ρ_w (réflectances océaniques) de ces eaux. Selon notre interprétation les spectres des ρ_w de la classe 5 devraient être différents sur le plateau continental de la petite côte où ils devraient être liés à des eaux du cas-2 et au large entre $16^{\circ}N$ et $19^{\circ}N$ où ils devraient refléter celui des eaux du cas-1.

Chapitre 5

Calcul d'une reflectance océanique normalisée (nLw) ne dépendant que de la chl-a à l'aide d'un MLP

5.1 Introduction

Le chapitre précédent a permis de caractériser toutes les eaux associées à la classe (1, 2, 3, 4, et 6), d'après les analyses de SOM2 puis SOM3, cependant les eaux de la classe 5 présentent une particularité (classe atypique) car ces eaux sont à la fois riches en chlorophylles et chaudes. Dans ces eaux, situées sur le plateau continental de la petite côte, il pourrait exister les eaux du cas-2.

Cette partie de la thèse aborde une autre méthode neuronale en occurrence les perceptrons multicouches (Multi Layer Perceptron - dénotée MLP). L'application du MLP combinée à un modèle simple de normalisation va nous permettre d'analyse les eaux de la classe 5 et 6 afin de les caractériser.

5.2 Données couleur de l'océan : nLw (normalized water leaving radiance)

L'analyse faite dans le chapitre précédent a montré que, les six classes que nous avons identifiées correspondent à des eaux ayant des caractéristiques différentes (SST, l'ISST et Chl-a). Par ailleurs, les capteurs à bord de satellite couleur de l'océan fournissent des informations supplémentaires telles que des mesures spectrales de radiances nLw (normalized water leaving radiance), à différentes longueurs d'onde. Ces mesures ont été utilisées pour estimer la concentration en chlorophylle-a (O'Reilly et al., 1998). Les données nLw (normalized water leaving radiance) contiennent des informations supplémentaires sur la composition biogéochimique de l'eau (Alvain et al., 2005). Nous proposons d'étudier ces informations supplémentaires en relation avec les six classes. Nous avons utilisé les données nLw à cinq longueurs d'onde (412 nm, 443 nm, 490 nm, 510 nm, 555 nm) fournies par la base de données SeaWiFS (NASA/GSFC/DAAC). Les données pour une période de 10 ans (1998-2007) ont été recueillies.

5.2.1 Statistique classique

Une étude statistique préalable a été effectuée afin de comprendre les pixels associés à chaque classe (Figure 5.1). En effet, ces pixels ont été extraits à partir des images de cartes géographiques de classes de 10 ans de la classification de SOM3. L'étude statistique nous a montré que les formes (allures) de spectres moyens de la classe 5 se distinguent nettement des autres (spectres moyens de la classe 1, 2, 3, 4 et 6) et ce, pour toutes les années étudiées. Cette analyse statistique de la forme de spectre de la classe 5 laisserait penser à l'effet dominant de la chlorophylle-a. Nous avons ainsi appliqué une procédure de normalisation proposée dans (Alvain et al., 2005) permettant d'atténuer l'effet de la chlorophylle-a. La normalisation

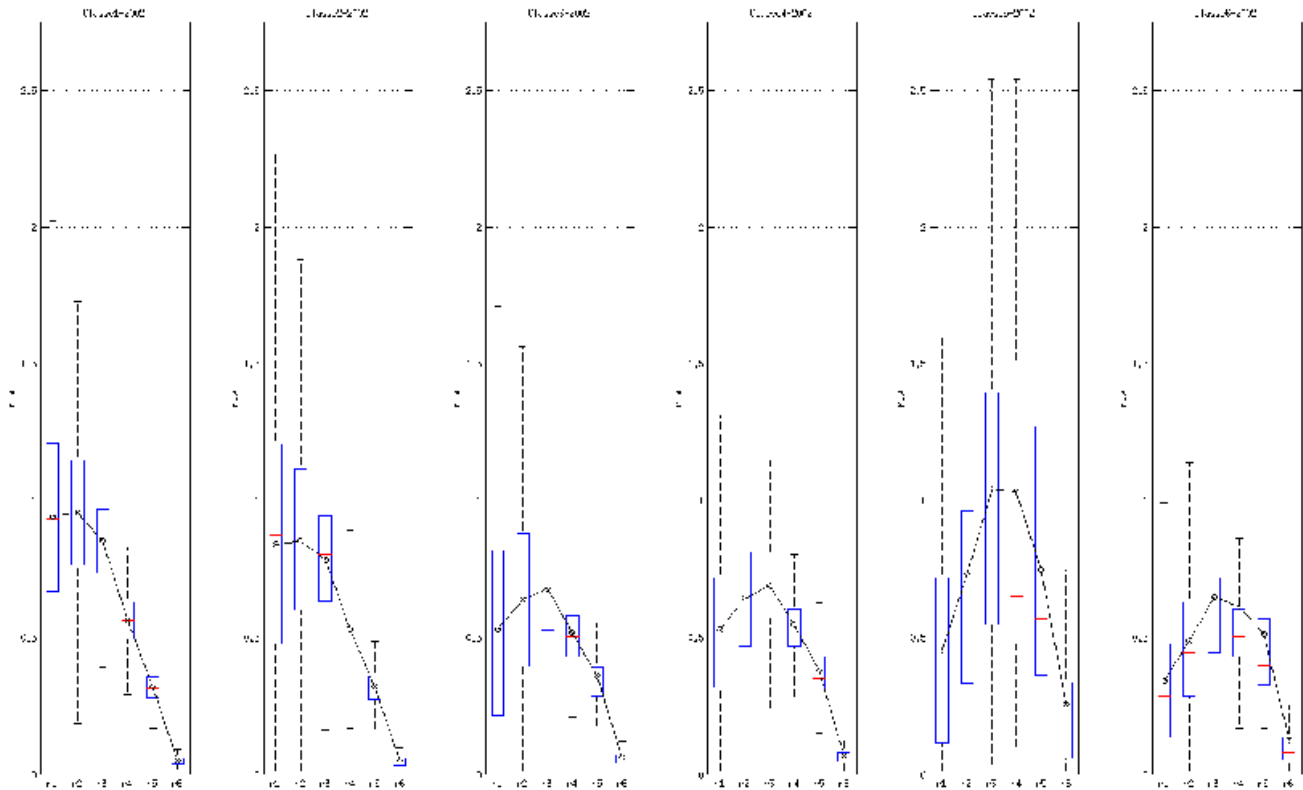


Figure 5.1 – Boxplots montrant la distribution des classes, année 2002.

consiste en une utilisation d'une table de référence (Alvain et al., 2005). Dans cette base de données nLw (normalized water leaving radiance), on a un modèle simple qui lie les données nLw à la concentration en chlorophylle-a. Les données de nLw de cette base ont été corrigées de l'effet atmosphérique (Gordon and Wang, 1994). Afin d'extraire les effets de second ordre en raison des caractéristiques spécifiques de l'abondance du phytoplancton, nous avons calculé un spectre spécifique dénotée nLw^* définie par :

$$nLw^*(\lambda) = \frac{nLw(\lambda)}{nLw^{ref}(\lambda, Chl-a)} \tag{5.1}$$

où $nLw^{ref}(\lambda, Chl-a)$ est une relation liant les valeurs de concentrations en chlorophylle-a aux valeurs spectrales de nLw fournies par (Alvain et al., 2005). La détermination de $nLw^*(\lambda)$ implique l'utilisation du modèle de $nLw^{ref}(\lambda, Chl-a)$ qui donne la valeur de spectre pour chaque valeur de la concentration en chlorophylle-a. Or la table de référence fournit les valeurs spectrales pour des concentrations de

chlorophylle-a allant de $0,005$ jusqu'à $4mgm^{-3}$, ce qui est bien en deçà des concentrations qu'on peut observer dans la région étudiée, qui peuvent aller au-delà de $10mgm^{-3}$. Nous avons donc décidé de développer un modèle $nLw^{ref}(\lambda, Chl-a)$ adapté à la région étudiée. Pour cela nous avons utilisé une autre classe de réseaux de neurones, le perceptron multicouche [Evêque, 2006], décrit dans le chapitre 2 qui sont bien adaptés pour la modélisation de relation non linéaire (Thiria et al., 1993) et présentent l'avantage de fournir des fonctions continues au lieu d'utiliser la table de référence (Alvain et al., 2005) qui implique une interpolation linéaire entre deux valeurs discrètes pour obtenir la valeur correcte pour une valeur donnée de la concentration en chlorophylle-a qui n'a pas le spectre nLw correspondant dans cette table.

5.2.2 Développement du MLP

Pour développer le perceptron multicouche, nous avons utilisé les données climatologiques de la concentration en chlorophylle-a et les données de spectres nLw , aux six longueurs d'onde ($\lambda = 412$ à $670nm$).

5.2.3 Prétraitements de données

L'étude abordée dans cette section consiste à examiner les données afin de connaître leur distribution. La figure 5.2a montre l'histogramme des données brutes de la concentration en chlorophylle-a du mois

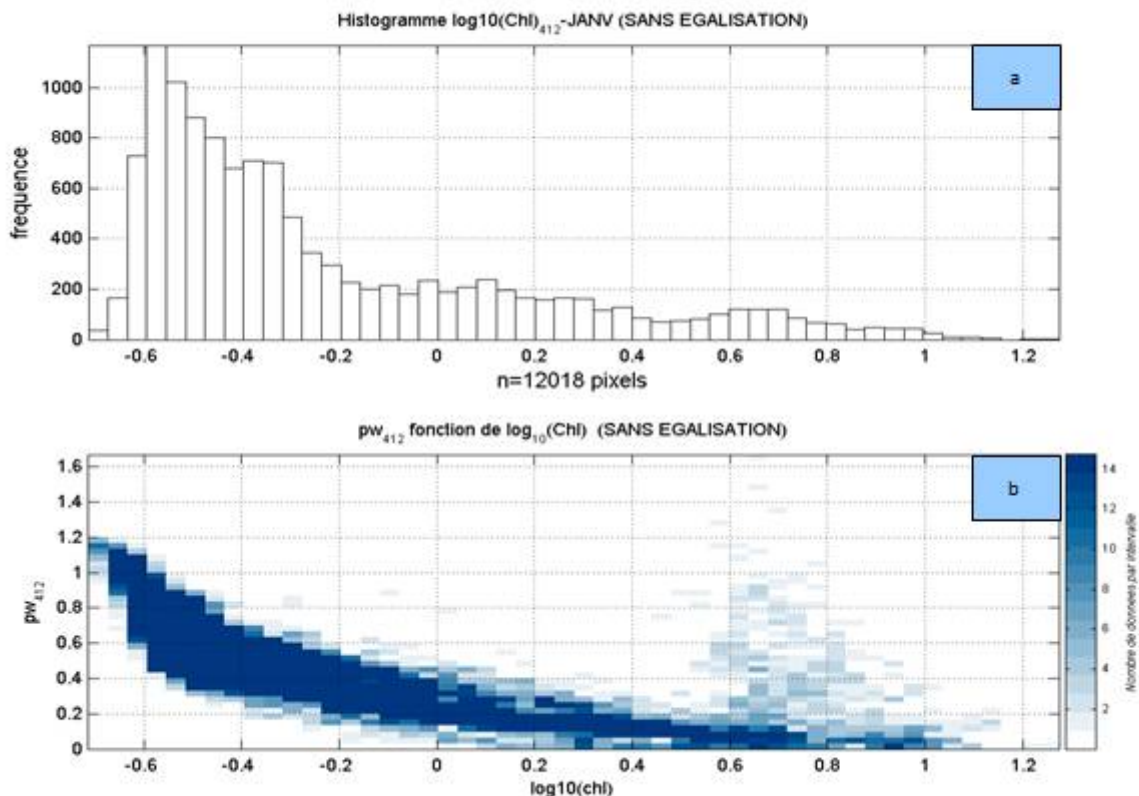


Figure 5.2 – Données brutes mois de janvier

janvier alors que la Figure 5.2b présente les données nLw en fonction de la concentration en chlorophylle-a en Log10. La Figure 5.2a montre que l'histogramme est plus tiré vers des valeurs négatives alors que les valeurs positives sont moins représentées, ceci peut influencer l'apprentissage du réseau car les données de la base d'apprentissage ne seront pas représentatives. Pour palier à cela, on procédera à une opération dite « **égalisation de l'histogramme** » qui consiste à avoir un histogramme "**plat**" afin d'avoir le même nombre de données par intervalle de valeurs. Après l'égalisation (Figure 5.3), on remarque que les données sont presque bien distribuées sur l'intervalle définie et ceci est important pour l'apprentissage du MLP. Malgré cette opération, on y voit sur la Figure 5.3b, une traînée de données par

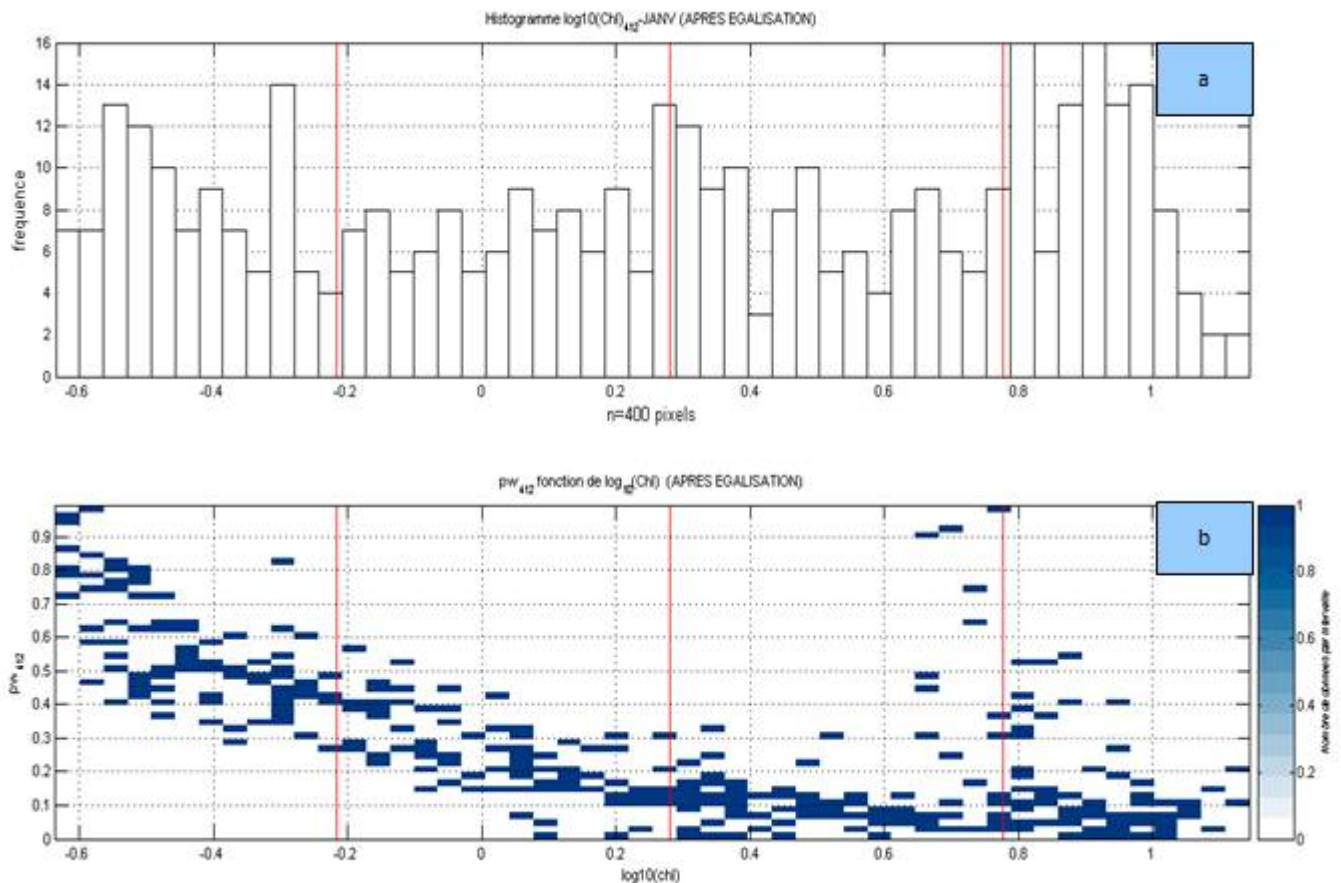


Figure 5.3 – Données brutes mois de janvier (APRES EGALISATION)

rapport à l'allure de la courbe ayant de fortes valeurs de nLw ($\lambda=412$), qualifiées des valeurs aberrantes qu'il faut supprimer par rapport aux restes de données.

Nous avons alors développé des algorithmes dont la procédure est décrite d'une manière synoptique ci-après, pour filtrer ces données aberrantes.

Algorithme : Filtre

Entree : V_{ch_moy} , V_{nLw_moy}

Sortie : Ch_moy , nLw_moy

Debut

1. Tracer $V_{nLw_moy}=f(V_{ch_moy})$
2. Identifier la zone des valeurs aberrantes
3. Prendre le vecteur V_{nLw_moy}
4. Récupérer les indices correspondants
5. Mettre les valeurs de V_{nLw_moy} à nan correspondant aux indices
6. Affecter les résultats à nLw_moy
7. Reprendre les étapes 5 et 6 avec le vecteur V_{ch_moy}
8. Affecter les résultats à Ch_moy

Fin

La Figure 5.4 présente le résultat du traitement.

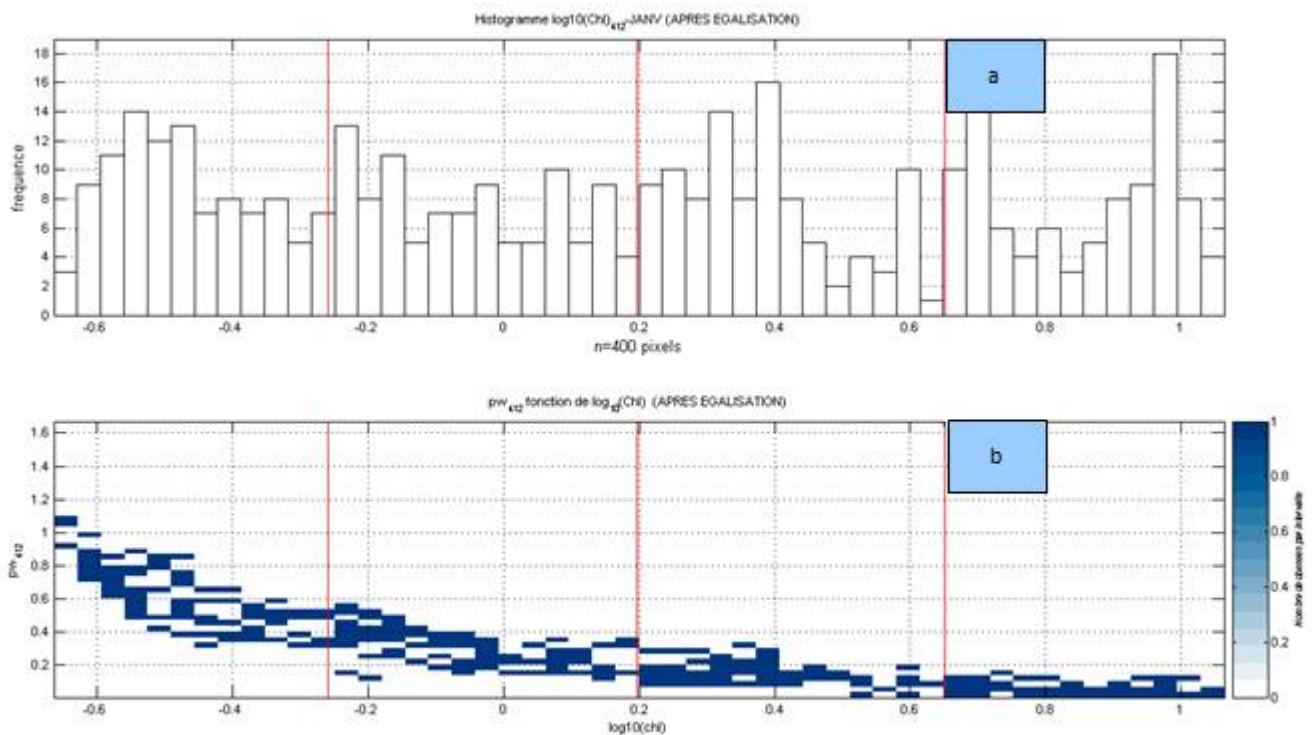


Figure 5.4 – Données filtrées mois de janvier (APRES EGALISATION)

Tableau 5.1 – Statistiques descriptives des caractéristiques (Chl-a) de chacune des six classes.

nLw($\lambda=412\text{nm}$)					
Neurones	Iterations	RMS_All	RMS_App	RMS_Val	RMS_test
3	100	0,1715	0,1714	0,1736	0,1697
3	200	0,1712	0,1712	0,1710	0,1715
3	400	0,1711	0,1708	0,1703	0,1730
3	600	0,1712	0,1713	0,1727	0,1689
3	800	0,1711	0,1713	0,1710	0,1704
3	1000	0,1712	0,1714	0,1708	0,1704

5.2.4 Base d'apprentissage MLP

Étant donné que nous avons six longueurs d'onde, nous avons formé six bases de données composées pour chacune une longueur d'onde plus la chlorophylle-a. Nous avons considéré au total 4800 données soit $100 \times 4 \times 12$ c'est-à-dire 100 données par intervalle et on a 4 intervalles définis sur les 12 mois de l'année climatologique. Ce choix se justifie par un tirage aléatoire de 100 données dans chaque intervalle dans l'optique d'avoir une base de données représentative. Ainsi, nous avons formé six bases et dans chaque base, on a la longueur d'onde λ plus la chlorophylle-a dénotée [base100_nLw(λ)_mois] ($\lambda = 412\text{nm}$ à 670nm). Après cette étape, nous allons faire un apprentissage MLP en utilisant l'outil **NFTOOL** de matlab. Les paramètres utilisés sont :

- matrice d'entrée : inputs(Xchl)
- matrice de sortie : outputs(XnLw)

Nous avons choisi un échantillonnage des données comme suit :

- Apprentissage : 70%
- Validation : 15%
- Test : 15%

5.2.5 Apprentissage

Conditions d'apprentissage : Nous avons utilisé des fonctions d'activation tangentes hyperboliques pour les neurones de la couche cachée, et linéaire pour la cellule de sortie. Un maximum de 100 à 1000 itérations ont été faits pour chaque apprentissage pour une bonne appréciation de l'expérience. Afin de définir une architecture optimale de MLP, nous avons essayé plusieurs architectures et nous avons effectué plusieurs apprentissages en faisant varier le nombre de neurones dans la couche cachée. L'apprentissage est effectuée avec les données nLw(λ) normalisées et les données de concentrations en chlorophylle-a normalisées en log 10. La normalisation a pour effet de ramener la majorité des valeurs de paramètres dans l'intervalle [-1, 1] qui est l'intervalle pour lequel la fonction d'activation du réseau est la plus variable. Les résultats de l'apprentissage MLP sont résumés dans le tableau 5.1 et les résultats pour les autres longueurs d'ondes sont données dans l'annexe F).

5.2.6 Détermination de l'architecture optimale

Le tableau 5.1 présente pour chaque apprentissage $nLw(\lambda)$ le nombre de neurones cachés, le nombre d'itérations, les RMS respectifs sur l'ensemble de données d'apprentissage, de validation et de test. Le tableau 5.1 permet, en examinant la RMS sur l'ensemble de validation de retenir le réseau optimal. Nous allons appliquer les réseaux optimaux retenus sur des données de test afin de voir le comportement du réseau (Voir Annexe F). La Figure 5.5 montre que le réseau se comporte bien en général et les corrélations

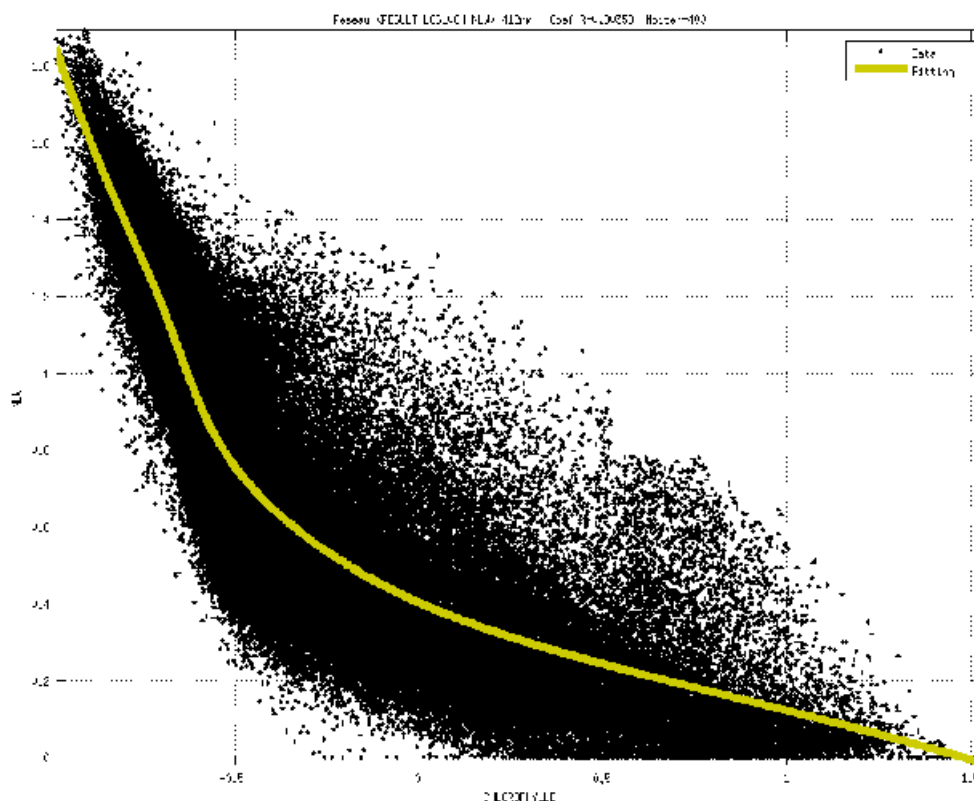


Figure 5.5 – tests du réseau [$nLw(\lambda=412nm)$]

entre les valeurs désirées et calculées par les réseaux sont fortes notamment pour les longueurs d'onde $\lambda=412, 443, 510$ et $670nm$ et moyenne pour $\lambda=555nm$ (Voir Annexe F).

5.2.7 Application du réseau MLP appris

Dans cette section, nous nous proposons de mettre en œuvre le réseau appris dans le paragraphe précédent. Pour cela, nous allons utiliser les données issues de la base de l'apprentissage de SOM3 (ISST, SST, Chl-a). Nous considérons pour cela la carte de Kohonen issue de l'apprentissage SOM3 où il a été projeté une base mensuelle. Puis, on examine les référents des classes 1 à 6. Pour chaque référent, on récupère les pixels associés aux spectres de référent, ensuite on calcule le spectre moyen correspondant. Ainsi on a un spectre moyen global qui sera associé à une valeur moyenne globale de la chlorophylle calculée en moyennant les valeurs de la chlorophylle associée au référent de la classe, on construit alors une matrice de la forme $[Chl-a \quad nLw(\lambda = 412), \dots, nLw(\lambda = 555)]$. On peut alors appliquer le réseau MLP, puis utiliser l'équation 5.1 pour calculer les $nLw^*(\lambda)$ (Figure 5.7). Les spectres calculés $nLw^*(\lambda)$

pour les six classes sont présentés dans la Figure 5.6. On remarque que les valeurs moyennes spectrales des classes 1, 2, 3, 4 sont proches les uns des autres pour les cinq longueurs d'onde de SeaWiFS, ce qui signifie que les données $nLw(\lambda)$ sont principalement affectées par des espèces de phytoplanctons contenant Chl-a.

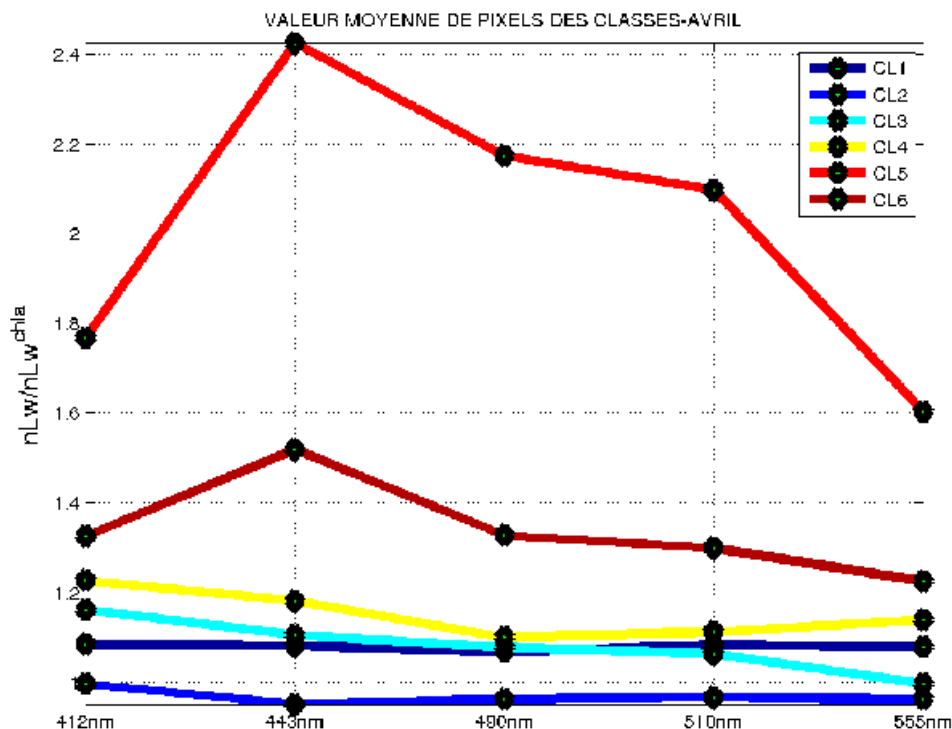
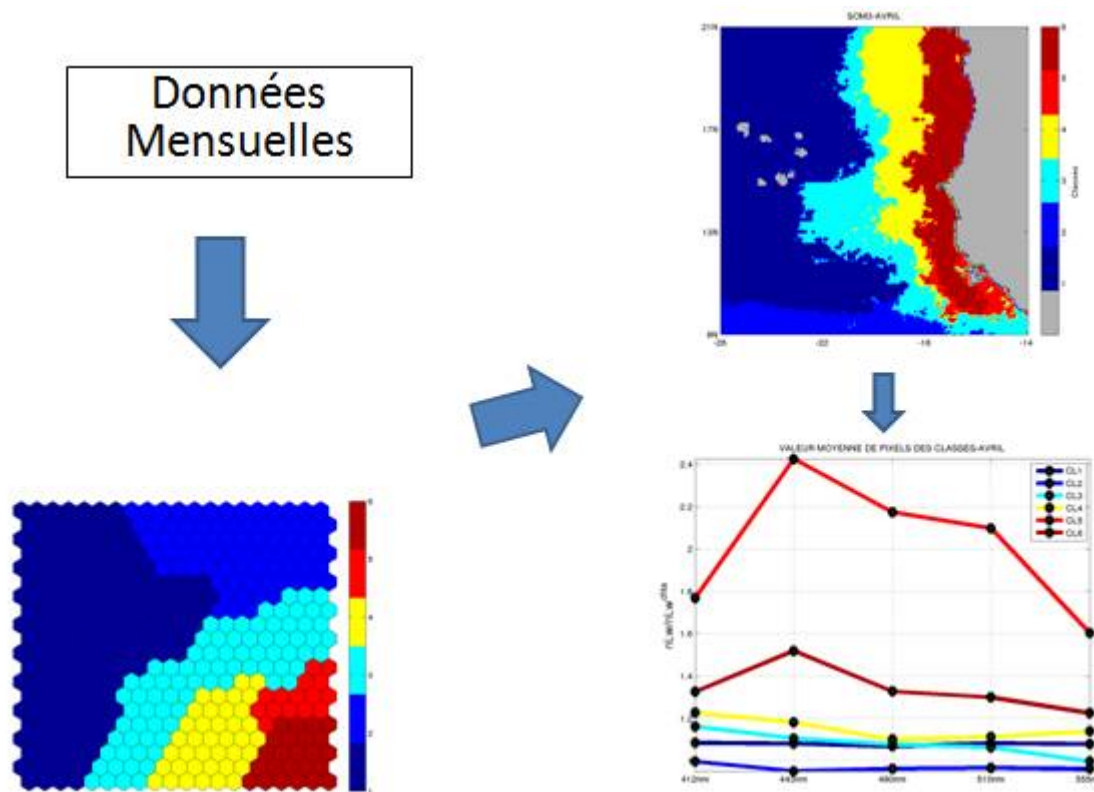


Figure 5.6 – spectres normalisés [$nLw^*(\lambda=412\text{nm à }555\text{nm})$] des classes 1 à 6.

Les valeurs moyennes de la classe 6 et de la classe 5 des spectres $nLw^*(\lambda)$ sont supérieures à 1 ce qui signifie que les données $nLw^*(\lambda)$ sont aussi affectées par d'autres constituants que Chl-a. Les fortes valeurs des $nLw^*(\lambda)$ observées dans les spectres de la classe 5 signifie que les eaux de la classe 5 reflètent plus la lumière solaire que les autres classes, et puisque les eaux de la classe 5 couvrent principalement, d'une part le plateau de la Guinée et d'autre part le Banc d'Arguin, on peut conclure que les eaux de la classe 5 sont essentiellement les eaux du cas-2. L'analyse des eaux de la classe 6 est encore plus difficile. Les valeurs spectrales $nLw^*(\lambda)$ de ces eaux sont également supérieures à 1, mais inférieures aux valeurs des spectres de la classe 5, ce qui signifie que les eaux de la classe 6 sont plus absorbante et/ou moins réfléchissante que les eaux de la classe 5. Les eaux de la classe 6 sont des eaux très productives, directement affectées par le système d'upwelling côtier. Ils favorisent le développement d'une grande variété de phytoplancton et d'espèces zooplanctoniques comme le montre la forte concentration de Chl-a) observée par le capteur SeaWiFS. Les valeurs spectrales $nLw^*(\lambda)$ de classe 6 pourrait être interprété par le développement d'espèces phytoplanctons et zooplanctons qui reflètent la lumière solaire incidente d'une part et d'autre par l'effet paquet, car ces eaux sont riches en phytoplancton. En raison de l'absence de mesure in situ dans cette zone, il est difficile de pousser plus loin l'interprétation des valeurs élevées de spectres $nLw^*(\lambda)$ de la classe 6.



5.3 Conclusion

Dans cette partie du travail de thèse, nous avons argumenté et justifié l'utilisation du réseau multicouche (MLP) par les limites de la table de référence (Alvain et al., 2005). Le développement des algorithmes a permis les prétraitements efficaces de données SeaWiFS pour la préparation en amont des réseaux. Nous avons développé six (6) MLP ayant comme entrée la concentration en chlorophylle-a et comme sortie l'une des valeurs de $nLw(\lambda=412\text{nm à } 555\text{nm})$, l'application du réseau développé sur les données de tests a donné des résultats satisfaisants ($R=0.90358$ pour $\lambda=412\text{nm}$). La mise œuvre du réseau sur les données SOM3 et l'application de la relation de l'équation 5.1 ont permis de calculer les spectres normalisés $nLw^*(\lambda)$ de classification SOM3. L'analyse de valeurs moyennes spectres de la classe 1, 2, 3, et 4 a montré qu'ils sont proches, ce qui a permis de conclure que ces classes sont principalement affectées par des espèces de phytoplanctons contenant Chl-a. Il a été remarqué que dans la classe 5, les spectres ont de hautes valeurs de $nLw^*(\lambda)$, signifie que les eaux de la classe 5 reflètent plus la lumière solaire que les autres classes, et puisque les eaux de la classe 5 couvrent principalement, d'une part le plateau de la Guinée et d'autre part le Banc d'Arguin, on a conclure que les eaux de la classe 5 sont essentiellement les eaux du cas-2. L'analyse des eaux de la classe 6 est très délicate. Les valeurs moyennes de spectres $nLw^*(\lambda)$ de ces eaux sont supérieures aux valeurs moyennes de spectres de classe 1 à 4, mais reste inférieures aux valeurs moyennes des spectres de la classe 5, ce qui signifie que les eaux de la classe 6 sont plus absorbante et/ou moins réfléchissante que les eaux de la classe 5. Les eaux de la classe 6 sont des eaux très productives directement affectées par le système d'upwelling côtier. Ils favorisent le développement d'une grande variété de phytoplancton et d'espèces. Les valeurs spectrales

$nLw^*(\lambda)$ de classe 6 pourrait donc être interprété par le développement d'espèces phytoplanctons et zooplanctons qui reflètent la lumière solaire incidente d'une part et d'autre par «**l'effet paquet**», car ces eaux sont riches en phytoplancton. En raison de l'absence de mesure in situ dans cette zone, il est difficile de pousser plus loin l'interprétation des valeurs élevées de spectres $nLw^*(\lambda)$ de la classe 6.

Chapitre 6

Classification spectrale pour la caractérisation d'espèces phytoplanctoniques

6.1 Introduction

Les données d'observation couleur de l'océan (concentration en chlorophylle-a) ont été intensivement utilisées ces dernières années pour étudier les eaux de surface des océans, les mers et des lacs. Ces données étant mesurées sur plusieurs bandes spectrales de la lumière, cette information spectrale peut être utilisée pour distinguer les différentes espèces de phytoplancton. En effet ces espèces présentent des signatures spectrales différentes en raison des propriétés d'absorption des pigments et ainsi que de diffusion qui sont spécifiques à chaque espèce. A titre d'exemple, les **diatomées**, qui ont un squelette de silice, rétrodiffuse plus fortement la lumière du soleil que les **haptophytes**.

Récemment, (Alvain et al., 2005) ont essayé de détecter les groupes de phytoplancton dans les eaux du cas-1 à partir de l'imagerie SeaWiFS. En comparant les concentrations en pigments issues de mesures in-situ avec les spectres de réflectance normalisés de SeaWiFS, pour s'affranchir de la dépendance de la concentration en chlorophylle-a, (Alvain et al., 2005) ont réussi à regrouper les réflectances de SeaWiFS en quatre groupes liés à des espèces dominantes de phytoplancton. Dans cette partie de la thèse, on se propose d'étudier les possibilités de détecter les espèces à partir de données d'observation satellitaire faites sur des réflectances marines à différents longueurs d'onde au large de l'Afrique de l'Ouest entre $8 - 24^{\circ}N$ et $14 - 30^{\circ}W$ en utilisant une procédure similaire à celle développée dans (Alvain et al., 2005).

La région de l'upwelling Sénégal-Mauritanien a été récemment étudiée de manière intensive par traitement des données couleur de l'océan issue du capteur SeaWiFS et des données de température de la surface de la mer issues du capteur NOAA/AVHRR, (Farikou et al., 2011) et (Diouf et al., 2011). Nous avons noté une quasi-permanence de l'upwelling au voisinage de la côte à partir de deux paramètres géophysiques (couleur de l'océan et SST). Nous avons aussi remarqué une forte concentration en chlorophylle-a près de la côte, un gradient fort entre la côte et le large avec une forte variabilité saisonnière (Farikou et al., 2011) et (Sawadogo et al., 2009). L'upwelling est modulé par l'intensité de la composante parallèle à la côte du vent ; il s'affaiblit en Juillet-Août avec la remontée de l'ITCZ vers le nord, ce qui diminue l'intensité des alizés. En plusieurs endroits (plateau de banc d'Arguin et au large du delta du Sine-Saloum), nous avons également noté la présence d'eaux dont les spectres correspondent à

des eaux du cas-2. Dans les paragraphes suivants, nous allons analyser les spectres de couleur de l'océan fournis par le capteur SeaWiFS en se focalisant sur l'identification des espèces phytoplanctoniques.

6.2 Les Données

les données de spectres $\rho_w(\lambda)$ utilisées sont issues des travaux de (Diouf et al., 2011) et sont disponibles sur le site : <http://poacc.locean-ipsl.upmc.fr/>. Ces réflectances sont obtenues par le traitement des réflectances journalières du capteur SeaWiFS observées au-dessus de l'atmosphère (TOA) par l'algorithme Neuro-Varia sur la période de 13 ans (1998 à 2010). L'algorithme utilisé a été spécifique, et a pris en compte l'absorption des aérosols. Afin d'extraire les effets du second ordre dus aux caractéristiques spécifiques de l'eau autres que celle de l'abondance de phytoplancton, on a calculé une réflectance spécifique $\rho_w^*(\lambda)$ définie par :

$$\rho_w^*(\lambda) = \frac{\rho_w(\lambda)}{\rho_w^{ref}(\lambda, chl - a)} \quad (6.1)$$

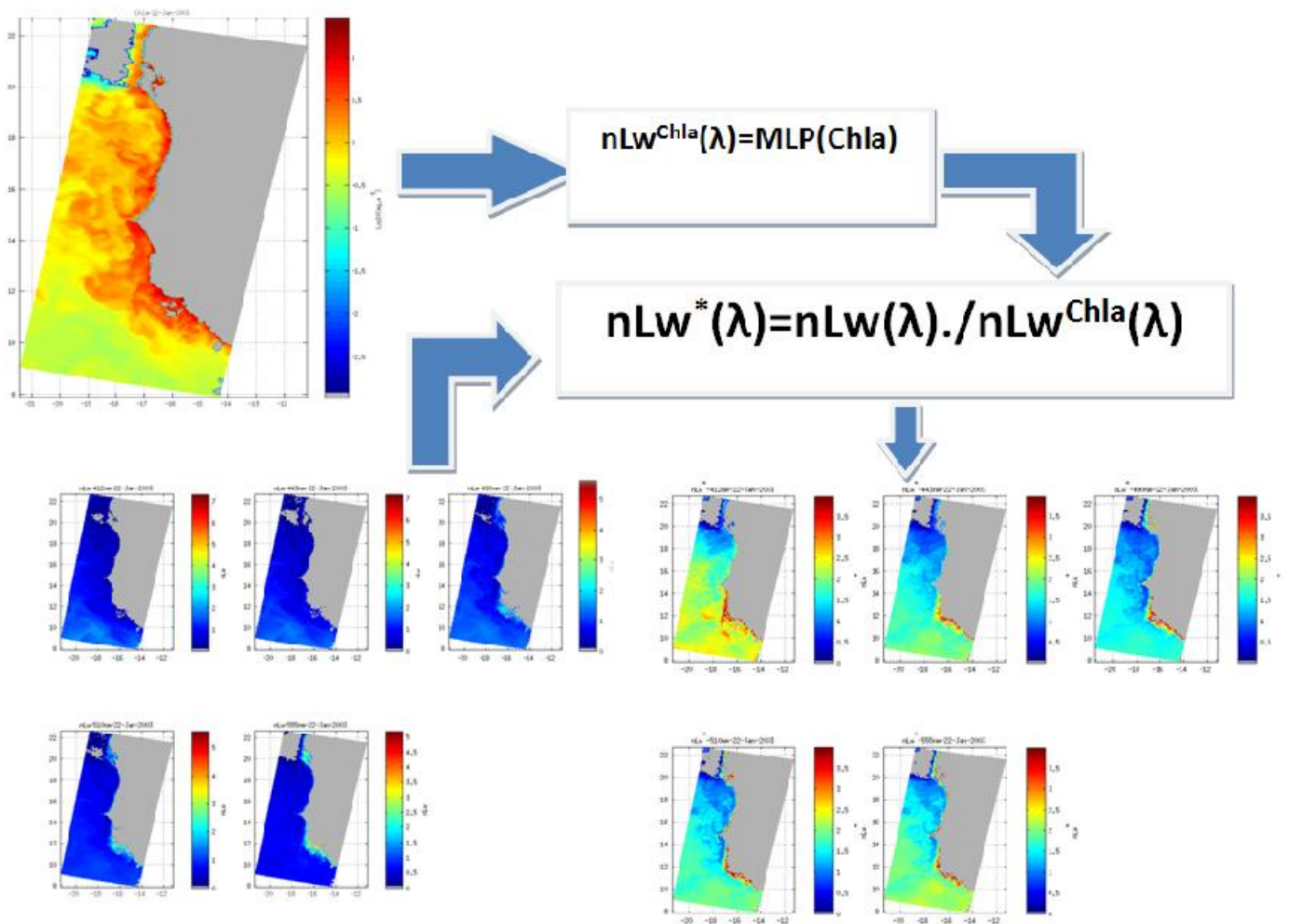
où $\rho_w^{ref}(\lambda, chl - a)$ est un modèle simple qui tient compte uniquement de la chlorophylle-a standard de SeaWiFS. Dans cette partie,

- $\rho_w^{ref}(\lambda, chl - a)$ est calculé comme $nLw^{ref}(\lambda, chl - a) \times E_0(\lambda)$ où $E_0(\lambda)$ est le rayonnement solaire,
- $nLw^{ref}(\lambda)$ ($nLw^{Chla}(\lambda)$) est fournie par le modèle développé par (Farikou et al., 2011) à l'aide de perceptrons multicouche (PMC) aptes à modéliser la relation non-linéaire ($Chl - a \rightarrow nLw^{ref}(\lambda)$) à partir d'une base de données. La Figure 6.1 montre le processus mise en œuvre pour calculer les réflectances normalisées $nLw^*(\lambda)$, L'exemple présenté pour la date de 22 janvier 2003 qui est peu nuageuse. On peut remarquer sur ces images de forts changements dans la gamme des longueurs d'onde 412nm à 555nm. Ceci est dû effectivement à la concentration en chlorophylle-a qui a une action dominante dans ce domaine spectral. On peut dire que la méthode de normalisation proposée a permis d'atténuer fortement l'effet de la concentration en chlorophylle-a sur les spectres. Ensuite, nous avons appliqué le processus de calcul à l'ensemble de données journalières disponibles sur le site : <http://poacc.locean-ipsl.upmc.fr/> (1997 à 2010).

Nous proposons maintenant de classier les spectres $nLw^*(\lambda)$ afin de tenter de différencier les groupes d'espèces phytoplanctoniques.

6.3 la méthodologie

La méthode adoptée est similaire à celle utilisée dans (Farikou et al., 2011). Nous avons d'abord appliqué une classification non supervisée analogue à celle effectuée par (Niang et al., 2003) et (Niang et al., 2006) sur une base de données journalière composée de $nLw^*(\lambda)$ de l'année 2003 qui est une année où la couverture nuageuse était très faible (Diouf et al., 2011). L'objectif est de résumer l'information contenue dans la base de données D par les cinq composantes vectorielles ($nLw^*(\lambda = 412nm \text{ à } 555nm)$) en produisant un nombre de vecteurs réduit de référence (rv) qui sont statistiquement représentatif des


 Figure 6.1 – processus de calcul de $nLw^*(\lambda)$

données. Chaque vecteur référent rv représente un ensemble de vecteur de D qui ont des caractéristiques similaires. Ces vecteurs référents sont statistiquement représentatifs de différentes classes de vecteurs de la base de données D . Dans le cadre de ce travail, nous avons choisi de déterminer les rvs en minimisant la distance Euclidienne entre rv et les vecteurs qui lui sont associés. Pour ce faire, nous avons utilisé un modèle de réseau de neurone spécifique basé sur la carte auto-organisatrice SOM introduite par (Kohonen, 1982). Dans l'étude que nous avons menée, nous avons utilisé une carte topologique ayant un grand nombre de neurones (20×20) correspondant à 400 groupes, fournissant une représentation très discriminante des observations. Les pixels de l'ensemble de données D sont donc regroupés en 400 groupes. Nous avons utilisé la version de SOM chargée sur le site web : <http://www.cis.hut.fi/projects/somtoolbox/download/>. La carte topologique a été apprise selon la procédure décrite dans Kohonen (2001). Le nombre de neurones a été déterminé empiriquement à partir des solutions similaires au problème et ensuite ajustées comme décrit dans (Badran et al., 2005). Si le nombre élevé de groupe nous permet de prendre en compte toute la complexité de l'ensemble, il s'avère être un obstacle pour extraire des interprétations géophysiques simples des phénomènes liés aux données. Pour pallier cela nous avons agrégé des groupes ayant des propriétés

communes en un nombre limité de groupe. Nous avons utilisé pour cela une classification hiérarchique ascendante (CAH) selon la procédure décrite dans (Jain and Dubes, 1998). Cette procédure est itérative : à chaque itération, elle agrège deux groupes similaires selon le critère de Ward. Dans le cas présent nous avons stoppé le processus à dix classes. Les 400 neurones sont donc agrégés en 10 classes significatives.

6.3.1 Résultats de l'apprentissage

Nous avons utilisé 329 images journalières de l'année 2003 comme indiqué dans (Diouf et al., 2011). La base est formée de 5 variables $nLw^*(\lambda = 412nm \text{ à } 555nm)$ et de taille 1279713. Cette base a été échantillonnée en prenant une ligne sur 10. La taille de la base d'apprentissage est donc de 127972 données de dimension 5. La base d'apprentissage, après normalisation, est présentée à la carte SOM de taille 20x20 pour un apprentissage. A l'issue de l'apprentissage on récupère en sortie : la carte des référents, la carte des cardinalités et le vecteur des neurones gagnant. Les cartes de distribution

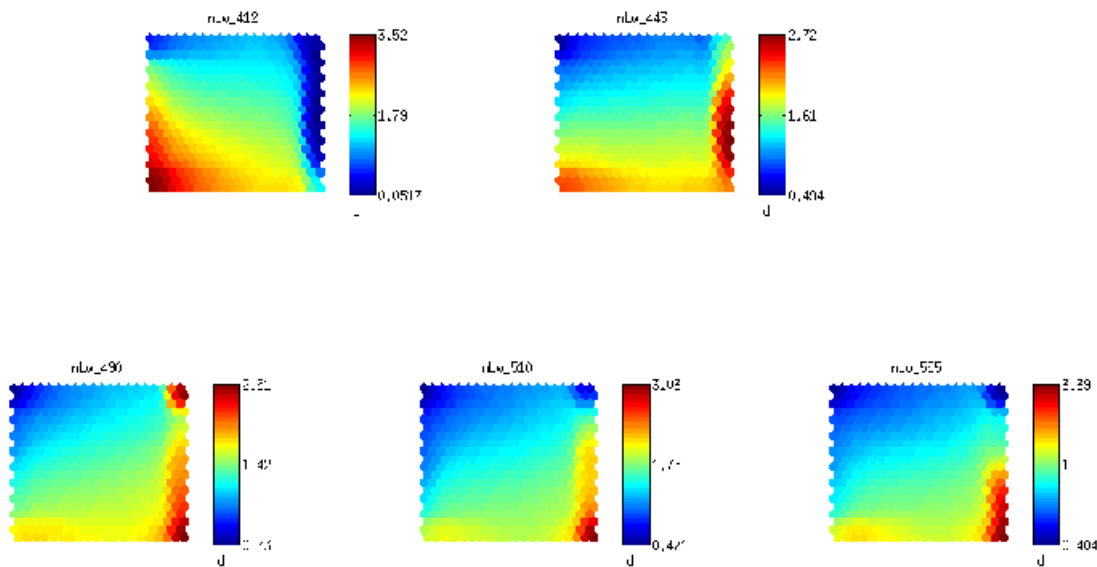


Figure 6.2 – Cartes de distribution issues de la carte auto-organisatrice de Kohonen. Les cellules bleues foncées représentent des valeurs faibles, alors que les cellules rouges foncées (couleur chocolat) représentent des valeurs élevées. L'échelle d indique les valeurs transformées de l'unité $nLw^*(\lambda)$

(« distribution maps » ou « component planes »), issues de la carte de Kohonen, sont représentées à la Figure 6.2. Elles permettent de visualiser la répartition des $nLw^*(\lambda)$ en fonction de leurs caractéristiques. On peut remarquer une bonne organisation des neurones dont les paramètres varient de façon continue.

6.3.2 Détermination des classes de $nLw^*(\lambda)$

La **U-matrix** et la classification descendante hiérarchique de Ward ont ensuite permis de regrouper les cellules en classes (agrégation de groupes) de $nLw^*(\lambda)$. La carte topologique obtenue est représentée sur la Figure 6.3. On remarque que la carte topologique associée à la CAH est très cohérente, puisque les classes représentent des groupes dont les neurones sont contigus sur la carte topologique. En plus,

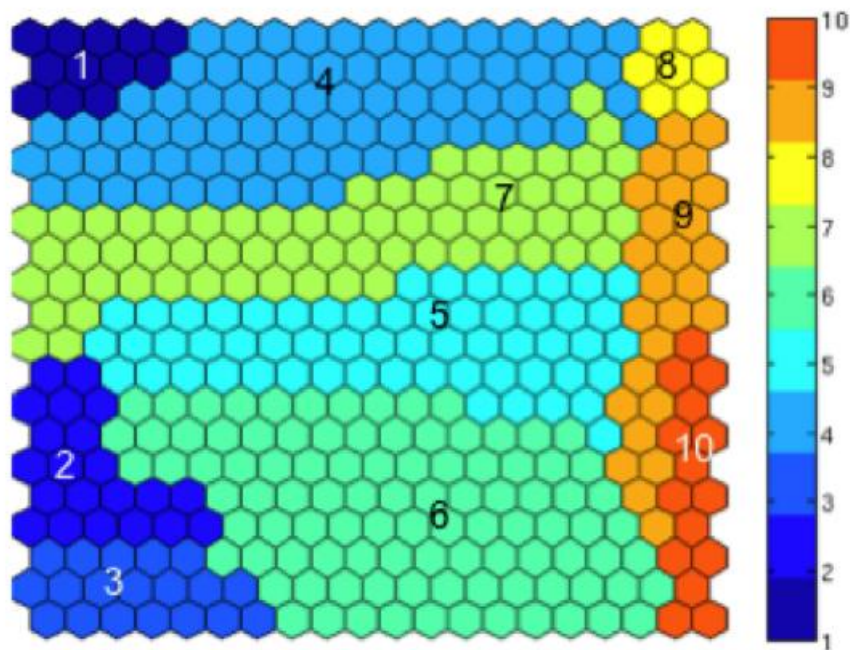


Figure 6.3 – Résultat pour les dix classes de $nLw^*(\lambda)$ formés à partir de la carte auto-organisatrice de Kohonen.

les spectres associés à chaque neurone (ou rv) déterminent des classes homogènes sur SOM. Le nombre de groupes (10) a été choisi d'une part parce qu'il présente la partition la plus discriminante à l'égard du dendrogramme de la CAH (non représentée) et d'autre part par l'homogénéité des spectres.

6.3.3 Analyse spectrale

Les formes spectrales des dix classes $nLw^*(\lambda)$ obtenues après la CAH sont représentées sur la figure 6.4. Les trois derniers spectres (8, 9, 10) sont caractéristiques des eaux du cas-2 avec une absorption élevée à 412 nm due à des substances organiques dissoutes. Les valeurs élevées du spectre de la classe 10 aux grandes longueurs d'onde correspondent à la rétrodiffusion due aux particules en suspension. Les faibles valeurs des spectres de la classe 8 aux grandes longueurs d'onde pourraient correspondre à l'absorption due à la matière organique dans les eaux très côtières. Les autres classes peuvent être analysées à travers la classification proposée par (Alvain et al., 2005). Les classes 2 et 3 peuvent correspondre à des diatomées ; en raison de leur proximité sur les cartes topologiques et la forme similaire de leurs spectres, ces classes pourraient être fusionnées. La classe 1 peut correspondre à des haptophytes, les classes 4, 5 et 7 à des SLC et *Prochlorococcus* et la classe 6 à des diatomées comme les classes 2 et 3. L'analyse de ces classes est une procédure très délicate car les spectres pourraient avoir été contaminés par petits défauts lors de la correction atmosphérique. Afin de vérifier ce point, nous avons représenté l'épaisseur optique atmosphérique (AOT dans ce qui suit) moyenne qui est associée aux réflectances captées par chaque neurone (Figure 6.5). Nous voyons sur cette figure que les classes 2 et 3 sont associées à des valeurs de l'AOT assez élevées et par conséquent nous ne les avons pas prises

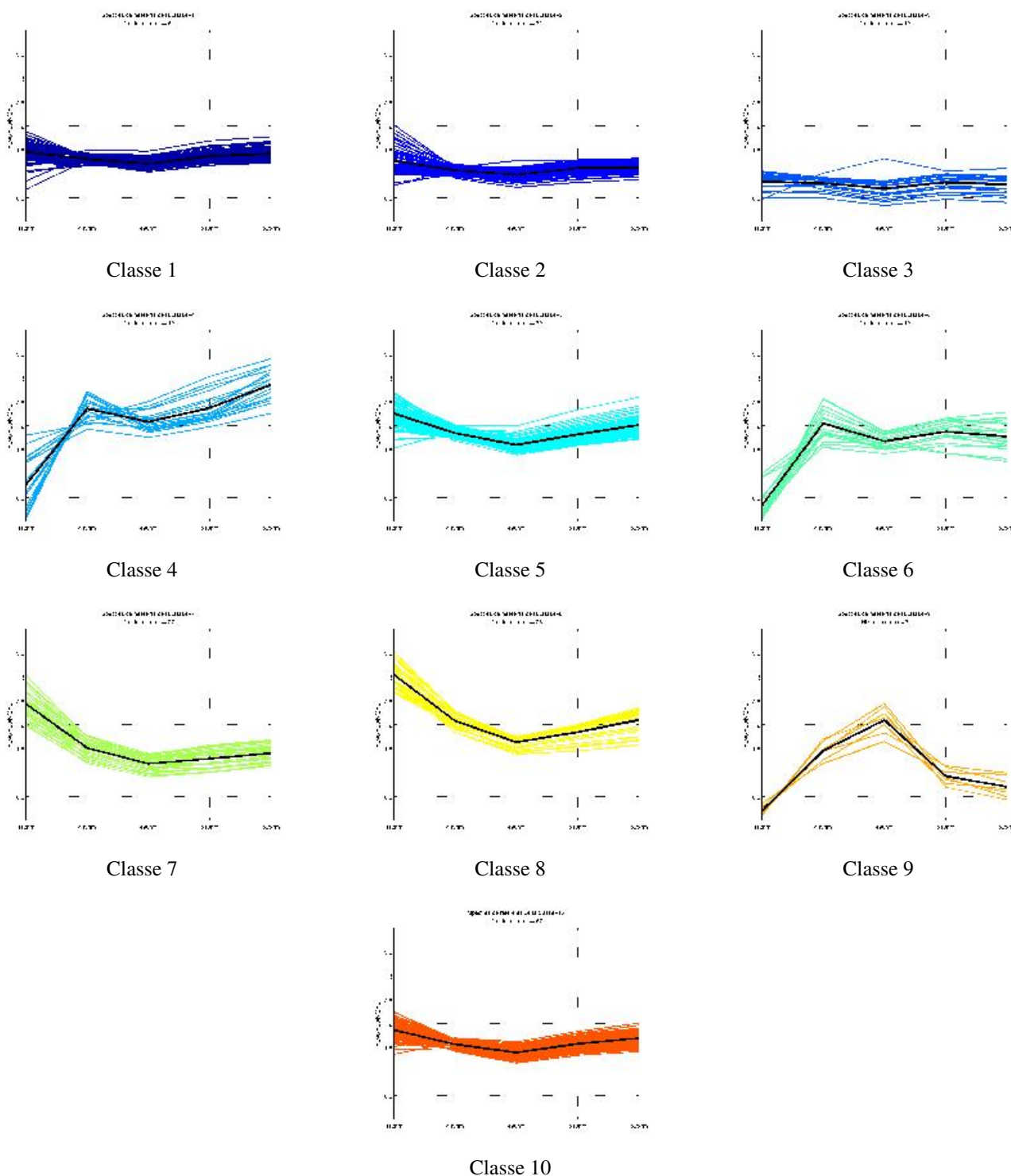


Figure 6.4 – Représentation des dix spectres de $nLw^*(\lambda)$.

en considération pour en faire une interprétation biologique et analyser leur répartition géographique, puisque ces deux classes peuvent être potentiellement contaminées par l'atmosphère. Pour les mêmes raisons nous n'avons pas interpréter les classes 8, 9 et 10. Nous avons donc décidé de concentrer notre analyse sur les classes 4, 5, 6, 7 seulement, qui sont associés à des valeurs AOT faible (moins de 0,25). Nous avons par ailleurs étudié la forme géographique de spectres de la classe 1 pour lequel, nous avons supprimer les données associées au neurone 1 qui sont très contaminées par l'atmosphère comme le

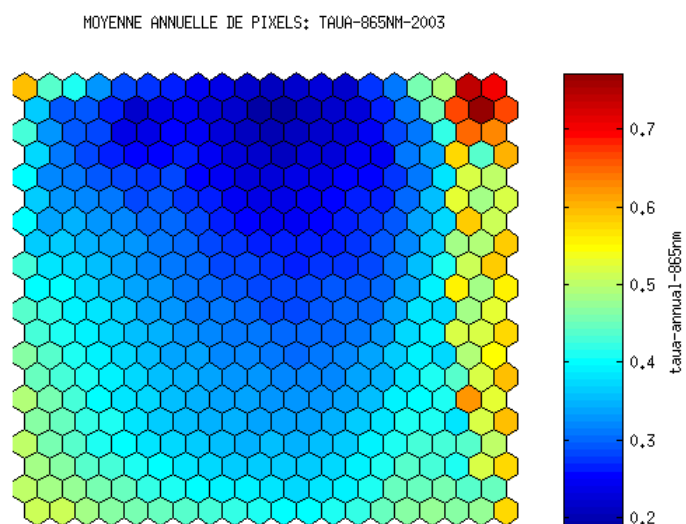


Figure 6.5 – Profondeur moyenne optique associée à chaque facteur de réflectance captée par chaque neurone de la carte SOM. La barre de couleur sur la droite est l'échelle la valeur AOT.

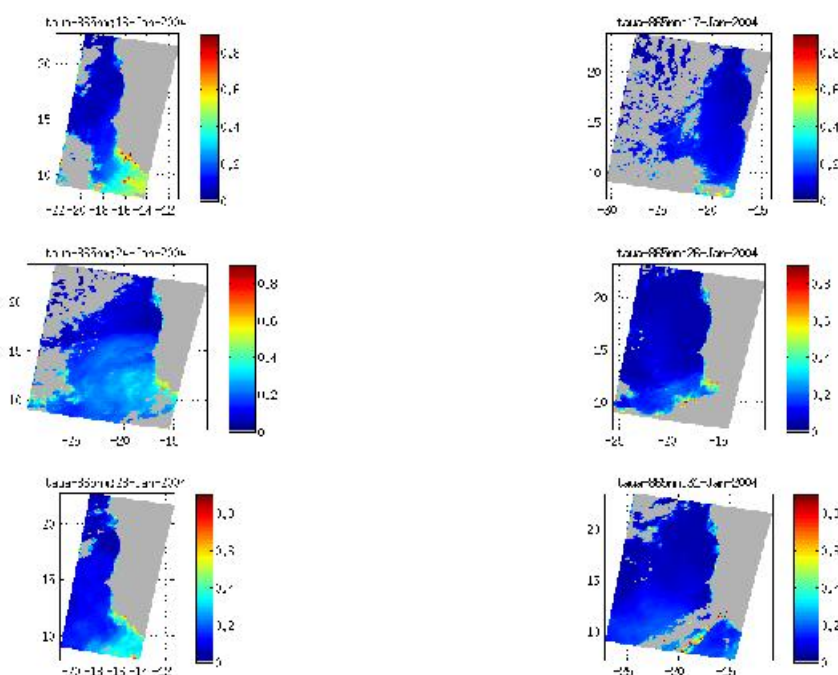


Figure 6.6 – Pattern de AOT pour les dates : 16, 17, 24, 26, 28, 31 Janvier 2004

montre les valeurs fortes moyennes de AOT.

Nous allons analyser les images satellitaires prises à plusieurs dates correspondant aux faibles valeurs de l'AOT. Parmi ces dates, nous avons choisi six jours de Janvier 2004 (16, 17, 24, 26, 28, 31) qui sont caractéristiques des situations observées au cours de la période de treize ans de données traitées. Les images de concentration de chlorophylle-a (Figure 6.7) montrent un développement de l'upwelling avec des valeurs de concentration très élevées à la côte. La concentration en chlorophylle-a décroît

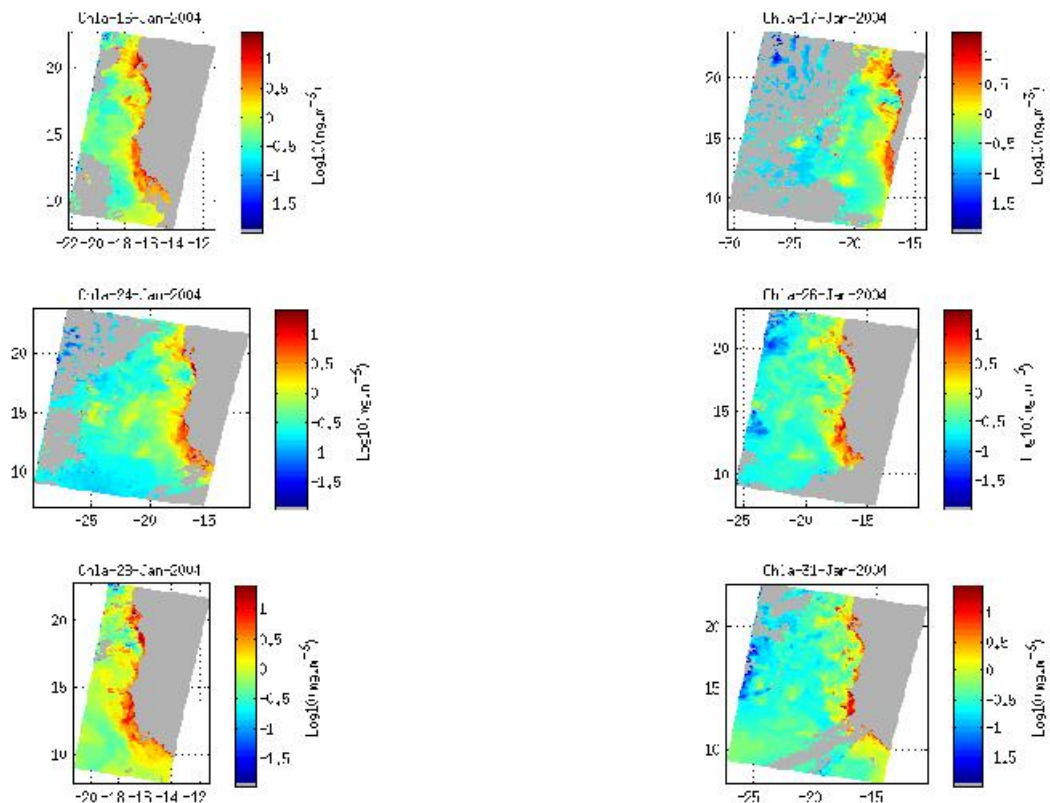


Figure 6.7 – Carte de concentration en chlorophylle-a pour les date de : 16, 17, 24, 26, 28, 31 Janvier 2004

vers le large où elle est advectée par les méandres liés à la dynamique non linéaire de l'upwelling, méandres qui peuvent persister plusieurs jours (Cap-Vert, images de janvier 17, 26, 31). Nous notons des concentrations de chlorophylle-a très élevées sur le banc d'Arguin ($20^{\circ}N$) et le Sud du Cap-Vert sur les bancs du Siné-Saloum et de la Guinée, la Guinée banque ($14^{\circ} - 11^{\circ}N$). Les cartes de signature spectrale (Figure 6.7, 6.8) montrent que les eaux de classe 1 sont situées près de la côte (dans les régions qui s'étendent de quelques dizaines de kilomètres des côtes) formant une bande parallèle à la côte ; elles sont bordées du côté ouest par les eaux de la classe 4, qui occupent l'océan profond.

Une interprétation possible de ces images peut-être la suivante : l'upwelling côtier génère des phytoplanctons dont le spectre moyen correspond à la classe 1 ; au large (océan profond) on trouve des phytoplanctons dont le spectre correspond à la classe 4, caractérisant la zone océanique de grande profondeur ayant une faible concentration en chlorophylle-a.

Les Bancs d'Arguin et du Siné-Saloum, les bancs de Guinée, qui sont des régions très peu profondes, sont caractérisés par des eaux de classes 8, 9, 10 dont les spectres correspondent aux eaux du cas-2 comme argumenté plus haut. Dans la partie sud de la région, on observe au large des eaux de classe 6, qui sont bien marquées sur la figure 6.8.

Le spectre de la classe 6 correspond à des diatomées comme indiqué dans (Alvain et al., 2005). Ces diatomées commencent à se développer le 17 Janvier près de la côte vers $10^{\circ}N$, présentent une

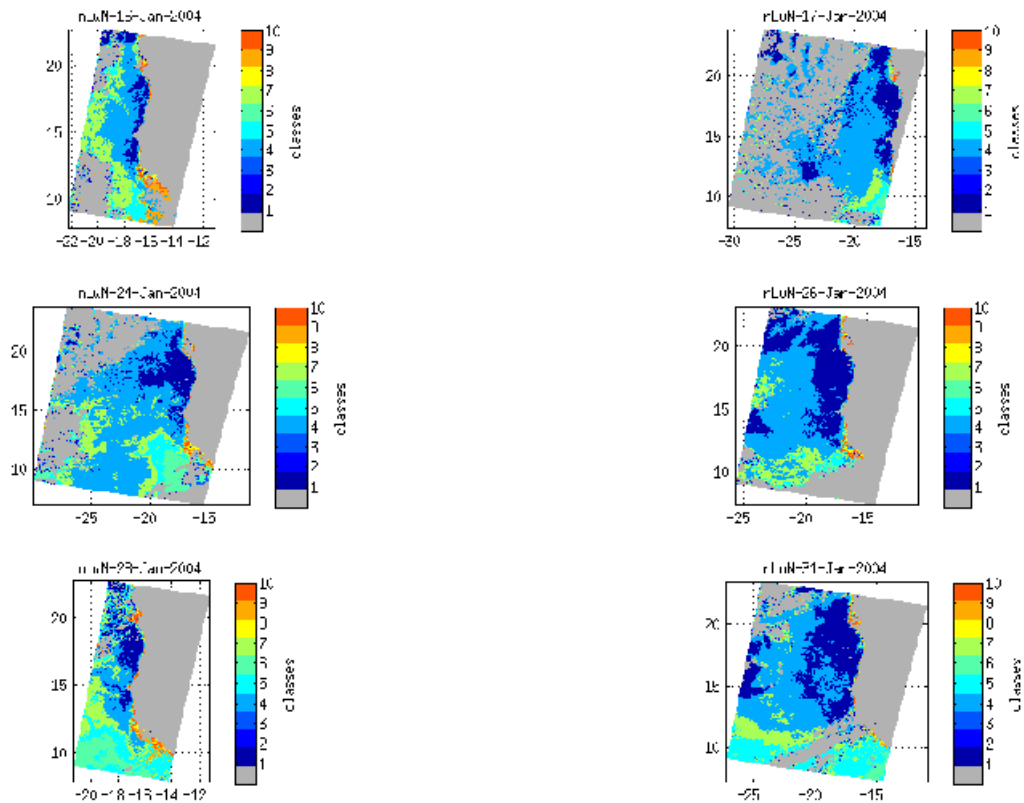


Figure 6.8 – Patterns de spectres pour le 16, 17, 24, 26, 28, 31 Janvier 2003 pour lesquels une couverture AOT est très faible.

extension bien définie le 24 Janvier, progressent vers l'ouest en date du 28 Janvier et sont encore visibles le 31 Janvier. Cette séquence dans le temps peut être identifiée comme un "bloom" de diatomées se développant au niveau du courant ECC qui est présent au sud de la région étudiée (Farikou et al., 2011). Ces situations sont observées plusieurs fois au cours de la période de treize ans.

6.3.4 Discussions et Conclusions

La procédure de la classification neuronale a permis d'identifier 10 classes spécifiques de spectres $\rho_w^{ref}(\lambda)$ (le spectre de réflectance $\rho_w(\lambda)$ normalisée par $\rho_w^{ref}(\lambda)$). En projetant les AOTs sur le classifieur, nous avons trouvé que plusieurs classes (2,3 et 8,9,10) sont contaminées par l'atmosphère. Les classes 2 et 3 sont rencontrées au large de la région étudiée. Nous n'avons pas pu interpréter ces classes en termes d'espèces phytoplanctoniques car cette interprétation pourrait être biaisée par le signal atmosphérique. Concernant les classes 8, 9 et 10 qui correspondent aux eaux côtières, elles occupent une région bien définie près de la côte, ces eaux sont identifiées comme étant des eaux du cas-2 car ayant une forte absorption à 412nm.

Ces eaux sont principalement rencontrées sur le Banc d'Arguin ($20^\circ N$) et sur le plateau continental au large du Siné-Saloum et de la côte de la Guinée qui sont des régions très peu profondes et très sensibles aux matières dissoutes (spectres de la classe 8) apportées par le ruissellement des eaux continentales et

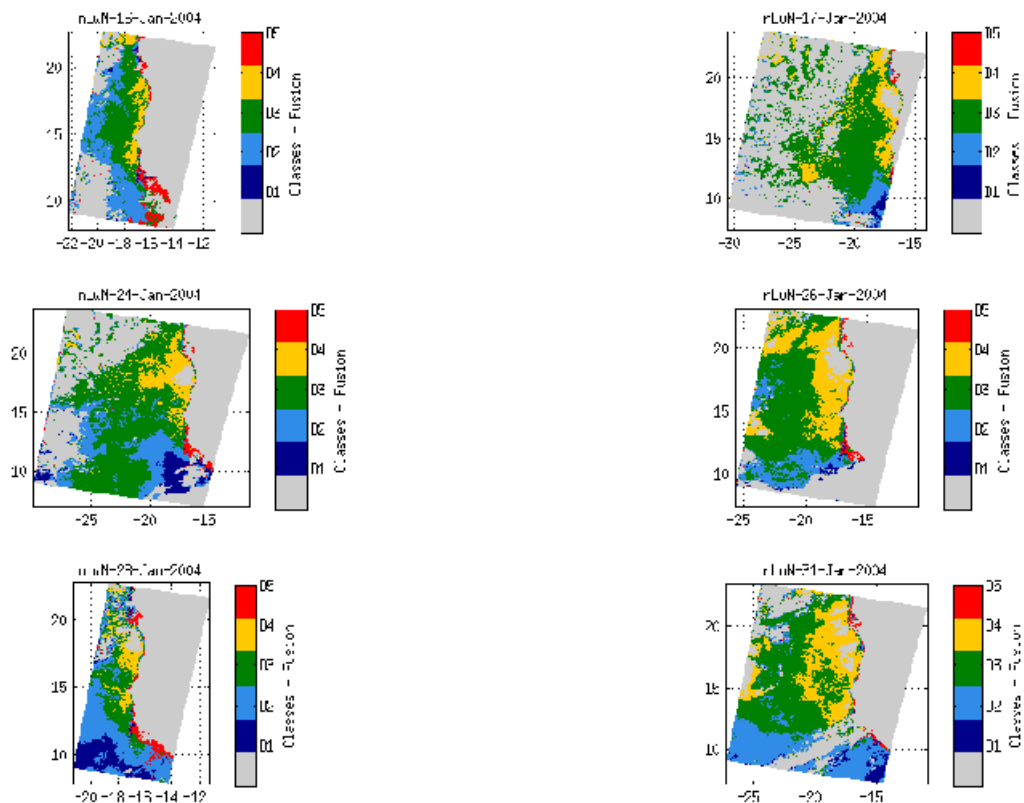


Figure 6.9 – Patterns des spectres pour les journées du 16, 17, 24, 26, 28, 31 Janvier 2003 pour lesquels nous avons amélioré la classe 1 (D4, haptophytes), classe 5 (D1, diatomées), classe 4 (D3, SLC). Nous avons fusionné les classes 2, 3, 5 et 7 qui ne peuvent être déterminées maintenant en D2 et les classes 8, 9 et 10 sont associées aux eaux du cas-2 en D5.

aux particules en suspension réfléchissant la lumière (spectres de la classe 10).

Après avoir examiné un grand nombre de situations pour lesquelles les AOTs sont faibles, nous sommes arrivés à décrire le scénario suivant pour l'interprétation de la variabilité spatiale des spectres en terme de paramètres biogéochimiques dans la région d'upwelling :

L'upwelling produit des espèces phytoplanctoniques qui sont caractérisées par les spectres de la classe 1. Selon (Alvain et al., 2005), le spectre de la classe 1 peut-être associé aux haptophytes. Les eaux du large sont caractérisées par les spectres de la classe 4, qui sont associées à des prochlorococcus et des SLC (Synechococcus-like et cyanobacteria). Nous avons observé une présence quasi-permanente des diatomées dans les eaux de courant ECC. La difficulté d'interprétation des différents spectres et de les attribuer à des espèces phytoplanctoniques spécifiques est en partie dû à la difficulté d'obtenir une correction atmosphérique très précise. De plus, les classes sont des indices qualitatifs, le passage d'une classe à un autre est brusque et discontinu, une espèce ayant une faible concentration pouvant avoir la même signature spectrale que cette même espèce à forte concentration. Le mélange de plusieurs espèces peut aussi rendre leur interprétation difficile.

Conclusion Générale

L'objectif principal de ce travail était la classification neuronale de données satellitaires pour la caractérisation de l'upwelling Sénégal-mauritanien. Ce travail est une continuité de (Sawadogo et al., 2009) et a permis de confirmer les résultats obtenus sur une période de 5 ans. L'apport de ce travail réside principalement sur la caractérisation des zones marines basées non seulement sur les indicateurs statistiques mais surtout sur les spectres optiques des espèces phytoplanctoniques du domaine Sénégal-Mauritanien étudiées.

Ce travail s'est appuyé principalement sur les données satellitaires avec l'utilisation des méthodes neuronales en l'occurrence les cartes de Kohonen et les perceptrons multicouches.

Dans ce travail, nous nous sommes essentiellement focalisés sur le phénomène de l'upwelling et ses conséquences dans la zone Sénégal-mauritanien. Ce travail a permis de caractériser l'upwelling côtier de la zone d'étude et de décrire sa variabilité saisonnière et interannuelle à partir des observations satellitaires de température de la surface de la mer SST et de couleur de l'océan sur une période de 10 années (1998-2007).

Cette étude a montré que l'on pouvait décomposer les eaux de la zone Sénégal-Mauritanien en 6 classes à l'aide des cartes auto-organisatrices en analysant d'abord la chlorophylle et les gradients zonaux de température (SOM2) puis ensuite en analysant la chlorophylle, les gradients zonaux de température et les températures de surface (SOM3). Les six classes sont bien définies et correspondent chacune à une zone géographique bien délimitée montrant la robustesse de la méthode utilisée. Les six classes peuvent être considérées comme des indicateurs statistiques pour analyser la variabilité spatio-temporelle de l'upwelling sénégal-mauritanien. Les caractéristiques de ces 6 classes sont :

- La classe 1 correspond à des eaux où la concentration en Chl-a est très faible et l'ISST voisine de zéro. Elle est associée aux eaux du large.
- La classe 2 correspondent à des eaux dont la concentration en Chl-a est un peu plus forte et dont la température moyenne est un peu plus forte. Les eaux de la classe 2 sont situées au sud de la région étudiée durant l'hiver et le printemps. Elles se déplacent vers le nord au début de l'été (Juin) et atteignent 17°N en Septembre, reste à cette latitude jusqu'en Octobre et redescendent vers le sud à la fin du mois de novembre. Les eaux caractéristiques de ces eaux pourraient suggérer qu'elles sont liées au contre-courant équatorial Nord Atlantique (North Atlantic Equatorial Counter Current NECC) qui est dirigé vers l'Est et s'intensifie en été pour former le Dôme de Guinée. Durant leur progression vers le nord en été, ces eaux n'envahissent jamais le plateau continental du Golfe de

Guinée qui reste recouvert des eaux de la classe 5 et de la classe 6.

- La classe 3 et la classe 4 correspondent respectivement à des eaux dont la concentration en Chl-a va en décroissant de l'est vers l'ouest et dont les ISST ont tendance à diminuer de l'est vers l'ouest.
- La classe 5 est atypique. Elle est caractérisée par des eaux riches en chlorophylle et de ISST et de SST chaude. On la trouve surtout sur le plateau de la petite côte et sur le plateau continental du Golfe de Guinée, et aussi sur le banc d'Arguin. Elle présente une grande variabilité saisonnière.
- La classe 6 correspond à des eaux qui ont des très hautes concentrations en Chl-a et des ISST et SST très froides. Elle forme une bande étroite située le long de la côte et présente une grande variabilité saisonnière. Elle caractérise les eaux de l'upwelling côtier.

L'analyse de ces classes a permis de mettre en évidence que cette zone présente une très forte variabilité saisonnière qui est due à la fois au vent, à l'ensoleillement, et aussi à la circulation océanique de grande échelle (NECC). Le vent est responsable de la variabilité saisonnière de l'upwelling, en particulier de la position des classes 6, 4 et 3. La disparition de ces classes au sud de 20°N en été est due à l'arrivée de l'ITCZ et de la disparition des alizées et il apparaît d'après nos études qu'il pourrait exister des eaux du cas-2 associées à la classe 5 ; ces eaux sont principalement situées sur le plateau continental de la petite côte. Ceci pourrait être analysé en considérant les spectres de réflectance marine.

Nous avons aussi introduit des données nLw (normalized water leaving radiance) qui contiennent des informations supplémentaires sur la composition biogéochimique de l'eau (Alvain et al., 2005) afin de compléter les analyses des eaux de la classe 5. Des études statistiques de spectres des différentes classes ont été effectuées et ont montré une forte dominance de l'effet de la chlorophylle-a sur les spectres de classe 5. Pour s'affranchir de ces effets, une méthode de normalisation proposée par (Alvain et al., 2005) et appliquée sur les données de la classification SOM3, a permis d'atténuer l'effet de la chlorophylle-a l'équation 5.1. Cependant l'algorithme (Alvain et al., 2005) est limité par les valeurs de concentration en chlorophylle-a (de 0,005 à 4mgm^{-3}), ce qui est bien en deçà des concentrations en chlorophylle-a qu'on peut observer dans le domaine océanique Sénégal-Mauritanien où les concentrations peuvent aller au-delà 10mgm^{-3} . Ce modèle de normalisation exige par ailleurs une interpolation linéaire entre deux valeurs discrètes pour obtenir la valeur correcte de nLw(Chl-a) pour une valeur donnée de la concentration en chlorophylle-a car la fonction nLw(Chl-a) est tabulée.

Nous avons ainsi fait recours aux perceptrons multicouches qui sont bien adaptés pour la modélisation de relation non linéaire (Thiria et al., 1993) et présentent l'avantage de fournir des fonctions continues au lieu d'utiliser le modèle tabulé proposé dans (Alvain et al., 2005), pour développer un modèle de normalisation adaptée de la zone Sénégal-Mauritanien et calculer alors les spectres normalisés $nLw * (\lambda)$. Ainsi six (6) réseaux de neurones multicouches ayant comme entrée la concentration en chlorophylle-a et une sortie l'une de nLw ($\lambda=412\text{nm}$ à 555nm) ont été développés. La mise œuvre de ces réseaux sur les données SOM3 combinée à l'équation 5.1 ont permis de calculer les spectres normalisés $nLw * (\lambda)$ pour les classes données par SOM3. L'analyse de valeurs moyennes spectres normalisés de la classe 1, 2, 3, et 4 a montré qu'ils sont proches, ce qui a permis de conclure que les eaux de ces classes sont principalement affectées par des espèces de phytoplanctons contenant de la Chl-a. Les spectres normalisés de la classe 5 ont de grandes valeurs de $nLw * (\lambda)$, signifiant que les eaux de

la classe 5 reflètent plus la lumière solaire que de classe 1 à 4, et puisque les eaux de la classe 5 couvrent principalement, le plateau de la Guinée d'une part et d'autre part le Banc d'Arguin, qui sont des régions océaniques de petit fond (25m), on en a conclu que les eaux de la classe 5 sont essentiellement des eaux du cas-2.

L'analyse des eaux de la classe 6 est très délicate. Les valeurs moyennes de spectres $nL_w * (\lambda)$ de ces eaux sont supérieures aux valeurs moyennes des spectres des classes 1 à 4, mais restent inférieures aux valeurs moyennes des spectres de la classe 5, ce qui signifie que les eaux de la classe 6 sont plus absorbantes et/ou moins réfléchissantes que les eaux de la classe 5. Les eaux de la classe 6 sont des eaux très productives et directement affectées par le système d'upwelling côtier. Elles favorisent le développement d'une grande variété de phytoplancton et d'espèces. Les valeurs spectrales $nL_w * (\lambda)$ de la classe 6 pourraient donc être interprétées par le développement d'espèces de phytoplanctons qui reflètent la lumière solaire incidente, car ces eaux sont riches en phytoplancton.

En raison de l'absence de mesure in situ dans cette zone, il est difficile de pousser plus loin l'interprétation des spectres de $nL_w * (\lambda)$. D'autre part les données analysées sont des données mensuelles qui peuvent être le mélange d'une large variété d'organismes phytoplanctoniques difficiles à identifier et on a décidé d'arrêter les analyses des données mensuelles.

Nous avons, dans une seconde étude, exploité les données journalières de SeaWiFS qui devraient être plus 'pures' (caractéristiques) car moins mélangées que les données mensuelles. Pour cela nous avons utilisé les données journalières issues de travaux de (Diouf et al., 2011).

Associée à la procédure de normalisation spécifique des $nL_w * (\lambda)$ cette deuxième étude constitue l'apport original de ce travail de thèse.

Cette procédure a permis d'identifier à l'aide d'un classifieur neuronal, 10 classes spécifiques de spectres $\rho_w^{ref}(\lambda)$. En projetant les AOTs sur ce nouveau classifieur, nous avons trouvé que plusieurs classes (2,3 et 8,9,10) sont contaminées par l'atmosphère. Les classes 2 et 3 sont rencontrées au large de la région. Les classes 8, 9 et 10 qui correspondent aux eaux côtières, occupent une région bien définie près de la côte ; ces eaux sont identifiées comme les eaux du cas-2 ayant une forte absorption à 412nm. Ces eaux sont principalement rencontrées sur le Banc d'Arguin (20°N) et sur le plateau continental au large de Siné-Saloum et au large de la côte de la Guinée, qui sont des régions très peu profonde et très sensible à la matière dissoute. Après avoir examiné un grand nombre de situation pour lesquelles les AOTs sont faibles, nous sommes arrivés à décrire les scénarios suivants pour l'interprétation de la variabilité spatiale de spectres en termes de paramètres biogéochimiques dans la région d'upwelling :

- l'upwelling produit des espèces de phytoplanctoniques qui sont caractérisés par le spectres de la classe 1. Selon (Alvain et al., 2005), le spectres de la classe 1 peut-être associées à des haptophytes. Les eaux du large sont caractérisées par le spectres de la classe 4, qui sont associées à prochlorococcus et SLC (Synechococcus-like et cyanobacteria). Nous avons observé une présence quasi-permanente des diatomées dans les eaux de courant ECC.
- La difficulté d'interprétation des différents spectres et de les attribuer à des espèces phytoplanctoniques spécifiques est en partie dû à la difficulté d'obtenir une correction atmosphérique très précise d'une part, au mélange des espèces d'autre part. De plus, les classes sont des indices qualitatifs, le

passage d'une classe à une autre est brusque et discontinue.

Ce travail constitue une contribution à la caractérisation des phénomènes qui régissent la variabilité l'upwelling de la zone Sénégal-Mauritanien. Il s'inscrit plus largement dans une dynamique d'étude globale du Système de Courant de Canaris (SCC) et des upwellings de bord Est. Du fait de leur importance en termes de ressources économiques pour les pays riverains, ces systèmes côtiers sont au cœur de chantiers de recherche de plusieurs laboratoires (recherches sur les systèmes d'upwellings d'Afrique de l'Ouest). Ces recherches visent à mieux comprendre la variabilité climatique et océanique de ces régions clés ainsi que leur influence sur les ressources vivantes. Le lien entre la variabilité physique du milieu océanique et la gestion de ces ressources vivantes n'est pas direct. Cependant la connaissance de cette variabilité est la première étape nécessaire à la compréhension de la chaîne de mécanismes chimiques et biologiques qui mènent jusqu'aux ressources de la pêche, sans oublier le rôle de l'exploitation par l'homme, qui est peut être le plus mal connu.

Ce travail pourrait être poursuivi en appliquant le modèle de normalisation proposé sur les données de modèles biogéochimiques (les sorties de NEMO-PISCES par exemple) couplés à des modèles de transfert radiatifs ou de mesures in situ de la zone d'étude pour comparer les résultats des modèles sur les jeux de données satellites et in situ et arriver à valider ainsi les modèles.

Si cette thèse a pu contribuer de façon même très minime, alors elle aura rempli son but fixé.

Annexe A

Articles JSPI

**Variabilité spatio-temporelle de l'upwelling Sénégal-Mauritanien
à partir d'observations satellitaires (pp. 60 - 71) ; ISSN : 0851-4453**

O. Farikou, S. Sawadogo, A. Niang, J. Brajard, M. Crépon, S. Thiria

Variabilité spatio-temporelle de l'upwelling Sénégal-Mauritanien à partir d'observations satellitaires

O. Farikou^{1,3}, S. Sawadogo², A. Niang³, J. Brajard⁴, M. Crépon⁴, S. Thiria⁴

¹Institut Universitaire des Sciences et Techniques d'Abéché (IUSTA), BP 6077 N'Djamèna-Tchad

²Laboratoire de Traitement de l'Information (LTI) – EPT, BP A10 Thiès, Sénégal

³Laboratoire de Traitement de l'Information (LTI) – ESP- UCAD, BP : 5085 Dakar-Fann (Sénégal)

⁴Institut Pierre Simone Laplace/Laboratoire d'Océanographie et Climat : Expérimentations Approches Numériques (IPSL/LOCEAN), Université. Paris 6, 75252 Paris, France

E-mail : farikou@yahoo.fr, s_sawadogo@yahoo.fr, awaniang@ucad.sn, Julien.Brajard@locean-ipsl.upmc.fr

RÉSUMÉ: L'upwelling Sénégal-Mauritanien est un upwelling côtier très productif situé le long de la côte ouest Africaine. Sa variabilité saisonnière et inter annuelle au sud de 20°N est analysée de 1998 à 2007 à partir d'images satellitaires de couleur de l'océan et de température de surface. Pour cela on utilise une méthode de classification neuronale (SOM) suivie d'une classification hiérarchique ascendante. On montre que six classes suffisent à décrire l'upwelling et à en expliquer la variabilité. Son extension est maximum en Janvier, Février, Mars, minimum en Aout, Septembre. La variabilité saisonnière est liée à celle du vent. Les six classes peuvent être considérées comme des indices statistiques permettant d'analyser la variabilité spatio-temporelle de l'upwelling Sénégal-Mauritanien..

ABSTRACT: The Senegalo-Mauritanian upwelling is a highly productive coastal upwelling located along the West African coast. Seasonal variability and inter-annual south of 20 ° N is analyzed from 1998 to 2007 from satellite images of ocean color and sea surface temperature. For this we use a classifying neuronal method (SOM) followed by an ascending hierarchical classification. We show that six classes are sufficient to describe the upwelling and to explain the variability. Its extension is maximum in January, February, March, minimum in August, September. Seasonal variability is related to the wind. The six classes can be regarded as statistical indices to analyze temporal and spatial variability of the upwelling Senegalo-Mauritanian.

MOTS-CLÉS : upwelling Sénégal-Mauritanien, couleur de l'océan, température de surface, classification neuronale (SOM), variabilité spatio-temporelle

KEYWORDS: senegalo-Mauritanian upwelling, ocean color, sea surface temperature, classifying neuronal (SOM), spatio-temporal variability.

1. Introduction

Les upwellings côtiers sont des zones océaniques très productives du point de vue biologiques. Plus de 80% de la productivité océanique se passe dans les zones d'upwelling. L'upwelling Sénégal-Mauritanien est un upwelling côtier situé le long de la côte ouest Africaine. Il s'étend de 30°N à 10°N. Il a été bien étudié durant les 20 dernières années à partir de campagnes à la mer puis en utilisant les observations satellitaires [7], [15], [5], [9], [14] et plus récemment à l'aide de modèles numériques [14]. L'upwelling Sénégal-Mauritanien fait partie des grands systèmes d'upwelling côtier situés le long des bords ouest des continents (Californie, Pérou-Chili, Benguela). Ces upwellings sont dus aux alizées

dont la composante orientée vers l'équateur génère un transport océanique de surface perpendiculaire à la côte et dirigé vers le large qui est compensé par une remonté des eaux sous jacentes qui sont froides et riches en éléments nutritifs favorisant le développement du phytoplancton. Les upwellings ont donc une forte signature de surface ce qui permet de les étudier à l'aide de capteurs satellitaires. L'analyse des températures de surface et des concentrations en chlorophylle montre que l'upwelling Sénégal-Mauritanien peut être divisé en deux régions de comportement différent [14]: l'une au nord de 21°N (figure 1) où les variations saisonnières sont faibles et où l'influence de l'upwelling peut s'étendre loin de la côte (quelques centaines de km au

niveau du cap Blanc, [7]) en présentant des structures filamenteuses et tourbillonnaires s'étalant vers le large

L'autre au sud de 21°N où l'upwelling présente une forte variabilité saisonnière ; son intensité est maximum en hiver et minimum en été lors que les alizées disparaissent avec la remontée vers le Nord de l'ITCZ (InterTropical Convergence Zone) et l'arrivée de la saison des pluies. La phase d'upwelling intense qui est caractérisée par une bande côtière étroite (de l'ordre de quelques dizaines de km) où la température est plus froide de plusieurs degrés que la température au large prend place pendant les mois de Décembre à Avril. Elle est associée à des eaux particulièrement riches en phytoplancton.

Le développement de l'imagerie satellitaire et le libre accès aux données permettent aujourd'hui l'utilisation des données de télédétection comme outil efficace pour la caractérisation des phénomènes océaniques à grande échelle spatio-temporelle. Cette étude s'appuie donc sur l'imagerie satellitaire.

Le but de l'étude que nous avons entreprise est d'affiner la compréhension des phénomènes qui régissent la variabilité spatio-temporelle de cet upwelling dans la zone 9°N-21°N. Pour cela on utilise des méthodes de classifications multi-variables

2. Présentation de la zone d'étude

La zone d'étude à laquelle nous allons nous intéresser se situe sur la côte nord-ouest du continent Africain [26°W-14°W et 9°N-21°N], sur la façade atlantique (figure 1). L'intérêt premier de notre zone d'étude est d'être une zone côtière. Elle se situe donc à la frontière entre un bassin océanique, l'Atlantique Nord, et un continent aride à semi-aride, l'Afrique du Nord. La troisième composante climatique est le forçage atmosphérique. La combinaison de ces particularités initie le phénomène d'upwelling côtier, c'est-à-dire, selon une première définition, une remontée en surface d'eaux profondes et froides le long de la côte

2.1 Type de marge et topographie

Il s'agit d'une marge passive, caractérisée par une vaste plate-forme continentale dont la largeur varie entre 30 et 150 km. Entre 15°N et 30°N, la profondeur du point de rupture de pente, position entre le plateau continental et le talus, est en moyenne de 103 m, ce qui est relativement peu profond [19]. Toutefois, la plate-forme comporte de nombreuses dépressions de grande profondeur. Au delà du point de rupture de pente, le talus continental conduit rapidement à des profondeurs de l'ordre de 4000 m. C'est sur cette plate-forme continentale peu profonde que se met en place

l'upwelling côtier sous l'action du régime de vent régional [21].

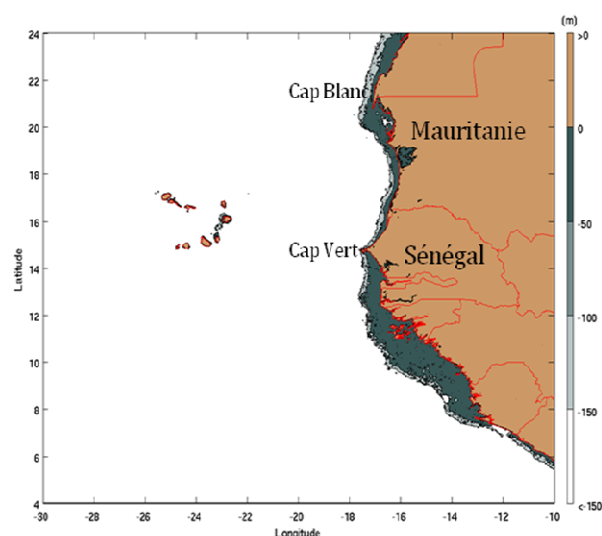


Fig.1 : Localisation et topographie de la zone d'étude. La terre est représentée en rouge et la bathymétrie en mètres selon l'échelle située à droite de la figure.

2.2 Situation climatique et hydrologique

Les vents dominants de notre zone d'étude sont essentiellement influencés par la Zone de Convergence Intertropicale (ITCZ). Il s'agit d'une zone dépressionnaire délimitant l'extension météorologique de chaque hémisphère. Deux points sont importants pour comprendre la relation entre les systèmes de vents et l'ITCZ. Tout d'abord, il s'agit d'une zone de convergence, c'est-à-dire que, à proximité de l'ITCZ, les vents vont du nord vers le sud dans la partie nord, et inversement dans la partie sud, le deuxième point est que l'ITCZ n'est pas en position fixe, mais suit un rythme de déplacement saisonnier. L'ITCZ est en position nord durant l'été boréal (figure 2a) et en position sud durant l'été austral (figure 2b). Le balancement des saisons d'un hémisphère à l'autre s'accompagne de la migration de l'ITCZ et donc du changement des régimes de vents.

Dans l'hémisphère Nord, sous l'effet de la force de Coriolis, les mouvements de fluide sont déviés sur leur droite (et sur leur gauche dans l'hémisphère sud). Au nord de l'ITCZ, le mouvement des vents convergents, d'orientation nord - sud, est alors dévié vers l'ouest, donnant ainsi naissance aux alizés (figure 2).

Il faut noter de suite que l'intensité des vents et leur orientation vont être fonction de leur position par rapport à l'ITCZ (en l'absence d'éventuelle influence orographique).

Ainsi, lorsque l'ITCZ va migrer vers le sud, un point quelconque de la côte va subir la migration des conditions climatiques. Toujours en considérant cette position nord de l'ITCZ, les vents de la partie sud ont un comportement différent (figure 2a et 3c). En effet, étant donné qu'ils convergent vers l'ITCZ, les vents sont orientés du sud vers le nord alors qu'ils sont dans l'hémisphère nord. Ils sont donc déviés vers l'est et forment le phénomène de mousson du sud-ouest. Au sud de l'ITCZ, le courant côtier est alors dirigé du sud vers le nord et rencontre le courant des Canaries (de sens opposé) aux environs du Cap Blanc (figure 3c et 3d).

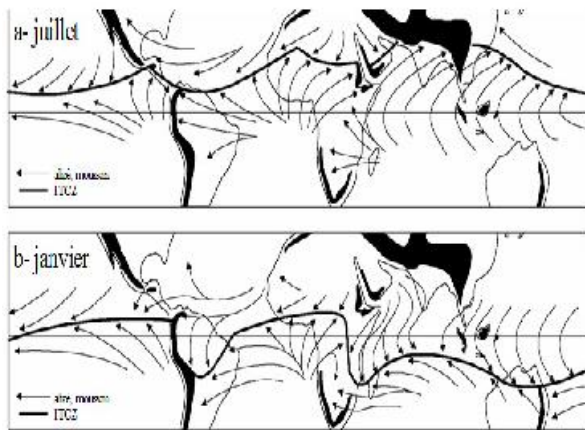


Fig.2 : Circulation tropicale du vent en surface : (a) en juillet (b) en janvier. (d'après [20])

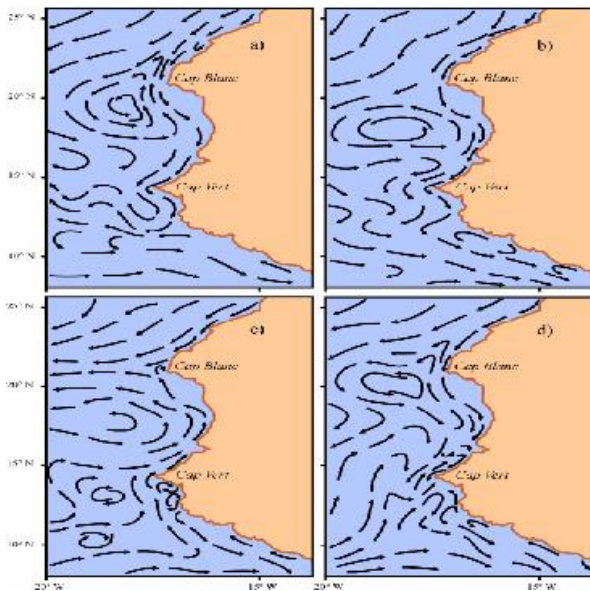


Fig.3 : Schéma de circulation de surface pour la côte Nord-Ouest Africaine : (a) en hiver boréal, (b) au printemps, (c) en été et (d) en automne. (D'après [15])

3 Données d'études

La signature de l'upwelling est bien marquée sur les observations satellitaires de température de surface (SST) et de couleur de l'océan marquant la teneur en phytoplancton, ce qui permet de suivre leur variabilité spatio-temporelle.

Due à la nécessité de moyenner les images satellitaires sur des périodes suffisamment longues pour filtrer la couverture nuageuse qui est importante pendant la saison pluvieuse d'été (Juillet, Août, Septembre), nous avons considéré dans notre étude les images mensuelles uniquement. A cette échelle temporelle, la SST et la concentration en chlorophylle (dénotee Chl-a dans ce qui suit) mesurés par satellite sont des indicateurs de l'activité de l'upwelling, la SST reflétant les caractéristiques dynamiques de l'upwelling, tandis que la Chl-a étant une signature de l'activité biologique et aussi de la dynamique via l'apport en sels nutritifs.

Deux bases de données ont été utilisées pour réaliser cette étude, l'une concerne la température de surface de l'océan, et l'autre les données couleur de l'océan.

3.1 Température de surface (SST)

Les données de SST sont issues du capteur NOAA/AVHRR et sont récupérées via le site web (<http://las.pfeg.noaa.gov/OceanWatch/>). La résolution spatiale est : 0.1 degré de latitude et la résolution temporelle est mensuelle. La précision est de l'ordre de 0.1°C. La période analysée va de 1998 à 2007, le type de donnée récupérée est codé en ASCII. Chaque mois de donnée est représenté par une matrice de dimension 121x121 donnant les températures SST_{xy} aux Pixels P_{xy} (les indices x et y représentant les coordonnées géographiques : x = latitude, y = longitude). Chaque Pixel P_{xy} correspond à un point géographique bien précis qui est le même pour chaque P_{xy} de chaque matrice. Pour une année on aura donc 12 matrices (janvier à décembre). Le nombre total de matrice de données pour ces 10 ans est 12x10

3.2 Indice de température (ISST)

Puisque l'upwelling se manifeste par un gradient de température entre la côte et le large, une idée simple est de le caractériser par un indice de SST (dénotee ISST dans ce qui suit [9]). Cet indice est défini comme la différence, à une latitude donnée, entre la SST moyenne dans la bande 22° - 26°W et la température en un point donné situé entre la côte et ladite bande. Il est calculé pour chaque mois de l'année en faisant la différence entre la moyenne de la SST dans la bande (22° et 24°W) au large (valeur dénotée SST_y) et la SST_{xy} en un pixel de la côte.

$$ISST_{xy} = SST_{xy} - SSTW_y$$

Cette procédure à l'avantage de filtrer les variations saisonnières de température de surface qui peuvent être importante et celles dues à des phénomènes de très grande échelle.

3.3 Données couleur de l'océan

Les données chlorophylle sont issues du capteur SEAWIFS et sont recueillies sur le site web : (<http://reason.gsfc.nasa.gov/OPS/Giovanni/ocean.seawifs.shtml>).

Etant donné ses variations importantes et sa distribution statistique (beaucoup de petites valeurs et peu de grandes) Chl-a est caractérisée par son Log10, quantité fréquemment utilisée par les bio-physiciens.

Ces données ont une structure identique à celle des SST en terme de résolution spatiale et temporelle. Le nombre total de données matricielles est également le même.

3.4 Détermination de la climatologie de la base de données.

Etant donné que les images couvrant la zone d'étude sont bruitées (nuages), nous avons utilisé au cours de ce travail des données moyennées appelées climatologie.

Afin d'explicitier la procédure de calcul de la moyenne, nous avons pris comme exemple le mois de janvier de chaque année de la période étudiée de la variable SST (1998-2007). Pour calculer la climatologie SST du mois de janvier, on considère l'ensemble des mois de janvier de la période (janvier_1998 à janvier_2007); chaque matrice de ces données est constituée à la fois des pixels valides (ayant une valeur SST) et pixels non valides (nuages ou la terre). Le calcul de la moyenne des températures sur les dix années à chaque Pixel se fait en considérant chaque pixel P_{xy} valide du mois correspondant; les 10 SST_{xy} associé au pixel P_{xy} des 10 matrices sont moyennées de (janvier_1998 à janvier_2007); la moyenne des SST_{xy} est calculée en considérant uniquement les températures valides; après cette opération on obtient une température moyenne **MSST** dénommée climatologie de janvier. La procédure est répétée pour les autres mois (fevrier_1998 à fevrier_2007 jusqu'à decembre_1998 à decembre_2007).

A la fin, nous obtenons 12 matrices (climato_janvier à climato_decembre). De ce fait, les images de la climatologie sont lissées par cette démarche. Un traitement similaire est fait sur les données de la concentration en chlorophylle-a, ainsi que sur les données index de température pour obtenir les climatologies mensuelles **MChl-a** et **MISST**.

Les pixels des bases de données que nous utilisons ont une superficie d'environ 11.1x11.1 km de telle sorte que chaque image analysée est composée de 121x121 pixels. Chaque pixel de la zone d'upwelling est donc caractérisé par un vecteur à trois dimensions dont les composantes sont [Log10 (Chl-a), ISST et SST].

Pour résumer, on dispose de 2 types de fichiers de données :

- le fichier F_1 composé de SST, ISST et Chl-a qui comprennent chacun 121x121x12 x10 valeurs. Ce sont les fichiers que l'on va analyser. Ils sont constitués de valeurs mensuelles s'étalant sur les 10 ans d'observation.
- le Fichier F_2 composé de **MSST**, **MISST** et **MChl-a** qui comprennent chacun 121x121x12 valeurs.

4 Méthodologie

Etant donné le grand nombre de vecteur de la base de données décrivant l'upwelling et la tridimensionnalité de ces vecteurs, nous avons choisi d'utiliser une méthode de classification capable de faire ressortir les principales caractéristiques de l'upwelling. Le but est de trouver un certain nombre de vecteurs caractéristiques correspondant à des situations 'type' résumant les phases principales de l'upwelling.

Pour cela nous avons utilisé une méthode de classification neuronale appelée carte auto-organisatrice "Self Organizing Map – SOM en anglais" qui a été introduite pour la première fois par [11]. SOM compresse l'information contenue dans une grande base de données en un nombre réduit de vecteurs référents (**rvs**). Chaque vecteur référent **rv** représente un ensemble de vecteur de **F** qui ont des caractéristiques similaires. Ces vecteurs référents sont statistiquement représentatifs de différentes classes de vecteurs de la base de données **F**. Dans le cas qui nous concerne, nous avons choisi de déterminer les **rvs** en minimisant la distance Euclidienne entre **rv** et les vecteurs qui lui sont associés. Typiquement SOM est composé d'un treillis de neurones le plus souvent bi-dimensionnel. Ce treillis est composé de deux couches de neurones (figure 4). Chaque neurone de la couche supérieure du treillis est associé à un vecteur référent **rv**, et est relié à une couche d'entrée dont les neurones représentent les composantes du vecteur de **D** (ensemble de données d'entrées à résumer) que l'on cherche à classer. Les différents neurones du treillis sont connectés entre eux et déterminent une relation topologique entre les **rvs**. Les **rvs** qui ont même dimension que les vecteurs de **F** et caractérisent des groupes de vecteurs de **F** qui sont reliées entre eux par la relation topologique.

Dans l'étude que nous avons faite, nous avons utilisé une carte topologique ayant un grand nombre de neurones (20x20) correspondant à 400 groupes, permettant une partition fine des différentes situations océaniques liées à l'upwelling. Nous avons utilisé la version de SOM chargée sur le site web :

<http://www.cis.hut.fi/projects/somtoolbox/download/>.

classifieur de Kohonen + classification hiérarchique

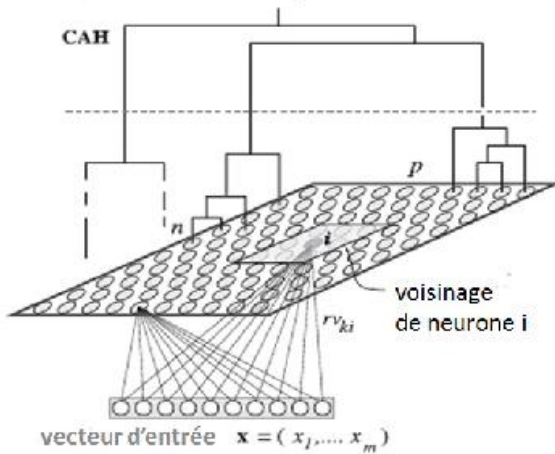


Fig.4 : Structure de la carte auto-organisatrice (SOM).

Afin d'avoir une base de donnée peu bruitée et d'en réduire la dimension, SOM a été appris à partir de la climatologie mensuelle F_2 selon la procédure décrite dans [3]. Les vecteurs de la base d'apprentissage sont des vecteurs à trois dimensions dont les composantes sont les moyennes mensuelles MISST, MSST et MChl-a au pixel P_{xy} . Afin de réduire la dimension de la base d'apprentissage, on échantillonne en prenant une ligne sur 3.

Si le nombre élevé de classe nous permet de prendre en compte toute la complexité de l'ensemble, il s'avère être un obstacle pour extraire des interprétations géophysiques simples des phénomènes liés aux données. Pour pallier cela nous avons agrégé des groupes ayant des propriétés communes en un nombre limité de classes. Nous avons utilisé pour cela une classification hiérarchique ascendante (CAH) selon la procédure décrite dans [10]. Cette procédure est itérative : à chaque itération, elle agrège deux groupes similaires selon le critère de Ward.

Dans le cas présent nous avons stoppé le processus à six classes. Les 20x20 neurones sont donc agrégés en 6 classes significatives.

La carte SOM ainsi obtenue est montrée sur la figure 5 (carte des variables et la carte montrant la décomposition en 6 classes). On remarque que les différents groupes sont statistiquement homogènes (ils comprennent à peu près le même nombre d'éléments de F_2) et que les six classes sont

contiguës sur le graphe représentant SOM, montrant ainsi que la relation topologique sur SOM fonctionne bien.

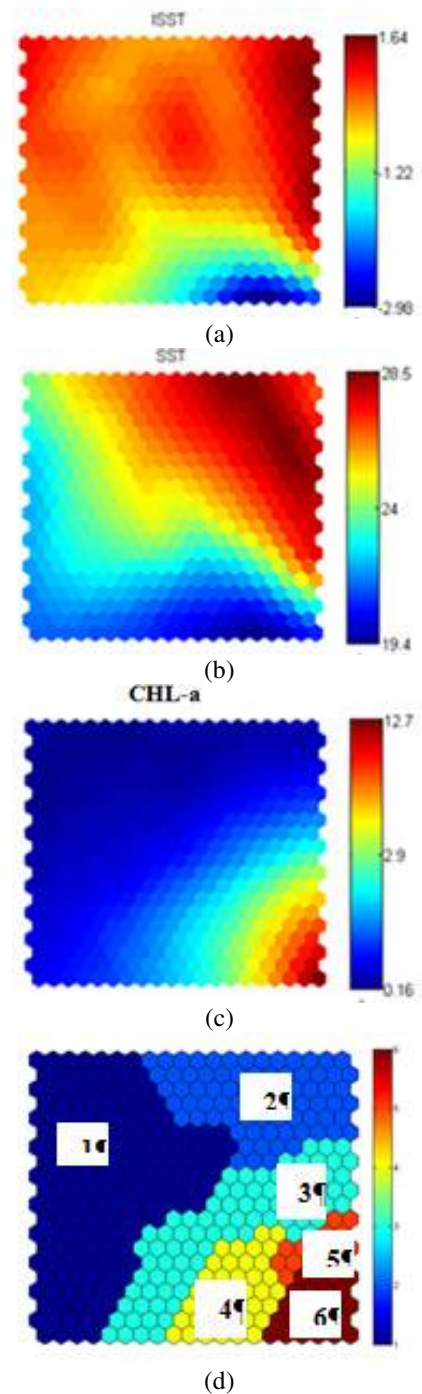


Fig.5 : Carte des variables, et carte montrant la décomposition en 6 classes après CAH

5 Résultats et Discussions

5.1 Analyse des six classes de SOM

Dans ce qui suit, nous analysons le comportement océanique de chacune des six classes. Chaque neurone de SOM a capturé un certain nombre de pixels de F_2 . Nous remarquons qu'il existe une bonne adéquation entre les classes, la concentration en Chl-a et les températures, montrant ainsi la pertinence de la décomposition en classes. Les chlorophylles les plus fortes et les températures les plus faibles sont captées par la classe 6 (upwelling) tandis que les chlorophylles les plus faibles et les températures les plus fortes (eaux du large) sont captées par les classes 1 et 2 (figure 5c).

Projetons les pixels associés à chaque classe sur une carte géographique (figure 6).

Nous remarquons que les six classes sont bien définies et correspondent chacune à une zone géographique bien délimitée montrant la cohérence de la classification. Cette cohérence géographique des pixels des classes qui est forte pour les données annuelles est aussi importante lorsque l'on projette les données mensuelles de chaque classe sur la carte géographique ce qui nous permet d'analyser le comportement saisonnier de l'upwelling. On remarque que la variabilité mensuelle des classes 4, 5 et 6 est très importante. Décrivons d'abord les caractéristiques des six classes.

La classe 1 (bleue foncée) correspond à des eaux où la concentration en Chl-a est très faible et la ISST voisine

de zéro. Elle est associée aux eaux du large. Elle occupe une large partie de la carte SOM et est géographiquement très étendue

Les classes suivantes sont influencées par l'upwelling à des degrés divers. Elles occupent chacune une petite partie de la carte SOM.

La classe 2 (bleu pale) est associée à des eaux chaudes pauvres en Chl-a et situées dans le sud de notre zone d'observation provenant probablement du contre-courant équatorial. Durant les mois de Juin, Juillet, Aout, Septembre, les eaux de la classe 2 envahissent les côtes Sénégalaises de 13°N à 17°N remplaçant ainsi les eaux des classes 6, 4 et 3, ce qui est en accord avec le schéma esquissé par [9]

La classe 3 et la classe 4 correspondent respectivement à des eaux dont la concentration en Chl-a va en croissant et dont les ISST deviennent respectivement un peu plus négatives.

La classe 5 est atypique. Elle est caractérisée par des eaux riches en chlorophylle et de ISST et de SST chaude. On la trouve surtout sur le plateau de la petite côte et sur le banc d'Arguin. Elle présente une grande variabilité saisonnière.

La classe 6 correspond à des eaux qui ont des très hautes concentrations en Chl-a et des ISST et SST très froides. Elle forme une bande étroite située le long de la côte et présente une grande variabilité saisonnière. Elle caractérise les eaux de l'upwelling côtier.

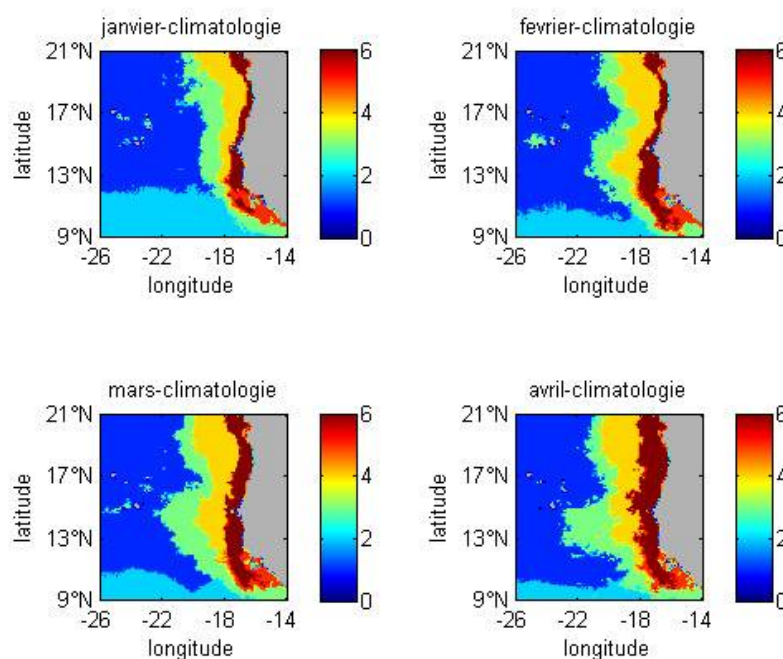


Fig.6a : Répartition mensuelle des six classes de janvier à avril

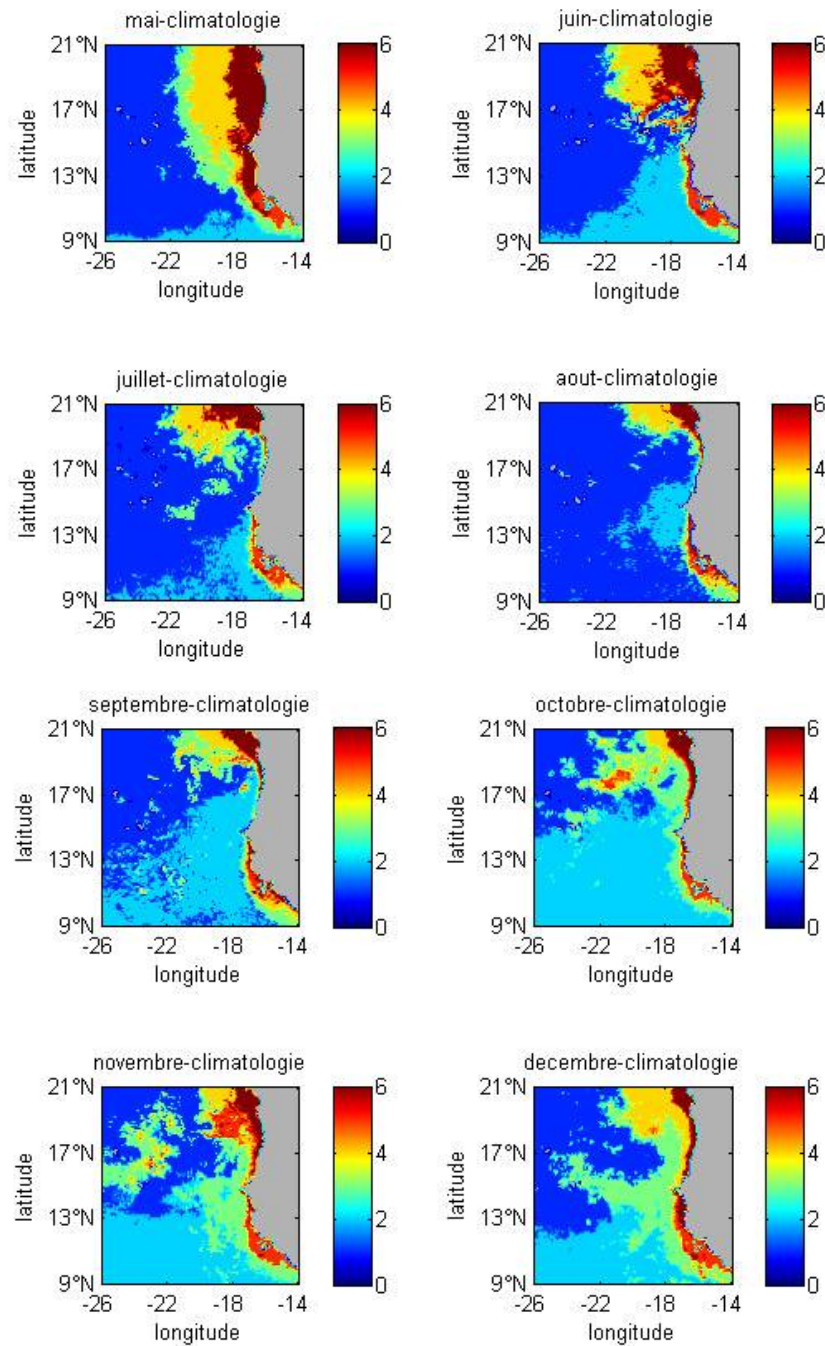


Fig.6b : Répartition mensuelle des six classes de mai à décembre

Les caractéristiques en ISST et en concentration de Chl-a pour les différentes classes sont données sous forme graphique par les “box plots” (figure 7) qui représentent la médiane des valeurs (trait plein dans la boîte) ; le premier quartile (25% des valeurs) et le dernier quartile (75% des valeurs) sont donnés par les bords de la boîtes.

Les moustaches représentent 1.5 de l'espace inter quartile.

Les classes sont toujours bien individualisées en Chl-a, quelque soit la saison (figure 7) avec une augmentation de la concentration en Chl-a de la classe 1 à la classe 6 lorsque l'on va du large vers la côte. Le comportement en température est plus compliqué. On observe bien en Janvier

une décroissance en ISST de la classe 1 à la classe 6 avec un comportement atypique de la classe 5 qui présente des ISST beaucoup plus chaudes que celles de la classe 4 et la classe 6

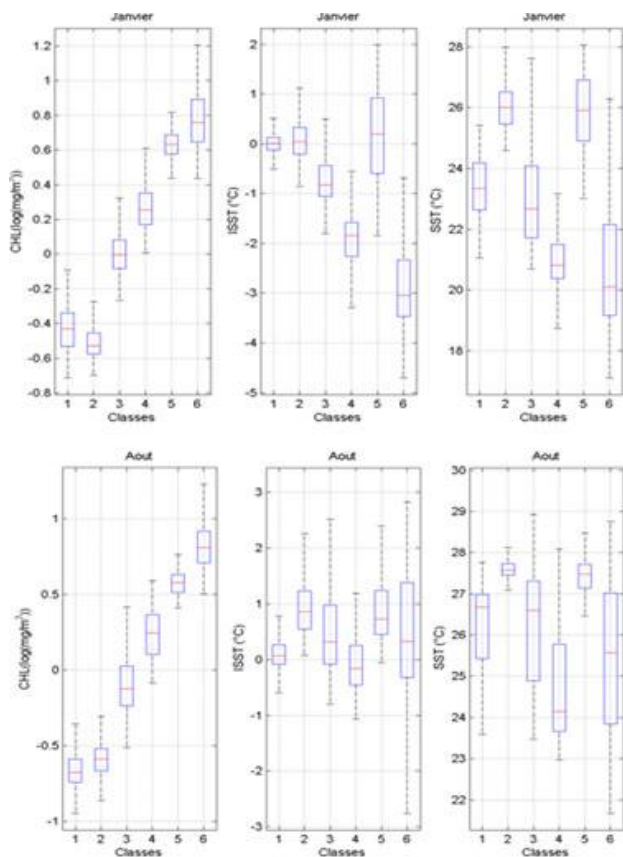


Fig.7 : Box-plots montrant les concentrations en Chl-a (en échelle log), les valeurs de ISST et de SST des différentes classes en Janvier et Aout.

5.2 Variabilité saisonnière de l'upwelling

Les figures 6 et 7 montrent une variabilité saisonnière bien marquée du comportement des traceurs océaniques associés à l'upwelling. Cette variabilité est visible sur les SST et ISST médianes des différentes classes (figure 8), beaucoup moins sur la concentration en Chl- a (figure 9) qui reste quasi constante.

Cette variabilité saisonnière bien marquée des SST et ISST des différentes classes pose la question de la pertinence de la décomposition en classes. En effet les caractéristiques hydrographiques des différentes classes présentent une certaine variabilité en ce qui concerne l'ISST et aussi de la SST. La question est alors de comprendre la signification d'une classe en terme d'hydrologie : la classe 1 est associée à des eaux dont l'ISST (figure 8) est voisine de zéro et dont la teneur en Chlorophylle est très basse (figure 9). La classe 1 regroupe les eaux du large, ceci quelque soit la saison.

Les eaux de la classe 1 sont éloignées de la côte pendant la saison d'upwelling (Décembre à Avril) et viennent la rejoindre pendant les mois d'été (Juillet, Aout, Septembre). La valeur médiane de la SST de la classe 1 (figure 8) est minimum en Avril et maximum en Octobre. Pour expliquer cela on peut incriminer l'ensoleillement qui est minimum en Hiver et maximum en été, avec un décalage temporel dû à l'inertie thermique. Cet effet est renforcé par le vent qui est maximum en hiver et donc brasse les eaux de surface jusqu'à la profondeur de la thermocline saisonnière contribuant ainsi au refroidissement des eaux de surface. L'effet inverse se passe l'été où le vent est faible favorisant la stratification et donc l'apparition d'eaux chaudes en surface. La classe 6 est associée à des eaux dont l'ISST et la SST médianes (figure 8) sont les plus négatives des 6 classe en hiver (-3° en Janvier) et voisine de zéro en été. La SST médiane est aussi la plus basse en hiver et chaude en été. La chl-a (figure 9) est la plus élevée des 6 classes et ne présente pas de variations significatives liées à la saison. La classe 6 (rouge foncé sur la figure 6) est confinée à la côte où elle forme une bande étroite (quelques dizaines de kilomètres) épousant le trait de côte. La classe 6 disparaît quasiment au Sud de 20°N en Juillet, Aout, Septembre. Il semble clair que la classe 6 représente les zones côtières où l'effet de l'upwelling est le plus intense. La très forte modulation saisonnière en température pourrait être due à l'ensoleillement d'été associé à l'atténuation des vents.

La classe 5 (rouge pale sur la figure 6) montre un comportement atypique. Des eaux riches en Chl-a et des ISST (et des SST) élevées par rapport aux classes adjacentes (6 et 4 et 3) comme cela se voit sur la figure 7. Il semble que la classe 5 caractérise, au moins durant les mois d'hiver (Décembre, Janvier, Février, Mars, Avril), les eaux situées sur des plateaux continentaux (plateau continental de la petite côte et banc d'Arguin). Ceci expliquerait les valeurs élevées de la Chl-a (eaux du cas-2) et la température (petits fonds). En Juin et Novembre on remarque des structures de classe 5 entre 18°-20°N s'étendant vers le large. Ces patterns ainsi que la bande de classe 5 trouvée en mai entre 17-19°N posent des problèmes d'interprétation. Une première explication pourrait provenir de la construction de la base de données qui moyenne des informations provenant d'années différentes, ce qui peut biaiser la continuité des patterns spatiaux. Une solution serait d'analyser ces types de structures sur chacune des années de la base de données.

Les classes 2, 3, 4 ont des ISST qui diminuent et des concentrations en Chl-a qui augmentent lorsque l'on se déplace de l'ouest (large) vers l'est. Leur structure spatiale suit approximativement celle de la classe 6 laissant supposer que les zones correspondant à ces classes sont influencées par l'upwelling via des structures de méso échelle (filaments, tourbillons).

En particulier, leur signature disparaît au Sud de 20°N en Juillet, Aout, Septembre. Nous avons aussi calculé la

variabilité saisonnière par l'extension spatiale (nombre de pixels) des classes 5 et 6 (figure 10)

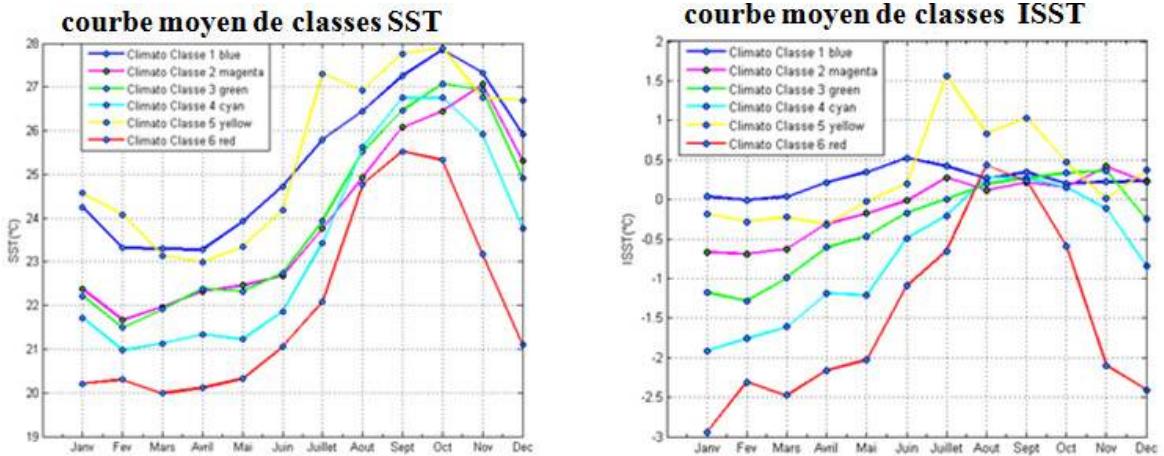


Fig.8 : Variation saisonnière des SST et ISST médianes des différentes classes.

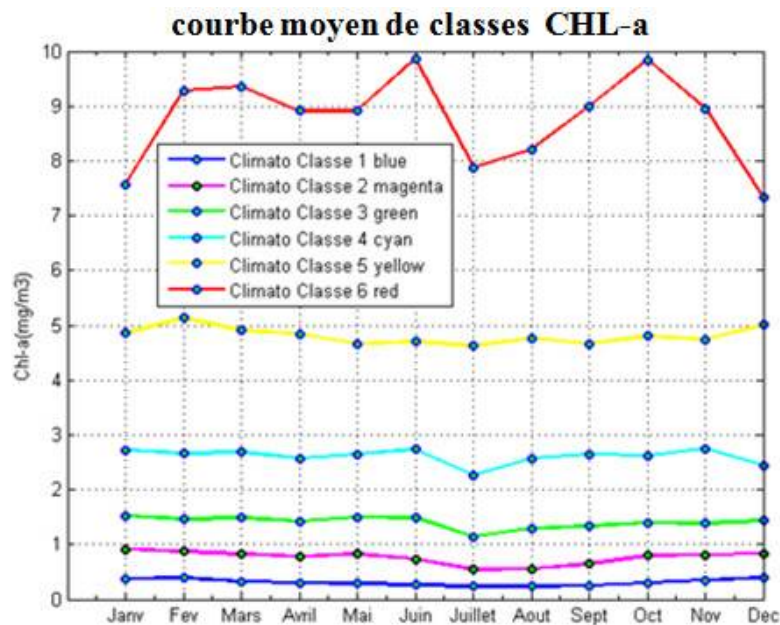


Fig.9 : Variation saisonnière des concentrations médianes en Chl-a des différentes classes

On remarque que ces deux classes présentent une forte variabilité saisonnière avec un maximum en mars, avril, mai, un second maximum en octobre novembre et un minimum pendant la saison des pluies (Juillet, Aout). Les variations saisonnières de l'upwelling sont en accord avec celles du vent. On a tracé sur la figure 11, la composante N-S du vent de 1990 à 2007 à Dakar. On

remarque la forte variabilité saisonnière du vent avec un maximum au mois de mars lorsque l'extension de la classe 6 est maximale et son ISST et SST minimales. La disparition quasi totale de la classe 6 en dessous de 20°N correspond à l'annulation du vent parallèle à la côte en Juillet-Aout. Cette valeur minimale du vent est associée avec le déplacement vers le Nord de l'ITCZ qui atteint sa latitude

maximale au mois d'aout. On ne remarque pas de variabilité inter-annuelle caractéristique du vent, à part une diminution sensible du pic hivernal en 1997

La variabilité saisonnière de la classe 6 peut s'expliquer selon le scénario suivant : les forts vents d'hiver génèrent un upwelling intense le long de la côte de 21° à 15°N

associé au développement de la classe 6. L'affaiblissement du vent en été (Juin, Juillet, Aout) conduit à la disparition de l'upwelling et donc de la classe 6 de 15°N jusque 20°N. La classe 6 est alors remplacée par la classe 1 qui s'étend maintenant du large jusqu'à la côte.

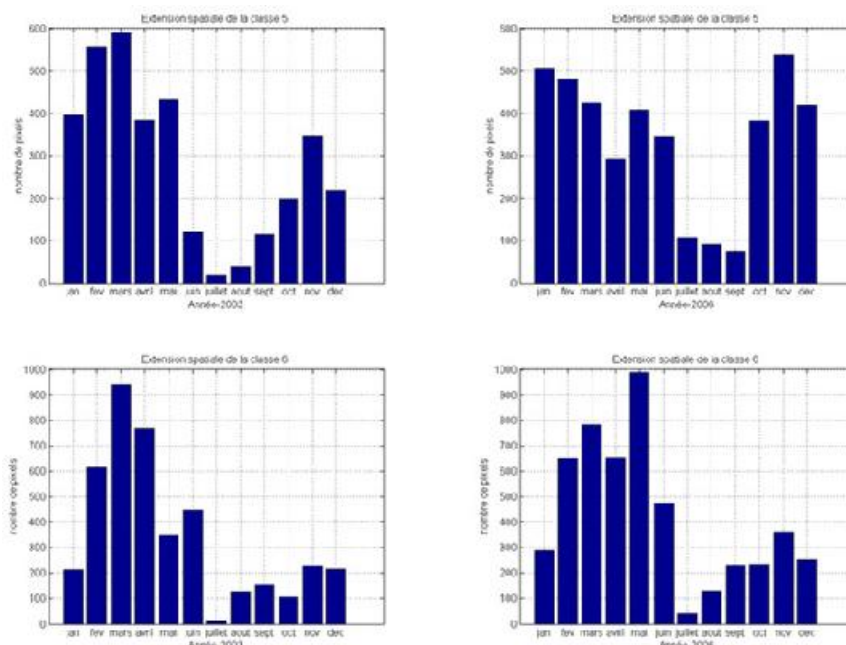


Fig.10 : Extension spatiale des classes 5 et 6 pour les années 2003 et 2006.

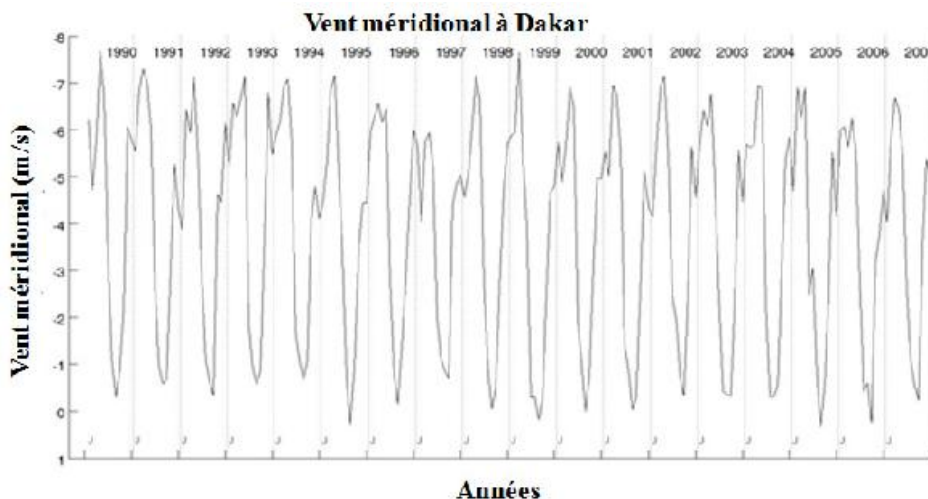


Fig.11 : Composante N-S du vent à Dakar. On remarque la forte variabilité Saisonnière avec un maximum au mois de mars.

6 Discussions

La classification multi-variables par carte topologique SOM des eaux de surface de la région de l'Upwelling Sénégal-Mauritanien entre 9°N et 21°N, nous a permis de faire une description de l'upwelling en terme de variabilité spatio-temporelle. En utilisant 3 variables accessibles par mesures satellitaires (ISST, SST et Chl-a), nous avons décomposé les situations hydrologiques de surface en 6 classes. Ces classes forment des ensembles statistiquement cohérents dans le sens où elles représentent des neurones contigus sur les cartes SOM (donc ayant des propriétés statistiques communes) et où elles sont aussi géographiquement continues formant des structures spatiales bien homogènes. On peut les considérer comme des indices multi-variés permettant de caractériser les différentes masses d'eaux. La classe 6 de SOM est associée aux eaux froides et riches en phytoplancton de l'upwelling côtier. Ces eaux longent la côte en formant une bande de quelques dizaines de kilomètres de large. Elles sont bien définies en hiver (Décembre-Avril) où on les retrouve jusque vers 13°N. Pendant les mois de Juillet, Aout, Septembre, ces eaux disparaissent au sud de 20°N et sont remplacées par des eaux du large qui en utilisant les résultats de SOM laissent à penser qu'elles proviennent du Sud et pourraient être dues à une remontée du Contre-Courant Equatorial provoquée par la remontée vers le Nord de l'ITCZ et la quasi-disparition des alizées. Les classes 4, 3, 2 ont un comportement similaire à celui de la classe 6, en accord avec le fait que les paramètres océaniques de ces classes seraient liés à une diffusion vers le large des paramètres de la classe 6.

La classe 5 est définie par des eaux qui sont chaudes et riches en chlorophylle. Elle est caractéristique des eaux situées sur le plateau continental de la petite côte où on les trouve en permanence. Elles apparaissent parfois au large (quelques dizaines de kilomètres des côtes) entre 18°-20°N. Ces structures ainsi que la bande de classe 5 trouvée en mai entre 17-19°N posent des problèmes d'interprétation. Une première explication pourrait provenir de la construction de la base de données qui moyenne des informations provenant d'années différentes, ce qui peut biaiser la continuité des patterns spatiaux. Une étude intéressante serait aussi d'analyser les spectres des pw (réflectances océaniques) de ces eaux. Selon notre interprétation les spectres des pw de la classe 5 devraient être différents sur le plateau continental de la petite côte où ils devraient être liés à des eaux du cas-2 et au large entre 16°N et 19°N où ils devraient plus refléter celui des eaux du cas-1

7 Conclusion

Cette étude a permis de caractériser l'upwelling côtier Sénégal-Mauritanien et décrire sa variabilité saisonnière à partir des observations satellitaires de SST et de couleur de l'océan. On peut dire aussi que la décomposition en classes nous a permis de mieux comprendre la dynamique de l'upwelling Sénégal-Mauritanien, en particulier sa variabilité spatiale en fonction des saisons. On a mis en évidence différentes zones aux caractéristiques bien définies dont l'étude permet de mieux saisir le comportement de cet upwelling. La zone étudiée présente une très forte variabilité saisonnière qui est due à la fois au vent et à l'ensoleillement. Le vent est responsable de la variabilité saisonnière de l'upwelling, en particulier de la position des classes 6, 4, 3 et 2. La disparition de ces classes au sud de 20°N est due à l'arrivée de l'ITCZ et de la disparition des alizées.

Il semble d'après cette étude qu'il pourrait exister des eaux du cas-2 associées à la classe 5 principalement et situées sur le plateau continental de la petite côte. Ceci pourrait être analysé en considérant les spectres de réflectances marines pw associés à chaque pixel.

Bibliographie

- [1] J.S. Allen, (1973) Upwelling and Coastal Jets in a Continuously Stratified Ocean - J. Phys. Oceanogr., p. 245-257.
- [2] J Aristegui, Alvarez-Salgado X.A., E.D. Barton, F.G. Figueiras, S. Hernandez-Leon, C. Roy and A.M.P. Santos (2004) Oceanography and fisheries of the Canary current/Iberian Region of the eastern North Atlantic, Chapter 23 in *The Sea*, Vol. 14, edited by A.R. Robinson and K.H. Brink, p. 877-931, John Wiley and Sons, Inc., New York.
- [3] F. Badran, M. Yacoub and S. Thiria (2005) Self-organizing Maps and Unsupervised Classification, Chapter 7 in *Neural Networks, Methodology and Applications*, edited by G. Dreyfus, Springer, Berlin Heidelberg New York, p 379-442,
- [4] J.A Barth (1989) Stability of a coastal upwelling front 2. Model results and comparison with observations J. Geophys. Res, p. 10844-10856.
- [5] E.D. Barton (1998) Eastern boundary of the north atlantic : Northwest Africa and Iberia, in *The Sea*, Vol. 11, edited by A.R. Robinson and K.H. Brink, p. 633-657, John Wiley and Sons, Inc., New York.

- [6] D. Beletsky, W.P. O'Connor, D.J. Schwab, D.E. Dietrich (1997) Numerical Simulation of Internal Kelvin Waves and Coastal Upwelling Fronts- *J. Phys. Oceanogr.*, p.1197-1215
- [7] A. Bricaud, A. Morel, and J.M. André (1987) Spatial/temporal variability of algal biomass in the mauritanian upwelling zone, as estimated from CZCS data, *Adv. Space Res.*, p. 5362-5384.
- [8] M. Crepon, C. Richez and M. Chartier (1984) Effect of coastline geometry on Upwellings - *J. Phys. Oceanogr.*, p. 1365-1382.
- [9] H. Demarcq and V. Faure, (2000) Coastal upwelling and associated retention indices from satellite SST. Application to Octopus vulgaris recruitment. *Oceano. Acta*, 23 , p.391-407.
- [10] A.K. Jain, and R.C. Dubes (1998) Algorithms for clustering data. Englewood. Cliffs, New Jersey, USA: Prentice Hall Advanced Reference Series, p.320
- [11] T. Kohonen (2001) Self Organizing Maps (3rd ed.). Springer, Berlin, Heidelberg, New York, Tokyo, p .459.
- [12] T. Kohonen (1982) Self-organized formation of topologically correct feature maps, *Biological Cybernetics* 43, p. 59–69
- [13] C.B. Lange, O.E. Romero, G. Wefer, AJ Gabric (1998) Offshore influence of coastal upwelling off Mauritania, NW Africa, as recorded by diatoms in - *Deep-Sea Res. Part I*, p. 985-1013.
- [14] C. Lathuiliere, V. Echevin, and M. Levy (2008), Seasonal and intraseasonal surface chlorophyll-a variability along the northwest African coast, *J. Geophys. Res.*, 113, C05007, doi:10.1029/2007JC004433.
- [15] E. Mittelstaedt (1991) The ocean boundary along the northwest African coast: Circulation and oceanographic properties at the sea surface, *Prog. Oceanogr.*, 26, 307– 355.
- [16] A. Niang, F. Badran, C. Moulin, M. Crepon and S. Thiria (2006) Retrieval of aerosol type and optical thickness over the Mediterranean from SeaWiFS images using an automatic neural classification method. *Remote Sensing of Environment*, p. 82-94.
- [17] A. Niang, L. Gross, S. Thiria, F. Badran and C. Moulin (2003) Automatic neural classification of ocean colour reflectance spectra at the top of atmosphere with introduction of expert knowledge. *Remote Sensing of Environment*, p. 257-271.
- [18] J. J. O'Brien and H. E. Hurlburt (1972) A Numerical Model of Coastal Upwelling - *J. Phys. Oceanogr.*, p. 14-26.
- [19] E. Caroli (1999) Analyse numérique de la morphologie des marges continentales et implication de l'évolution du niveau marin sur l'immersion des plates formes continentales est-africaines, Ecole Normale Supérieure de Lyon, Bordeaux.
- [20] M. Leroux (1996) La dynamique du temps et du climat. Enseignement des Sciences de la Terre. Masson, Paris, 310 pp.
- [21] X. Giraud (2001) Reconstitution paléocéanographique du signal $\delta^{15}N$: modélisation couplée physique et biogéochimique d'un upwelling côtier, thèse de Doctorat, Université de Bordeaux 1, p. 17-29.

Annexe B

Articles RJEES

**Research Journal of Environmental and Earth Sciences 5(12) :
756-768, 2013 ; ISSN : 2041-0484 ; e-ISSN : 2041-0492**

**Multivariate Analysis of the Senegalo-Mauritanian Area by
Merging Satellite Remote Sensing Ocean Color and SST
Observations**

**O. Farikou, S. Sawadogo, A. Niang, J. Brajard, C. Mejia, M. Crépon
and S. Thiria**

Multivariate Analysis of the Senegalo-Mauritanian Area by Merging Satellite Remote Sensing Ocean Color and SST Observations

^{1,3}O. Farikou, ²S. Sawadogo, ¹A. Niang, ⁴J. Brajard, ⁴C. Mejia, ⁴M. Crépon and ⁴S. Thiria

¹Ecole Supérieure Polytechnique, Université Cheikh Anta Diop (Dakar), BP 5085 Dakar Fann, Sénégal,

²Ecole Polytechnique de Thiès, BP A10 Thiès, Sénégal

³Institut Universitaire des Sciences et Techniques d'Abéché (IUSTA), Tchad

⁴IPSL/LOCEAN, unité mixte CNRS-IRD-UPMC-MNHN, Case 100, 4 Place Jussieu, 75005 Paris France

Abstract: The Senegalo-Mauritanian upwelling is a very productive upwelling occurring along the West coast of Africa. The seasonal and inter-annual variability of the upwelling region between 9° and 22°N and 14° and 25°W was studied by merging monthly ocean color data and sea surface temperature provided by satellite sensors during twelve years from 1998 up to 2010. We combined these two parameters to obtain a unique index describing the spatio-temporal variability of the upwelling. We used a classification methodology consisting in a neural network topological map and a hierarchical ascendant classification. Six classes can explain most of the variability of this region, one of them (class 6) being dedicated to the coastal upwelling water, another being the signature of the Gulf of Guinea dome water (class 2), a third one to case 2 water (class 5). The classes can be considered as multi-factorial statistical indices allowing us to characterize the different water types of this region and to investigate their variability. It is shown that the upwelling extent is maximum in February-March, minimum in August-September. Its variability is linked to that of the wind and to the ITCZ position. The Gulf of Guinea waters moves northward in June and relaxes to their southward position in December. During the twelve years of observation, we were not able to evidence climatic trends of the SST and Chl-a concentration. The methodology we have developed can be used in a large variety of problems implying multi sensor measurements.

Keywords: Data fusion, machine learning, oceanography, phytoplankton, remote sensing

INTRODUCTION

The Senegalo-Mauritanian upwelling is a very productive coastal region occurring along the West coast of Africa. It extends from 26°N down to 10°N. It has been well documented during the past decades from *in-situ* and satellite observations. For a review of the physics (Barton, Eastern Boundary of the North Atlantic, 1998) and the biogeochemical behavior (Aristegui *et al.*, 2004). The major forcing are the trade winds blowing South-westward, which generate an offshore Ekman transport generating an upwelling of deep cold waters rich in nutrients favorable to phytoplankton development. The analysis of satellite-derived chlorophyll concentration and Sea Surface Temperature (SST) (Lathuiliere *et al.*, 2008) showed that the Senegalo-Mauritanian upwelling can be split into two regions, one north of 21°N (Fig. 1) where the seasonality is very weak, the other south of that latitude, where the upwelling presents a strong seasonality. In the south part, the upwelling intensity is maximum at the beginning of spring (March to April)

showing an offshore SST gradient with cold SST along the coast and high phytoplankton concentration (Fig. 2 and 3). The intense upwelling period is associated with a strong phytoplankton bloom, which extends far away offshore, as shown in ocean color satellite images. It weakens in June and disappears from mid July to the end of summer, period for which the offshore horizontal SST and chlorophyll concentration gradients decrease (Fig. 2 and 3). The weakening in the upwelling is due to that weakening of the trade winds linked to the Northward displacement of the Inter Tropical Convergence Zone (ITCZ) driving the rainy season. North of 21°N, there is a quasi permanent upwelling of which the corresponding chlorophyll-*a* concentration extends far offshore associated with filaments and eddies.

In fact, this schematic description is, in reality, much more complicated. The studied area is the eastern termination of the North Atlantic Tropical gyre. It is influenced at its northern boundary by the Canary current (CanC) flowing southward and the North Equatorial Current (NEC) flowing southwestward and

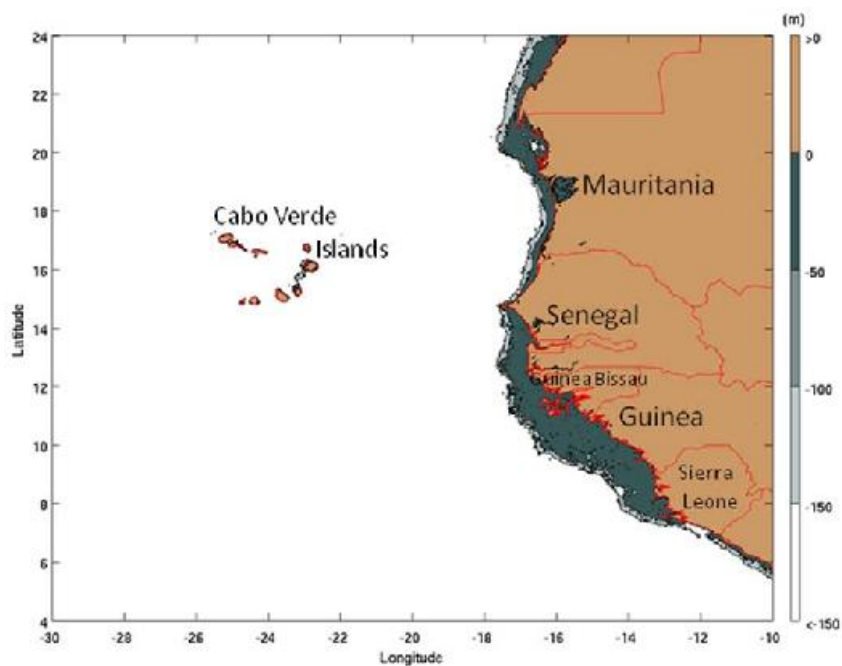


Fig. 1: Mauritania and Senegal coastal topography. The land is in brown and the ocean depth corresponds to the color scale in meters (right side of the figure)

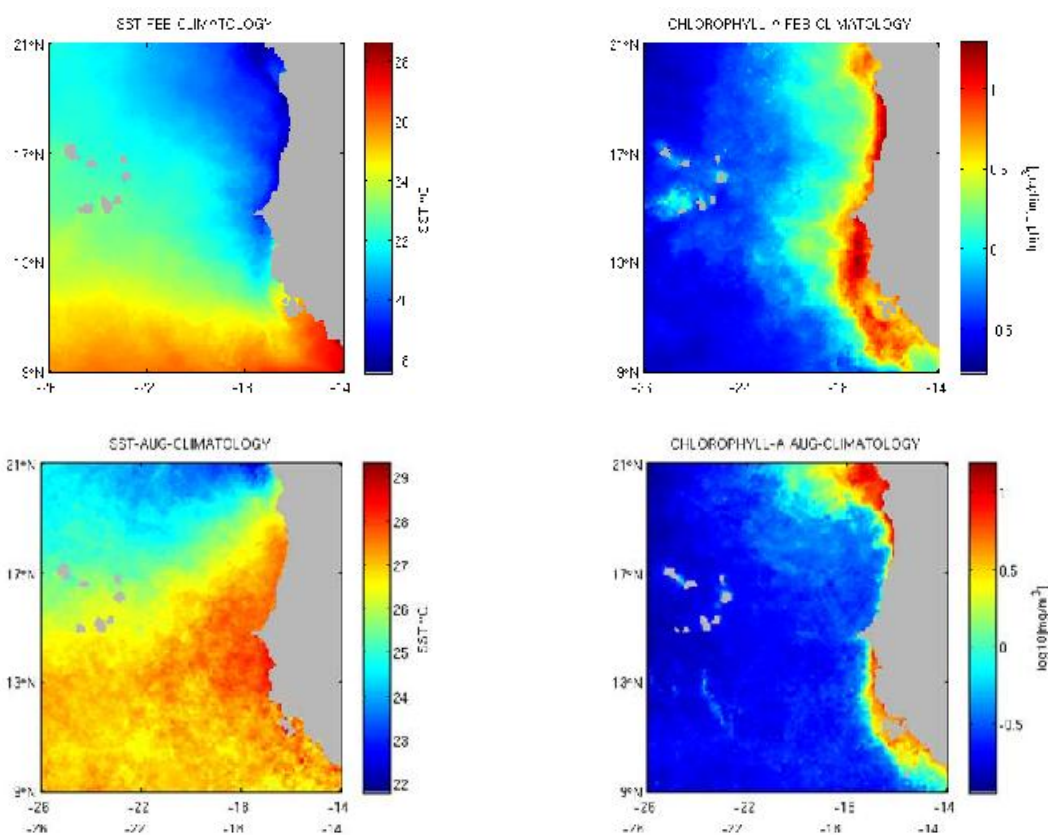


Fig. 2: Climatology of the SST (left panels) and chlorophyll-*a* concentration (right panels) for the months of February (upper panels) and August (lower panels) averaged for the ten years of observation

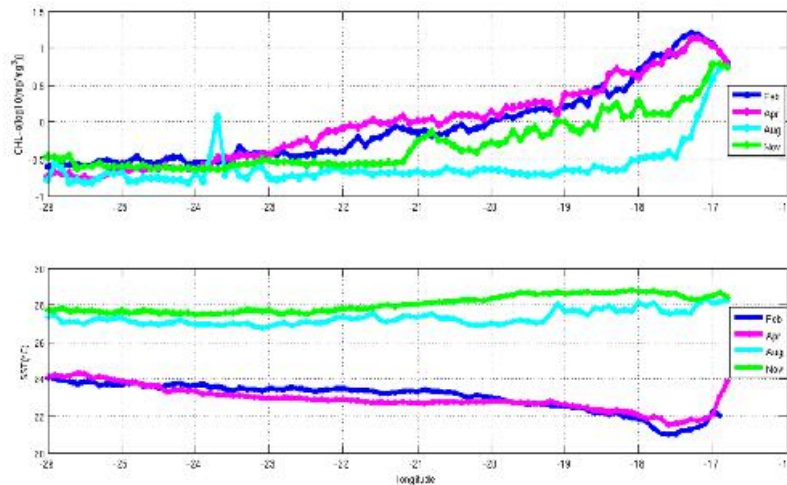


Fig. 3: SST (upper panel) and chlorophyll-*a* (lower panel) zonal sections (February, April, August and November) along 17° N

at its southern boundary by the North Equatorial Counter Current (NECC) flowing eastward. Seasonal fluctuation and meandering of these currents may modulate the upwelling region and the structures and characteristics of the associated waters (Lathuiliere *et al.*, 2008).

The aim of the present study is to combine ocean color and SST satellite observation in order to extract optimum information on the upwelling dynamics, in a much more efficient way than the information obtained by processing the two data sets separately. For that, we propose to cluster ocean situations presenting similarities, both with respect to chlorophyll concentration and to SST. For this, we chose a neural network clustering method, i.e., the Self Organizing Map (SOM) (Kohonen, 2001). SOM have been used in many geophysical studies to extract information of huge data set leading to understand complex phenomena. It has recently been applied to study the biological production in eastern boundary upwelling systems by Lachkar and Gruber (2012).

We focused our interest on monthly time scale due to the necessity to average the satellite images on a sufficient time to remove the effect of clouds, which can prevent the sea surface observation with satellite sensors, especially during the rainy season (July, August and September). At that time scale, SST and Chlorophyll-*a* concentration (Chl-*a* in the following) measured by satellite are proxies of the upwelling, SST being a signature of the dynamical behavior of the upwelling whereas Chl-*a* is a mixed signature of the dynamics and the integrated biological activity. Due to its own dynamics, Chl-*a* is able to enhance ocean physical structures not revealed by SST. Furthermore, we have processed the normalized ocean color spectra given by the SeaWiFS sensor in order to tentatively retrieve additional information such as the biogeochemical water type (case 2 water) or the phytoplankton species.

THE SENEGALO-MAURITANIAN UPWELLING

Coastal upwelling zones are very productive ocean regions. More than 80% of the ocean productivity is encountered in upwelling regions. The upwelling physical characteristics and behavior have been extensively described in the scientific literature (O'Brien and Hurlburt, 1972; Allen, 1973; Brink, 2005). Coastal upwelling is characterized by vertical motion of deep, hence cooler, water reaching the surface coastal layers. Its offshore extent is of the order of the first baroclinic radius of deformation (some tens of kilometers depending on the location and stratification). Its intensity is related to the along shore wind component which generates an offshore Ekman transport and consequently an upward vertical movement of water at the coast to satisfy the continuity equation. The upwelled water supplies nutrients to the surface layer, thus favoring the blooming of phytoplankton. The pattern of the upwelling is modulated by the alongshore variability of the wind, the coastline geometry (Crepon *et al.*, 1984; Beletsky *et al.*, 1977), the bottom topography and instabilities of the fluid motion. Schematically, one observes a strip of cold surface water with high chlorophyll concentration parallel to the coast and whose front with offshore surface waters may oscillate (Barth, 1989) and form filaments (Bricaud *et al.*, 1987; Lange *et al.*, 1998; Lathuiliere *et al.*, 2008). The signature of the upwelling is well observed on satellite Sea Surface Temperature (SST) and ocean color images, permitting the analysis of its variability.

In this research, we aimed at extracting the most pertinent information from satellite ocean color observation and SST on the Senegalo-Mauritanian upwelling in order to characterize its variability and the mechanisms driving it. For this, we analyzed 10 years (from 1998 up to 2007) of monthly sea surface chlorophyll concentration (Chl-*a* in the following) provided by the SeaWiFS (Sea-viewing Wide Field-of

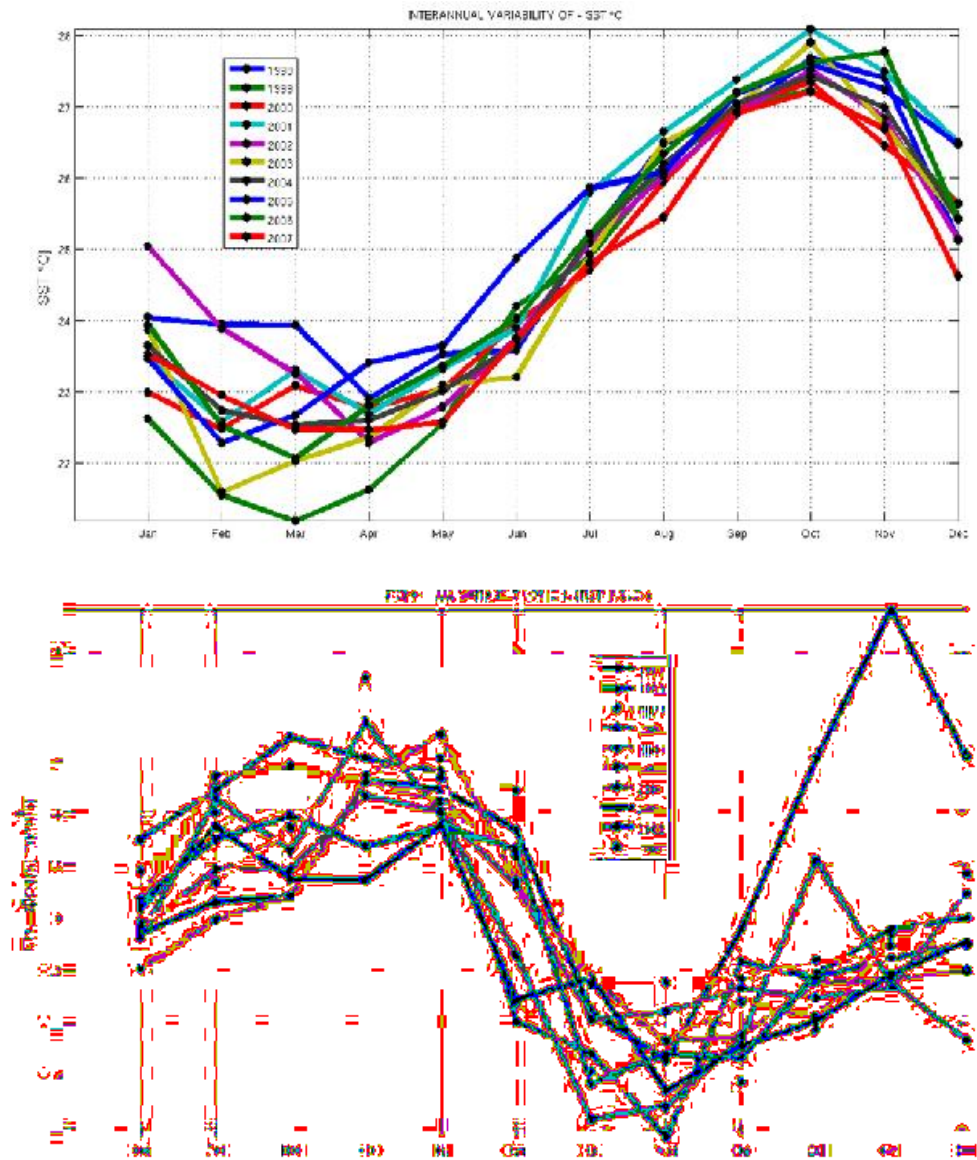


Fig. 4: Seasonal variability of the SST (upper panel) and chlorophyll-*a* concentration (lower panel) averaged over the studied zone for the different years

view Sensor) satellite radiometer and SST by the NOAA AVHRR instrument (Advanced Very High Resolution Radiometer (<http://las.pfeg.noaa.gov/Ocean Watch/>)). We propose to combine these two variables for obtaining an optimum comprehensive description of the upwelling.

Since the upwelling signature is mainly characterized by the offshore gradient of SST (Fig. 2), we thus decided to add a new variable defined as the difference at a given latitude (since the coast is approximately N-S) between the SST at an observed point and the mean SST in the longitude band 22°-26°W which is an offshore region far away from the influence of the coastal upwelling (Demarcq and Faure, 2000). This new variable, denoted ISST, filters out the

important SST seasonal variations (Fig. 4) due to air-sea interactions and large scale oceanic physical phenomena and permits the enhancement of the local dynamical processes. Due to its large variation range and its distribution (many small values and very few large values), Chl-*a* was expressed by its Log_{10} value, which is a quantity widely used by biophysicists. The images we processed extend from 9°N to 21°N and from 14°W to 26°W. As the pixel size of the data files is of the order of 11.1×11.1 km, each image is composed of 121×121 pixels. Each pixel of the upwelling is thus represented by a three dimensional vector D [Log_{10} (Chl-*a*), SST and ISST]. In order to overcome the problem of missing data due to clouds or to processing artifacts, which are major problems in

ocean remote sensing, we only considered, for the learning data set, monthly averaged images provided by the NOAA/SeaWiFs site. The learning dataset D was built up by averaging the monthly means for the 10 years of observations in order to decrease the number of missing pixels for each month. To reduce the size of the learning data set, we sampled one line out of three of the monthly satellite image matrix. The learning set D represents a mean climatology of the behavior of the upwelling for the 10 years under study. After removing the effect land and clouds, D consisted of 23893 three-dimensional vectors.

We now present the method we used.

THE METHODOLOGY

We first applied an unsupervised classification similar to that done by Niang *et al.* (2003, 2006). The aim was to summarize the information contained in the three-component vector data set D by producing a small number of reference vectors (rv) that are statistically representative of the data. Each reference vector (rv) represents a set of vectors of D that have a similar pattern. In the present case, we chose to determine the rv s by minimizing some distance between rv and the vectors of D it represents. To do this, we used a specific neural network model, the so-called topological map, which was first introduced by Kohonen (1982) and fully described in Kohonen (2001). Each neuron of the map (Fig. 5) is associated with a particular reference vector (rv) and thus corresponds to a group (in this study, a set of pixels belonging to D). The rv s approximate the density of the data set D . They are computed by minimizing a specific cost function as in the K -Means algorithm (Badran *et al.*, 2005). Besides the different neurons of the topological map C are connected together and determine a topological (neighborhood) relationship among the different groups (neurons). Close neurons on the maps correspond to rv s that are quite similar; very distant neurons correspond to rv s that are very different. The set of rv s represents the dataset D by compressing the information contained in it.

In the present study, we dealt with a two-dimensional map with quite a large number of neurons (20×20) and therefore of rv s, providing a highly discriminating representation of the observations. The pixels of the data set D are thus clustered into 400 groups. We used the SOM version available on the web site <http://www.cis.hut.fi/projects/somtoolbox/download/>. The topological map was learned according to the procedure described in Niang *et al.* (2003, 2006). The number of neurons was determined empirically from solutions of similar problems and then adjusted as described in Badran *et al.* (2005).

The large number of groups allowed us to take into account the complexity of the dataset but may have prevented us from synthesizing some geophysical information embedded in the data, such as spatial or seasonal specificities. To counteract this difficulty, we

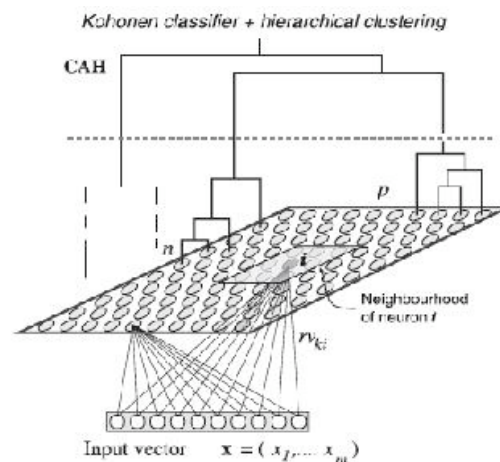


Fig. 5: Structure of the Self-Organizing Map (SOM). The network comprises two layers: an input layer used to present observations and an adaptation layer for which a neighbourhood system is defined. Each neuron i is fully connected to the input layer. It is associated with a group that is represented by a reference vector rv_i . In the figure, the neurons are cluster in 3 classes by HAC

decided to aggregate this large number of groups into a smaller number of classes based on the similarities of the groups. We therefore extracted a few pertinent classes from the groups by clustering groups having similar statistical properties, expecting that the classes could be associated with geophysical characteristics. For that we used a hierarchical ascendant classification (HAC in the following), which is a bottom-up hierarchical classification (Jain and Dubes, 1998). This method iteratively computes a partition hierarchy (Badran *et al.*, 2005). From the initial partition (the neurons on the map), two subsets of the computed partition are clustered at each iteration. These two subsets are selected by measuring their similarity according to the Ward criterion (Fig. 5).

We aggregated the 20×20 neurons into six significant classes. The resulting clustering of the three dimension vectors rv s associated with the neurons of the topological map is given in Fig. 6. We note that the topological map+HAC clustering is very coherent, since the classes represent clusters whose neurons are contiguous on the topological map. Moreover, the geophysical parameters (SST, ISST and chl- a) associated with each neuron (or rv) determine homogeneous fields on the SOM. Their gradients are smooth, well defined without any discontinuity. As an example, the chlorophyll- a concentration is maximum in the bottom right corner, minimum in the upper left corner of the SOM, the SST is minimum at the bottom, maximum at the top of the SOM. The number of groups (six) was selected because it presented the most significant discriminative partition with respect to the full dendrogram of the HAC (not shown) one the one hand and to the upwelling parameters, on the other hand.

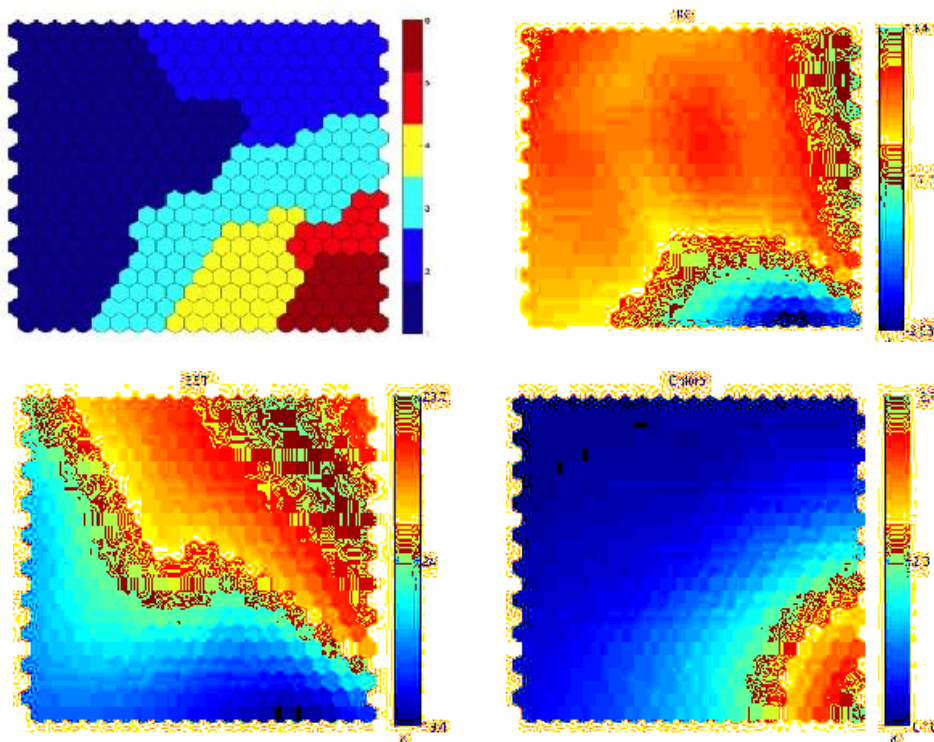


Fig. 6: Upper left panel: Representation of the SOM map clustering in six different classes. Bottom left panel: projection on SOM of SST; the coldest temperatures are in the bottom and highest ones in the upper right side of the map; Upper right panel, projection of the ISST; the coldest ISST are in the bottom right corner and the smallest ones in the upper right side of the map. Bottom right panel: Chlorophyll-*a* concentration; the highest chlorophyll-*a* concentration are located in the bottom right corner of the map, the smallest in the upper left corner of the map

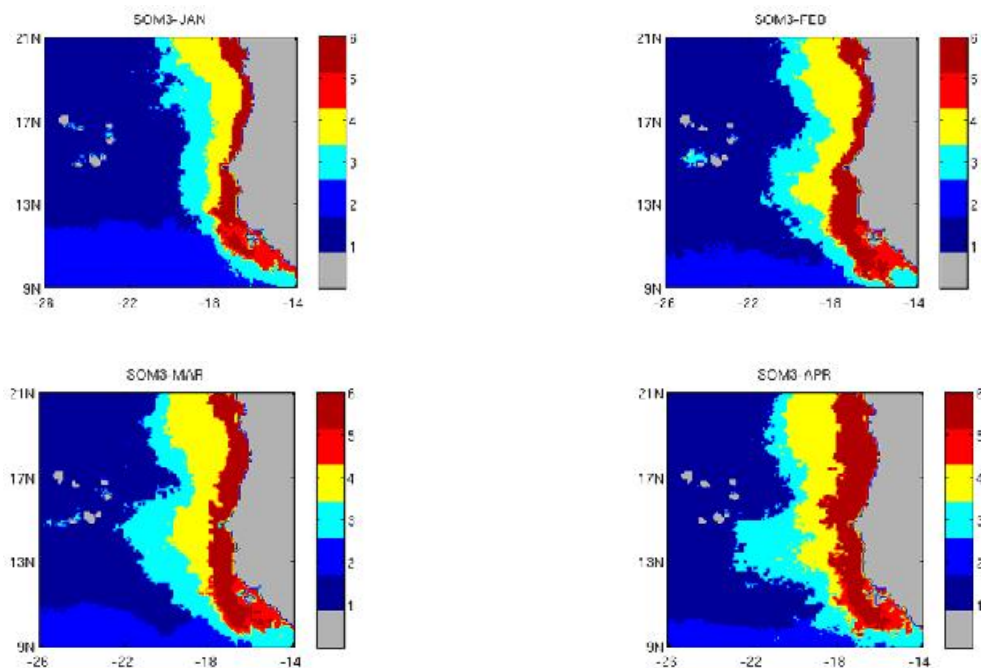


Fig. 7a: Geographical extend of the six classes for January, February, March and April. The classes are identified by the color bars on the right side of the cartoons

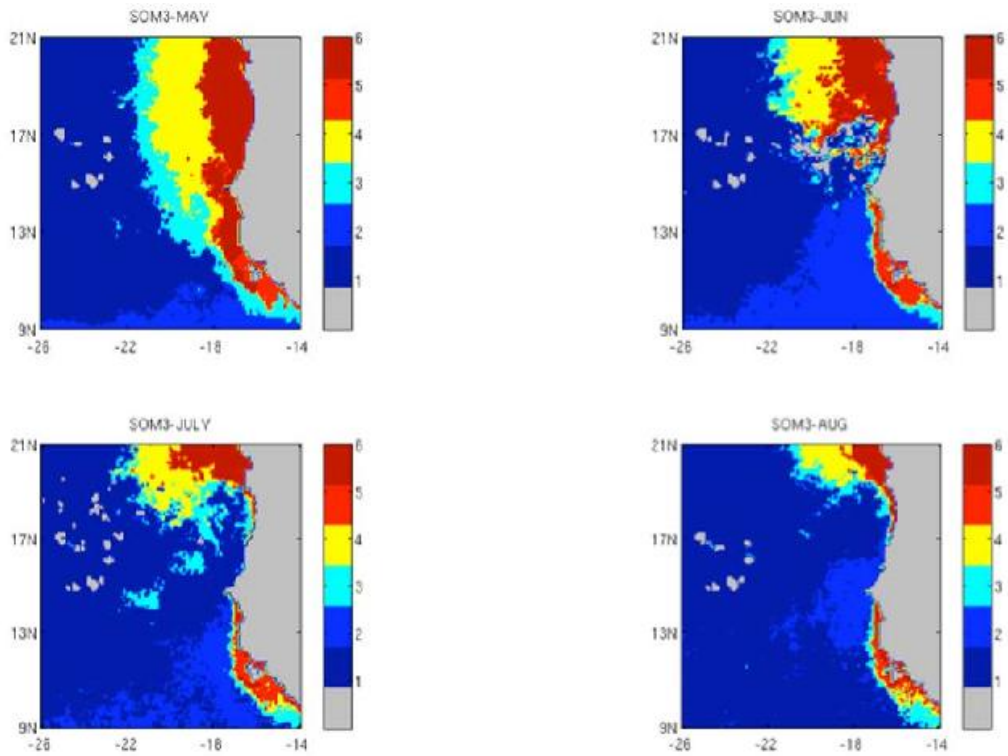


Fig. 7b: Geographical extend of the six classes for May, June, July and August. The classes are identified by the colors bars on the right side of the cartoons

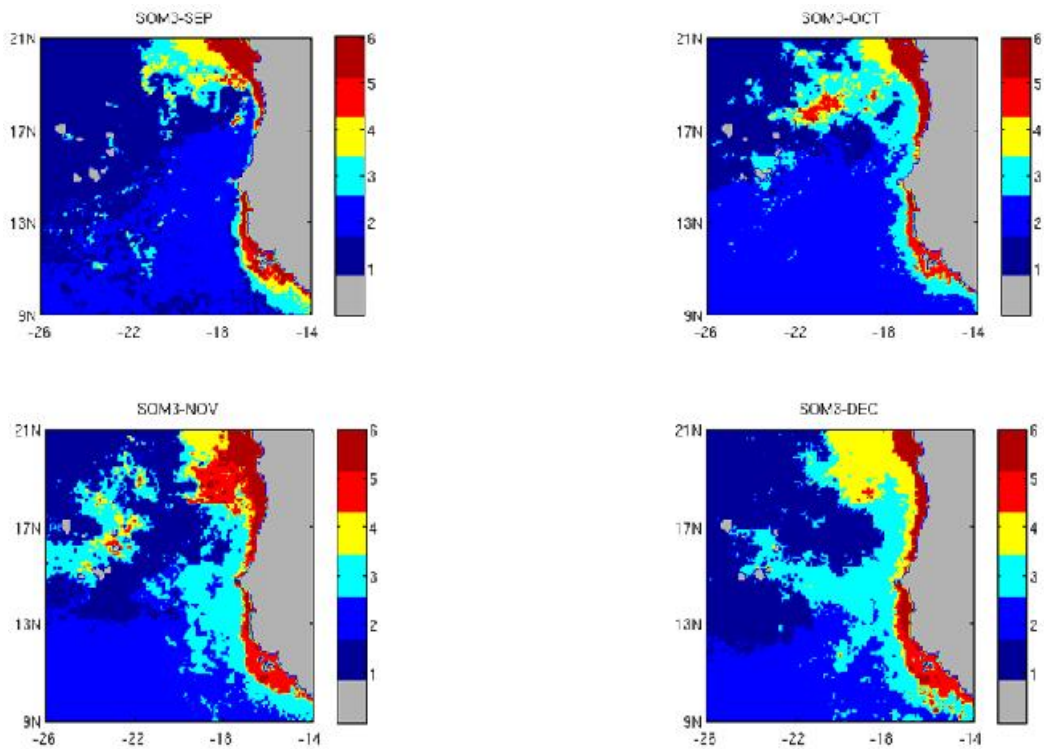


Fig. 7c: Geographical extent of the six classes for September, October, November and December. The classes are identified by the color bars on the right side of the cartoons

ANALYSIS OF THE SIX CLASSES

In the following, we present the mean geophysical characteristics of the six classes for the ten year period we analyzed. Each neuron of the topological map has captured a set of pixels of D . We project the set of pixels corresponding to each class for a specific month of the monthly climatology on a geographical map of the studied region (Fig. 7). We observe that the six

classes are geographically very coherent, since the pixels of a class are contiguous on the geographical map (Fig. 7). The classes, which have common statistical properties, are then associated with well-defined geographical areas. In Fig. 8, we have displayed the median chlorophyll- a content, the median ISST and SST values with their variances associated with the six classes in winter (January), when the upwelling is well developed and in August, when the

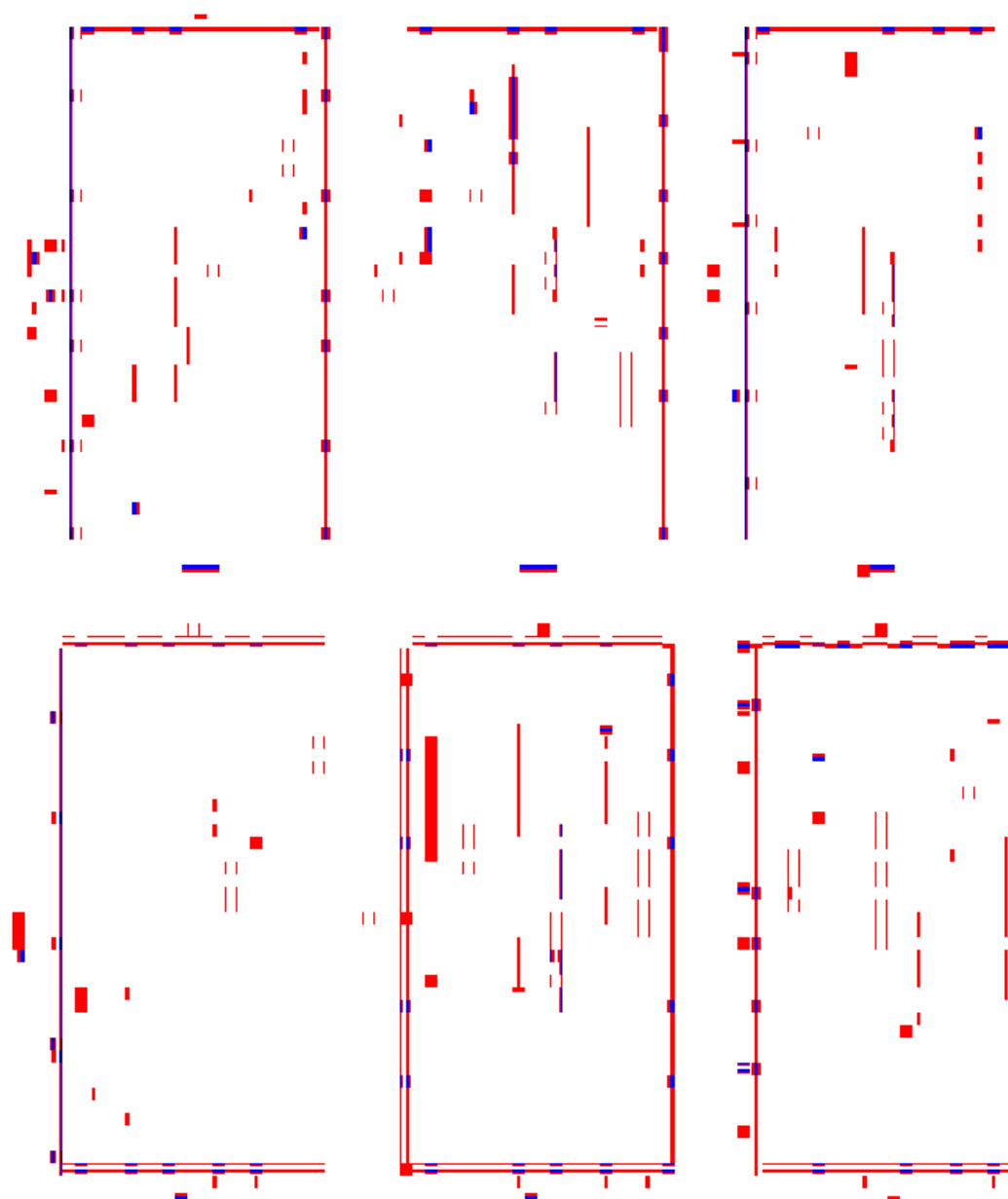


Fig. 8: Chlorophyll- a concentration (left hand panels), ISST (middle panels) and SST (right hand panels) values for the different classes in January (top) and August (bottom); Chlorophyll- a concentrations are given on a log scale; The little boxes on each panels includes 50% of the pixel values of the dedicated class and the line inside the box represents the median value of the class; The bars outside of each box correspond to the range of the remaining 50%

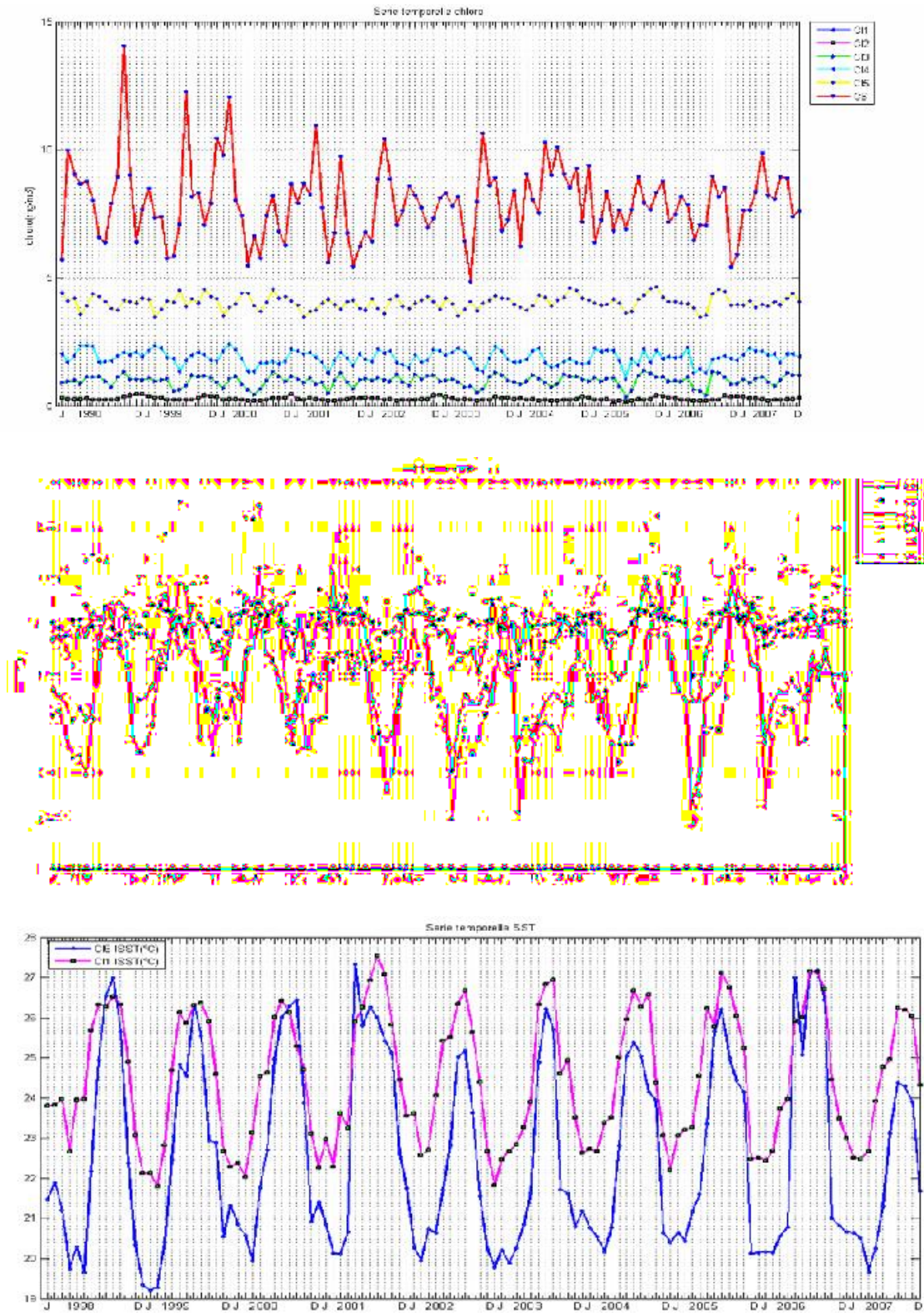


Fig. 9: Times series of the ocean parameters associated with each class: (upper panel) chlorophyll-*a* concentration, (middle panel) ISST, (bottom panel) SST (shown only for class 1 and 6). Note the strong seasonal variability for SST and ISST and chlorophyll-*a* concentration for class 6. The coldest SST were observed in 1999

upwelling signature is the weakest and practically non-existent. We first analyze the six classes from their Chl-a, ISST and SST characteristics:

- Class-1:** Corresponds to low chlorophyll-*a* concentration and warm SST. Its ISST is close to zero by construction.
- Class-2:** Corresponds to low chlorophyll-*a* concentration and slightly warmer SST and positive ISST.
- Class-3:** Corresponds to slightly higher chlorophyll-*a* concentration and slightly colder SST and negative ISST.
- Class-4:** Corresponds to higher chlorophyll-*a* concentration and cold SST and negative ISST.
- Class-5:** Corresponds to much higher chlorophyll-*a* concentration and quite warm SST and positive ISST.
- Class-6:** Corresponds to the highest chlorophyll-*a* concentration and the coldest SST and very high negative ISST.

Similar behavior is observed for the other months.

Figure 7 shows that class-1 (deep blue in the figure) is associated with offshore surface waters unaffected by coastal upwelling. This class extends far from the coast and is associated with a large number of neurons of the topological map (Fig. 6) corresponding to subtle differences in offshore water characteristics.

Due to their geographical location (Fig. 7) and ocean characteristics (Fig. 8) class-3, class-4 and class-6 are associated with the coastal upwelling. Class-6 (deep red in Fig. 7), represents costal upwelling waters with cold temperatures and very high Chl-*a* concentration. This class is concentrated along the coast and extends a few tens of kilometers offshore. It follows the coast north of Cap Verde (15°N) where the continental shelf is narrow and extends over the Guinea coastal shelf south of 15°N. It is well marked during the boreal winter (January, February, March), for which it can extend down to 12°N and disappears in summer (July, August, September) south of 17°N. Class-4 (yellow in Fig. 7) and class-3 (light blue in Fig. 7) extend quite far offshore. They correspond to waters that have been influenced by the costal upwelling through oceanic processes as instabilities and filaments. Class-3 waters are bounded by class-1 waters at its western limit (Fig. 7). Its offshore extent is important in winter and quite non-existent in summer.

Class-5 (light red in Fig. 7) corresponds to waters rich in chlorophyll and whose temperature is high. These waters are mainly located in shallow water area, such as the Guinea shelf south of 15°N and the Arguin bank at 21°N). The class-5 signature is visible up to 19°N in July, August and September replacing Class-6 waters at some location south of that latitude. In winter (January to April), class-5 waters (light red in Fig. 7)

are confined South of 12°N in the coastal region, covering the shallow continental Guinea shelf shown in Fig. 1. In fact, these waters are suspected to correspond to case-2 waters, which are coastal waters for which Chl-*a*, sediments and dissolved maters are mixed, leading to an overestimation of chlorophyll-*a* concentration values due to absorption of incoming solar radiation by the dissolved maters.

Class-2 waters (sky blue in Fig. 7) are waters whose chlorophyll-*a* concentration is quite small and temperature warm. They are located in the south part of the region during winter and spring. They move northward at the beginning of summer and reach 17°N in September, staying at that latitude until October and then moving back to south at the end of November. These waters may represent a signature of the North Atlantic Equatorial Counter Current (NECC), which intensifies in summer and forms the so-called Guinea Dome (Siedler *et al.*, 1992). During their northward progression in summer, class-2 waters never invade the Guinea continental shelf, which is mainly covered with class-5 and class 6 waters.

The SST of the six classes presents a well-marked seasonal variation, which is shown in (Fig. 8) and on monthly mean SST time series (Fig. 9). The SSTs of the six classes vary in phase. The SST is maximum in October and minimum in February-March-April. The seasonal variation is associated with the seasonal variation of the long shore component of the wind, which generates an offshore Ekman transport and thus a mass deficit, which is compensated by an upward advection of deep cold waters rich in nutriment at the coast. Surprisingly the chlorophyll concentration does not show well-marked seasonal variation as seen in Fig. 9, except for class 6, whose Chl-*a* values present high seasonal and inter annual variations during the first five years of the studied period which could not be related to any identifiable indices such as NAO, Enso or African monsoon indices.

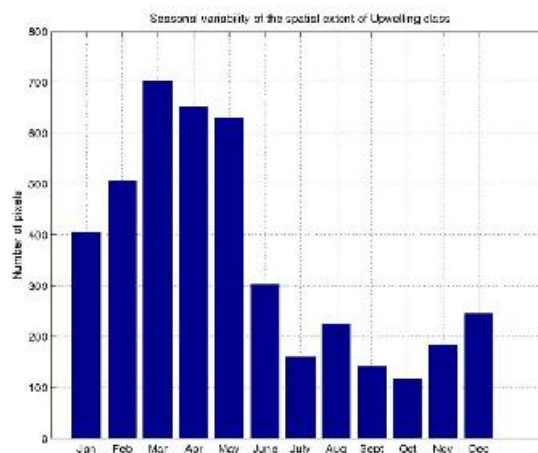


Fig. 10: Seasonal variability of the spatial extent of class 6 waters



Fig. 11: Time series of the monthly wind vector at Dakar from 2003 to 2006. The wind direction is given according to the axes drawn in the upper left panel

Moreover, we note that the extent of class 6 waters, which is associated with the coastal upwelling, presents a strong seasonal variation as shown in Fig. 10. This extent is minimum in summer (July to September). It starts to grow in November, is maximum in March, stays well developed until May and begins to decrease in June. This extent is associated with the long shore component of the trade winds. The decrease is linked to the northward progression of the ITCZ (Inter Tropical Convergence Zone) and the ensuing decrease in intensity of the trade winds that drive the upwelling dynamics (Fig. 11). The ratio between the maximum extent class-6 area and its minimum extent is about 6, showing the strong variability of the upwelling area.

In Fig. 11 shows the time series of the monthly wind vector at Dakar from 2003 to 2006. The wind presents a strong seasonal variability, with a maximum in winter (November to May) and a minimum in summer (July, August, September) when the ITCZ reaches its northernmost position. We note that the seasonal variability of the extent of class 6 waters shown in Fig. 10 varies with the southward component of the wind, in agreement with the upwelling theory (Allen, 1973). The more intense the North South wind component, the larger the extent of class 6 waters.

CONCLUSION

A Self Organizing Map (SOM) associated with a hierarchical ascending clustering provides an efficient index to investigate the variability of the Senegalo-Mauritanian upwelling by combining SST, ISST and chlorophyll-*a* values observed by satellite remote sensors using a multivariate statistical relationship. The grouping of the data in few (six) classes based on three major signatures of the upwelling, SST and ISST, as physical parameters and chlorophyll-*a* concentration as biogeochemical parameter, allowed us to study the variability of the upwelling. This method has enabled us to pick up the major characteristics of the Senegalo-Mauritanian upwelling. The values of the upwelling parameters are very distinct for the six classes (Fig. 9) justifying the CAH classification a posteriori. In Fig. 6,

we have displayed the decomposition in six classes on SOM and in Fig. 7 the geographical pattern of the six classes. The different classes can be related to ocean phenomena: class 6 corresponds costal upwelling waters (cold and high Chl-*a* concentration), while class 1 corresponds to offshore waters (warm and low Chl-*a* concentration). Class 4 and 3 correspond to waters, which have been influenced by the upwelling through temperature and chlorophyll-*a* diffusion due to filaments and instabilities. Class 2 can be related to the NECC waters moving northward in June-July with the ITCZ displacement and enhancing the dome of the Gulf of Guinea. Class 5 can be associated with shallow shelf waters corresponding to class-2 water. It mainly corresponds to case-2 waters and is found on the Arguin bank and on the shallow shelf extending off the Gambia and Guinea coasts. This was confirm by analyzing the nLw* spectrum (spectrum normalized by the Chl-*a* spectrum). The class 5 normalized spectra are always larger than unity, which could be due to the effect of the light backscattering onto the sediment particles present in the waters, confirming that these waters are case-2 waters. The class 6 normalized spectra, which are also larger than unity can interpreted by high chlorophyll-*a* concentration associated with a large variety of phytoplankton species. This may lead to an important package effect (Bricaud *et al.*, 1987) that makes the chlorophyll-*a* concentration-absorption relationship non-linear and the normalized spectrum different from unity. The very high variance for classes-5 and -6 indicates the complexity of the interaction between the incoming solar radiation and the biogeochemical constituents of the water column.

Our method has also enabled us to quantify the spatial extent of the upwelling in terms of pixel numbers and estimate its seasonal variability. The variability of this extent is linked to the variability of the southward component of the wind that is parallel to the coast. Nevertheless, for the ten years of observations, we were not able to detect any climatic trends in the SST of the different classes, nor in the chlorophyll-*a* concentration. This might be due than in strong upwelling area as fed by deep waters, the

characteristics of this deep water are much more stable than surface layers. Moreover, despite attempting to investigate the inter-annual variability of the upwelling in terms of spatial extent and duration (expressed in month of the year) in respect of class 6, which reflects the upwelling dynamics, we were not able to reach definite conclusions regarding the inter annual variability of this extent, which is quite small and is biased by the cloud coverage masking the relaxation of the upwelling in summer and very often its spatial extent all year around. This effect is important during the summer months (July, August, September), leading to spurious effect on the class extent.

The results of the method we have developed show the interest in combining the different variables through a unique multivariate statistical process rather than to combine the results of analysis done separately on the different variables. As an example, analysis of the chlorophyll-*a* concentration can provide classes having similar concentration; analysis of ISSTs or SSTs can also provide classes having similar ISSTs or SSTs. But the merging of three contours obtained separately is a very delicate operation whereas the multivariate analysis done by SOM provides an optimum combination of the information embedded in the three data sets. Class 5 and class 2 oceanic characteristics would not have been easily identified by applying mono variable statistical methods on each variable separately.

The method we have presented is relevant to analyze a large variety of phenomena, which have been observed with several different sensors. It permits to combine different measurements given by different sensors in a rational manner to extract pertinent information on the phenomena, which would not have been obtained by using the observation separately as shown in this study. Moreover the present method, which uses SOM, is very easy to implement and to handle, as there exists friendly software dedicated to the handling them. Other method such as k-means could also be used to make data fusion analyses. In fact SOM is an extension of the K-means method in which the different clusters are related together allowing a more efficient partitioning as shown in Badran *et al.* (2005), specially when the analyzed data set is related to complex physical laws.

ACKNOWLEDGMENT

This study has been supported by a grant of the French institution CNES (Centre National d'Etudes Spatiales) under the contract TOSCA-(2011-2012).

REFERENCES

Allen, J.S., 1973. Upwelling and coastal jets in a continuously stratified ocean. *J. Phys. Oceanogr.*, 3(3): 245-257.

- Aristegui, J., X.A. Alvarez-Salgado, E.D. Barton, F.G. Figueiras, S. Hernandez-Leon, C. Roy and A.M.P. Santos, 2004. Oceanography and Fisheries of the Canary Current/Iberian Region of the Eastern North Atlantic. In: A.R. Robinson and K.H. Brink (Eds.), *the Sea*. Chapter 23, John Wiley and Sons Inc., New York, 14: 877-931.
- Badran, F., M. Yacoub and S. Thiria, 2005. Self-organizing Maps and Unsupervised Classification. In: Dreyfus, G. (Ed.), *Neural Networks, Methodology and Applications*. Chapter 7, Springer, Berlin, Heidelberg, New York, pp: 379-442.
- Barth, J.A., 1989. Stability of a coastal upwelling front 2. Model results and comparison with observations. *J. Geophys. Res.*, 94(C8): 10844-10856.
- Barton, Eastern Boundary of the North Atlantic, 1998. Northwest Africa and Iberia. In: Robinson, A.R. and K.H. Brin (Eds.), *the Sea*. John Wiley and Sons Inc., New York, 11: 633-657.
- Beletsky, D., W.P. O'Connor, D.J. Schwab and D.E. Dietrich, 1997. Numerical simulation of internal Kelvin waves and coastal upwelling fronts. *J. Phys. Oceanogr.*, 27: 1197-1215.
- Bricaud, A., A. Morel and J.M. André, 1987. Spatial/temporal variability of algal biomass in the Mauritanian upwelling zone, as estimated from CZCS data. *Adv. Space Res.*, 7(2): 5362-5384.
- Brink, K., 2005. Coastal Physical Processes Overview. In: Robinson, A. and K. Brink (Eds.), *Harvard University Press*, NY.
- Crepon, M., C. Richez and M. Chartier, 1984. Effect of coastline geometry on upwellings. *J. Phys. Oceanogr.*, 14(8): 1365-1382.
- Demarcq, H. and V. Faure, 2000. Coastal upwelling and associated retention indices from satellite SST. Application to octopus vulgaris recruitment. *Oceano. Acta*, 23: 391-407.
- Jain, A.K. and R.C. Dubes, 1988. *Algorithms for Clustering Data*. Prentice Hall, Englewood, Cliffs, New Jersey, pp: 320.
- Kohonen, T., 1982. Self-organized formation of topologically correct feature maps. *Biol. Cybern.*, 43: 59-69.
- Kohonen, T., 2001. *Self Organizing Maps*. 3rd Edn., Springer, Berlin, Heidelberg, New York.
- Lachkar, Z. and N. Gruber, 2012. A comparative study of biological production in eastern boundary upwelling systems using an artificial neural network. *Biogeosciences*, 9: 293-308.
- Lange, C.B., O.E. Romero, G. Wefer and A.J. Gabric, 1998. Offshore influence of coastal upwelling off Mauritania, NW Africa, as recorded by diatoms in sediment traps at 2195 m water depth. *Deep Sea Res. I*, 45: 985-1013.
- Lathuiliere, C., V. Echevin and M. Levy, 2008. Seasonal and intraseasonal surface chlorophyll-*a* variability along the northwest African coast. *J. Geophys. Res.*, 113: C05007, Doi: 10.1029/2007J C004433.

- Niang, A., L. Gross, S. Thiria, F. Badran and C. Moulin, 2003. Automatic neural classification of ocean colour reflectance spectra at the top of atmosphere with introduction of expert knowledge. *Remote Sens. Env.*, 86(2): 257-271.
- Niang, A., F. Badran, C. Moulin, M. Crépon and S. Thiria, 2006. Retrieval of aerosol type and optical thickness over the Mediterranean from SeaWiFS images using an automatic neural classification method. *Remote Sens. Env.*, 100(15): 82-94.
- O'Brien, J.J. and H.E. Hurlburt, 1972. A numerical model of coastal upwelling. *J. Phys. Oceanogr.*, 2(1): 14-26.
- Siedler, G., N. Zangenberg, R. Onken and A. Morlière, 1992. Seasonal changes in the tropical atlantic circulation: Observation and simulation of the guinea dome. *J. Geophys. Res.*, 97(C1): 703-715.

Article à soumettre au journal RSE

Ocean color spectral clustering of the Senegalo-Mauritanian waters in relation with phytoplankton species.

O. Farikou^{1,3}, M. Crépon⁴, S. Thiria⁴, S. Sawadogo^{1,3}, A. Niang³,
J. Brajard⁴, M. Carlos⁴, Y. K. Mahamoud¹

17 mars 2013

¹Institut Universitaire des Sciences et Techniques d'Abéché (IUSTA), BP 6077 N'Djaména-Tchad.

²Laboratoire de Traitement de l' Information (LTI) – EPT, BP A10 Thiès, Sénégal.

³Laboratoire de Traitement de l' Information (LTI) – ESP- UCAD, BP : 5085 Dakar-Fann (Sénégal).

⁴Institut Pierre Simone Laplace/Laboratoire d' Océanographie et Climat : Expérimentations Approches Numériques (IPSL/LOCEAN), Université. Paris 6, 75252 Paris, France.

E-mail : farikou@yahoo.fr, s_sawadogo@yahoo.fr, awaniang@ucad.sn, Julien.Brajard@locean-ipsl.upmc.fr.

1 Introduction

Ocean color measurements have been intensively used to investigate Chlorophyll-a concentration (Chl-a in the following) of surface waters of the ocean, marginal seas and lakes. As these measurements are made at several wavelengths, we expect that this spectral information can be used to distinguish the different phytoplankton species. These species present different spectral signature due to the absorption properties of the pigments and to their backscattering properties that are specific to each species. As an example, diatoms, which have a silica skeleton more strongly backscatter the sun light than haptophytes.

This region has been recently intensively studied by processing SeaWiFS ocean color data and MODIS sea surface temperature as reported in [4] and [3]. We note the quasi presence of an intense upwelling both on ocean color signal and SST. We note very high chl-a concentration near the shore and a strong offshore gradient with, which present a strong seasonal

variation ([4], [10]). The Upwelling is modulated by the intensity of wind component parallel to the coast. It weakens in July-August with the arrival of the ICTZ coming from South, which decreases the intensity of the trade winds. On several locations (Arguin bank, shelf off the Sine Saloum delta), we note the presence of waters whose ocean color spectra correspond to cas

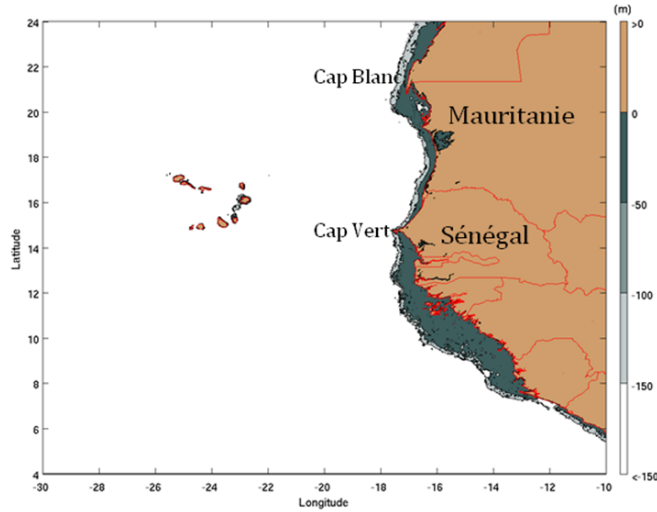


Figure: 1 – Mauritania and Senegal coastal topography. The land is in red and the ocean depth is represented with the color scale in meters (right of the figure)

2 waters. In the present paper we analyzed the ocean color spectrum provided by the SeaWiFS sensor with a special focus on phytoplankton species identification. Section 2 describes the ocean color data we processed. Section 3 presents the methodology we used. Section

4 is devoted to the analysis of the ocean color spectra. Section 5 analyses the results in the studied region. Section 6 is devoted to a critical summary and a conclusion.

2 The data

We used the water leaving reflectances $\rho_w(\lambda)$ computed by [3], which can be found on the web site : <http://poacc.locean-ipsl.upmc.fr/>. These water leaving reflectances were obtained by processing the SeaWiFS daily reflectances, $\rho_{toa}^{obs}(\lambda)$ observed at the top of the atmosphere (TOA) with the NeuroVaria algorithm [3] from 1997 up to 2010. This algorithm was specifically trained for taking into account absorbing aerosols. In order to extract second order effect due to specific water characteristics other than phytoplankton abundance, we computed a specific water leaving radiance $\rho_w^*(\lambda)$ [1] : defined as :

$$\rho_w^*(\lambda) = \frac{\rho_w(\lambda)}{\rho_w^{ref}(\lambda, Chl - a)}$$

where $\rho_w^{ref}(\lambda, Chl - a)$ is a simple model of $\rho_w(\lambda)$ that accounts only for the SeaWiFS standard Chl-a. In the present paper the $\rho_w^{ref}(\lambda, Chl - a)$ as $nLw^{ref}(\lambda, Chl - a) \times E_0(\lambda)$ where $E_0(\lambda)$ is the solar irradiance, nLw^{ref} is $Chl - a \rightarrow nLw^{ref}(\lambda, Chl - a)$ model computed by [4] by using MLPs which are neural networks dedicated to function estimation from data bases. We then computed the $\rho_w(\lambda)$ from the SeaWiFS daily values from 1997 up to 2010. We now propose to cluster the $\rho_w(\lambda)$ spectra in different groups to tentatively evidence different phytoplankton species.

3 The methodology

The methodology is similar to that used in [4]. We first applied an unsupervised classification similar to that done by [9] et [8] on a data base composed of the $\rho_w(\lambda)$ daily values for the year 2003 which is a year for which the cloud coverage was very low with respect to the 13 years of observation. The aim was to summarize the information contained in the five-component vector data set D by producing a small number of reference vectors (rvs) that are statistically representative of the data. Each reference vector (rv) represents a set of vectors of D that have a similar pattern. In the present case we chose to determine the rvs by minimizing some distance between rv and the vectors of D it represents. To do this, we used a specific neural network based model, the so-called topological

map, which was first introduced by [6]. Each neuron of the map (Figure ??) is associated with a particular reference vector (rv) and thus corresponds to a group (in this paper, a set of pixels belonging to D). The rvs approximate the density of the data set D . They are computed by minimizing a specific cost function as in the K-Means algorithm ([2]). Besides the different neurons of the topological map C are connected together and determine a topological (neighbourhood) relationship between the different groups (neurons). Close neurons on the maps correspond to rvs that are quite similar ; very distant neurons correspond to rvs that are very different. The set of rvs represents the dataset D by compressing the information contained in it.

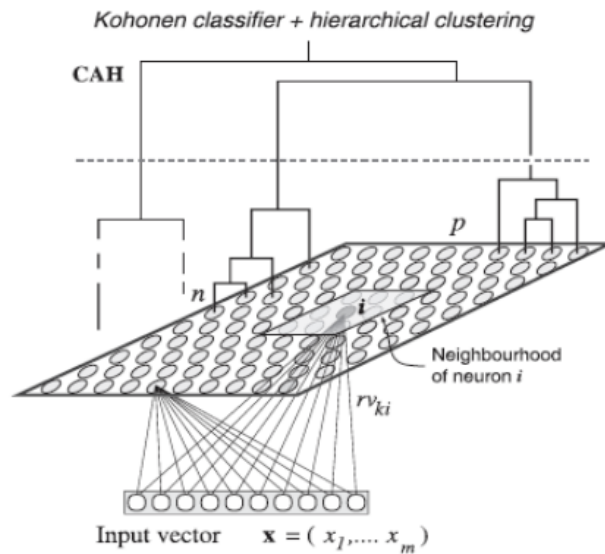


Figure: 2 – Structure of the self-organizing map (SOM). The network comprises two layers : an input layer used to present observations and an adaptation layer for which a neighbourhood system is defined (distance d between neurons and a neighbourhood function). Each neuron i is fully connected to the input layer. It is associated with a group that is represented by a reference vector rv_i . In the figure the neurons are cluster in 3 classes by HAC

In the present study, we dealt with a two-dimensional map with quite a large number of neurons (20x20) and thus of rvs , providing a highly discriminating representation of the

observations. The pixels of the data set D are thus clustered into 400 groups. We used the SOM version available on the web site <http://www.cis.hut.fi/projects/somtoolbox/download/>. The topological map was learnt according to the procedure described in [7]. The number of neurons was determined empirically from solutions of similar problems and then adjusted as described in [2]. The large number of groups allowed us to take into account the complexity of the dataset but may have prevented us from synthesizing some geophysical information embedded in the data, such as spatial or seasonal specificities. To counteract this difficulty, we decided to aggregate this large number of groups into a smaller number of classes based on the similarities of the

groups. We thus extracted a few pertinent classes from the groups by clustering groups having similar statistical properties expecting that the classes can be associated with geophysical characteristics. For that we used a hierarchical ascendant classification (HAC in the following), which is a bottom-up hierarchical classification [5]. This method iteratively computes a partition hierarchy. From the initial partition (the neurons on the map), two subsets of the computed partition are clustered at each iteration. These two subsets are selected by measuring their similarity according to the Ward criterion. We aggregated the 20x20 neurons into 10 significant classes. The resulting clustering of the five dimension vectors r_{vs} associated with the neurons of the topological map is given in

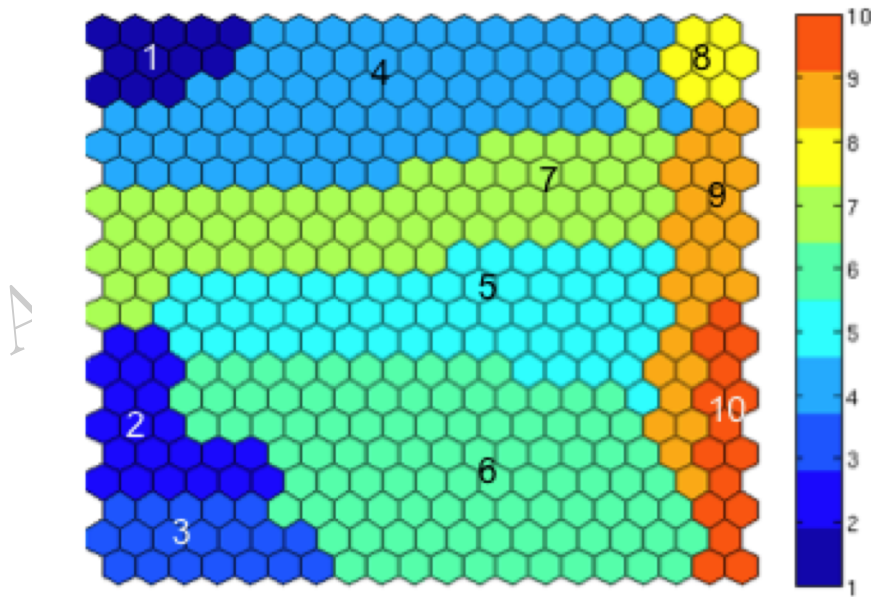


Figure: 3 – representation of the ten classes on the SOM map. Each class is represented by a different color.

We note that the topological map + HAC clustering is very coherent, since the classes represent clusters whose neurons are contiguous on the topological map. Moreover the spectrum associated with each neuron (or r_v) determine homogeneous fields on SOM. Their gradients are smooth, well

defined without any discontinuity. The number of groups (10) was selected because it presented the most significant discriminative partition with respect to the full dendrogram of the HAC (not shown) one the one hand and to the spectrum homogeneity on the other hand.

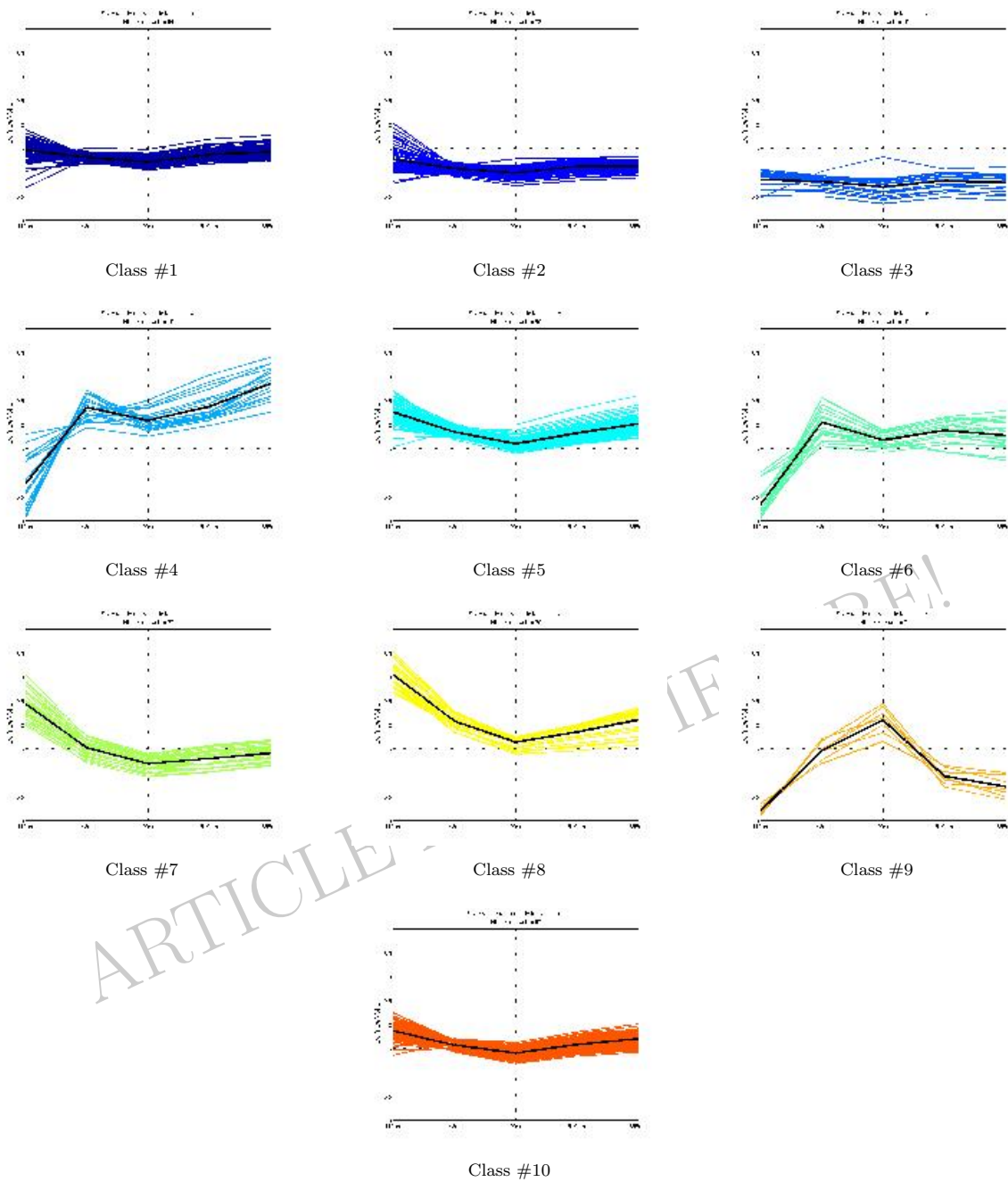


Figure: 4 – Representation of the ten specific water leaving radiance Spectra.

4 Spectrum analysis

The topology of the corresponding SOM map is shown in Figure 3. The spectral pattern of the ten classes of the specific water leaving radiance spectrum obtained after the CAH are shown in Figure 4. The last three spectra (8, 9, 10) are characteristic of

case 2 waters with a high absorption at 412 nm due to dissolve organic substances. The high values of spectrum 10 at higher wave numbers correspond to backscattering due to suspended particles. The low value of spectrum 8 at high wave numbers may correspond to absorption due to organic matter in very coastal waters. The other classes may be analyzed

through the classification proposed by [1]. Classes 2 and 3 may correspond to diatoms; owing to their proximity on the topological maps and the similar pattern of their spectrum they could be merged together. Class 1 may correspond to Haptophytes, class 4, 5 et 7 to SLC and Prochlorococcus and class 6 to diatoms as classes 2 and 3. Class analysis is

We see in that figure that classes 2 and 3 are associated with quite high AOT values and consequently were not considered for a biological interpretation of the geographical pattern since these two classes can be potentially contaminated by the atmosphere. Due to the same reasons we did not interpret classes 8, 9 and 10. We thus decided to focus our analysis on classes on classes 4, 5, 6, 7 only, which are associated with low AOT values (less than 0.25). We also studied the class 1 geographical patterns for which we removed neuron 1 data which are highly contaminated by the atmosphere as shown by the quite high value of its mean AOT(Figure 5). We analyzed several time periods for which AOT were small. Among them, we selected a six day period in January 2004 (16, 17, 24, 26, 28, 31) which is characteristics of the observed situations during the thirteen year period we

A possible interpretation of the can be the following : the costal upwelling generates phytoplanktons whose mean spectrum corresponds to class1. The offshore ocean (deep ocean) corresponds to class 4 spectrum, characterizing large ocean area with low Chl-a concentration. Arguin bank and Siné-Saloum, Guinea bank, which are both very shallow regions are characterized by classes 8, 9, 10 spectra corresponding to case 2 waters as argued above. In the south part of the region, we observe class 6 waters, which are well marked in Figure 9. Class 6 spectrum corresponds to diatoms [1]. These diatoms start to develop on the 17th of January near the coast, present a well define extent on the 24th January, progresses

a very delicate procedure as the spectra could have been contaminated by small defects in the atmospheric correction. In order to check that point, we have represented the mean atmospheric optical thickness (AOT in the following) that is associated with the reflectances captured by each neuron (Figure 5).

processed the data. The Chl-a concentration images (figure 7) show a well-developed upwelling with very high chl-a concentration at the coast. The mean Chl-a concentration decreases offshore. Chl-a concentration is advected offshore by meanders, which can persist several days; see as an example the **"hook like meander"** off Cap Verde in the 17, 26, 31 January images. We note very high Chl-a concentration in the Arguin bank (20°N) and South of Cap Verde in the Siné-Saloum, Guinea bank (14° - 11°N). The spectral signature maps (Figure 8, 9) show that class 1 waters are located near the coast (in regions extending several tenths of kilometers offshore) forming a strip like parallel to the coast; they are surrounded on their west side by class 4 waters, which occupy the deep ocean.

to west on the 28 January and are still visible on the 31 January. This time sequence can be identified as a bloom of diatoms developing in the ECC waters which occupy that region [4]. This above scenario is typical of many situations observed during the thirteen year period of observation.

5 Discussion and Conclusions

A neuronal classification procedure has permitted to identify 10 classes of specific water leaving radiance spectrum $\rho_w^*(\lambda)$ water leaving radiance spectrum $\rho_w(\lambda)$ normalized by a $\rho_w^{ref}(\lambda, Chl - a)$ spectrum which is a simple model of $\rho_w(\lambda)$ that accounts

MOYENNE ANNUELLE DE PIXELS: TAU-865NM-2003

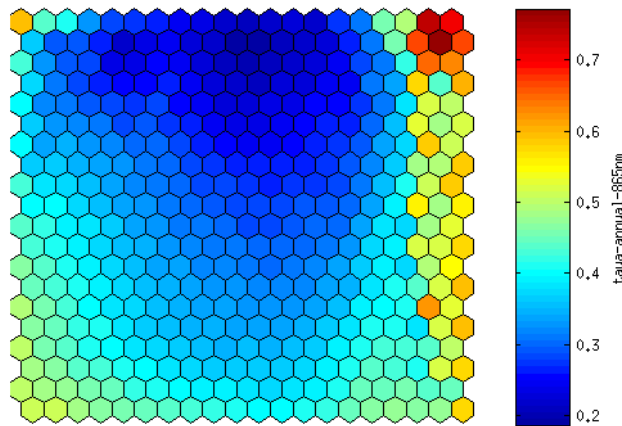


Figure: 5 – Mean optical depth associated with each reflectance captured by each neuron of SOM. The color bar on the right is AOT value scale

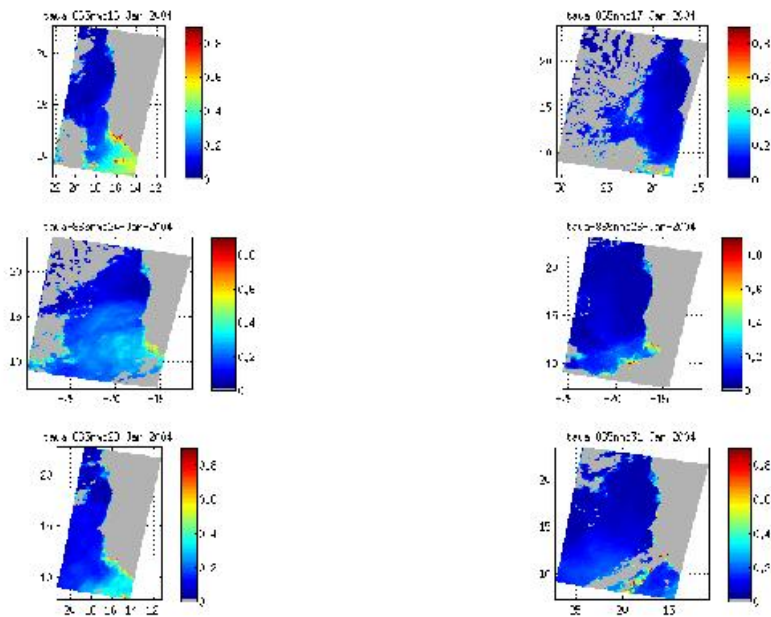


Figure: 6 – AOT Pattern for the 16, 17, 24, 26, 28, 31 January 2004,

only for the SeaWiFS standard Chl-a. By projecting the AOT onto the classifier, we found that several classes (2, 3, and 8, 9 and 10) were contaminated by the atmosphere. Classes 2 and 3 are encounter in offshore regions. We did not try to interpret them in term of phytoplankton species since this interpretation

could be biased by the atmospheric signal. Regarding classes 8, 9 and 10, which correspond to coastal water occupying well defined region near the coast, they were identified as case 2 waters having a strong absorption at 412 nm. They are mainly encounter on the Arguin bank ($20^{\circ}N$) and on the continental shelf

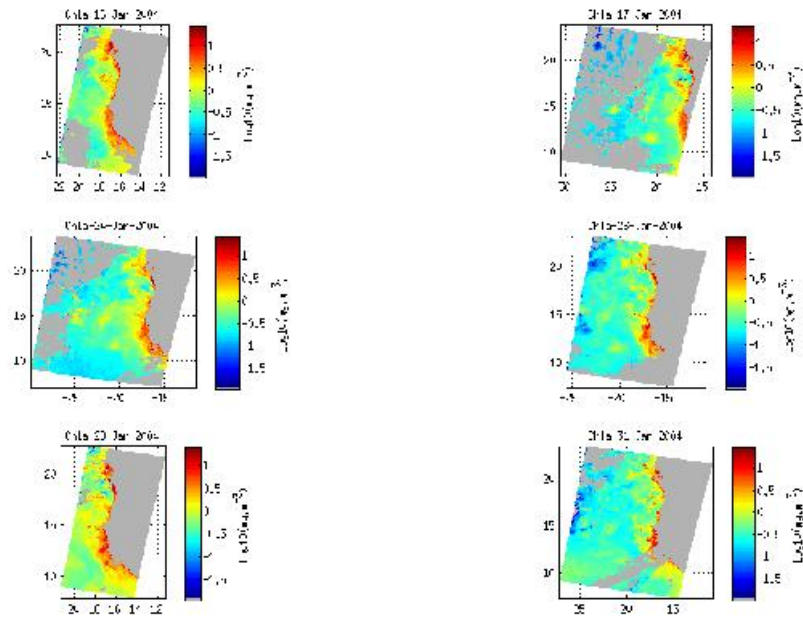


Figure: 7 – Chl-a concentration maps for the 16, 17, 24, 26, 28, 31 January 2004

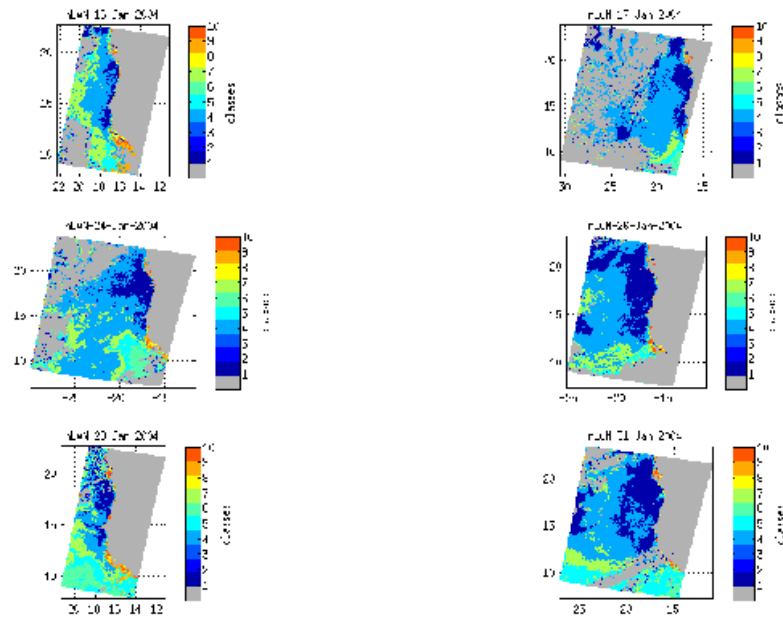


Figure: 8 – Spectrum Pattern for the 16, 17, 24, 26, 28, 31 January 2003, days for which AOT coverage is very low

off Siné-Saloum and Guinea coasts, which are very shallow regions very sensitive to dissolve matters (class

8 spectrum) and to suspended particles reflecting light (class 10 spectrum) on the other hand. After inspecting a large number of situations for which AOT

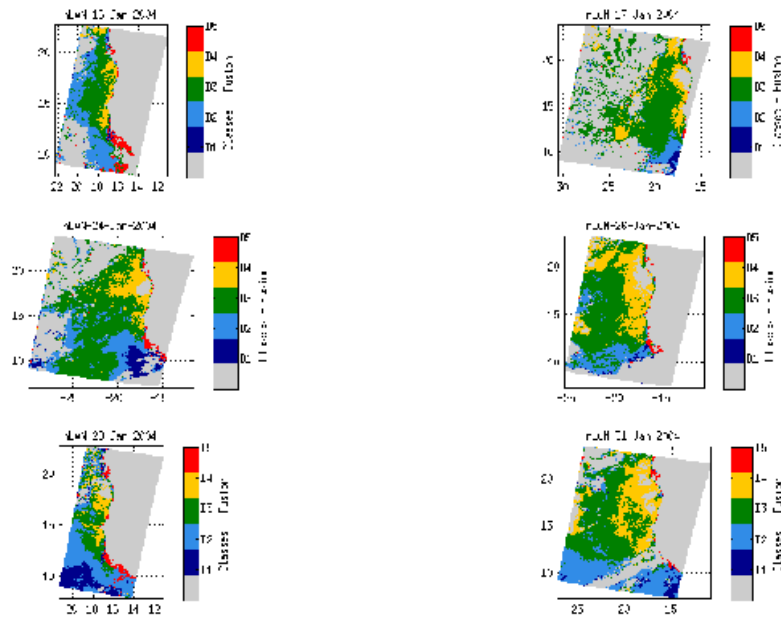


Figure: 9 – Spectrum Pattern for the 16, 17, 24, 26, 28, 31 January 2003, days for which we have enhanced class 1 (D4, haptophytes), class 5 (D1, diatoms), class 4 (D3, SLC). We have merged class 2, 3, 5, 7 which can not be determined at present time into D2 and class 8, 9, 10 which are associated with case 2 waters into D5

is small, we arrived to outline the following scenarios for interpreting the space variability of the spectra in terms of biogeochemical parameters in the upwelling region :

- The upwelling produces phytoplankton species, which are characterized by class 1 spectrum. According to [1], class 1 spectrum can be associated with haptophytes. The offshore waters are characterized by class 4 spectrum, which are associated with prochlorococcus and SLC (Synechococcus-like and cyanobacteria). We observed the presence quasi permanent of diatoms in the ECC waters.
- The difficulty in interpreting the different spectra and attributing them to specific phytoplankton species is partly due to the difficulty to obtain a very precise atmospheric correction. Moreover, the classes are qualitative indices ; the transition from one class to another abrupt, discontinuous.

Références

- [1] S. Alvain, C. Moulin, Y. Dandonneau, and F.M. Bréon. Remote sensing of phytoplankton groups in case 1 waters from global seawifs imagery. *Deep Sea Research Part I Oceanographic Research Papers* 52, 11 1989-2004”, 2005.
- [2] F. Badran, M. Yacoub, and S. Thiria. Self-organizing maps and unsupervised classification. *G. Dreyfus, Springer, Berlin Heidelberg New York*, 2005.
- [3] D. Diouf, A. Niang, S. Thiria, J. Brajard, M. Crépon, and S. Sawadogo. Monitoring aerosol optical thickness and chlorophyll-a to the west africa coast using a neural-variational method. *Journal des Sciences Pour l'Ingénieur, Vol. 13*, pp. 46-51. ISSN : 0851-4453., 2011.
- [4] O. Farikou, S. Sawadogo, A. Niang, J. Brajard, M. Crépon, and S. Thiria. Variabilité spatio-temporelle de l’upwelling sénégal-mauritanien à

- partir d'observations satellitaires. *Journal des Sciences Pour l'Ingénieur, Vol. 13, pp. 60-71. ISSN :*, 2011.
- [5] A.K. Jain and R.C Dubes. Algorithms for clustering data. *Englewood. Cliffs, New Jersey, USA : Prentice Hall Advanced Reference Series, 320 pp*, 1998.
- [6] T. Kohonen. *Self-organized formation of topologically correct feature maps*. Biological Cybernetics 43, pp. 59–69, 1982.
- [7] T. Kohonen. *Self Organizing Maps*. Springer, Berlin, Heidelberg, New York, Tokyo, 459pp., 2001.
- [8] A. Niang, F. Badran, C. Moulin, M. Crepon, and S. Thiria. Retrieval of aerosol type and optical thickness over the mediterranean from seawifs images using an automatic neural classification method. *Remote Sensing of Environment, pp. 82-94.*, 2006.
- [9] A. Niang, L. Gross, S. Thiria, F. Badran, and C. Moulin. Automatic neural classification of ocean colour reflectance spectra at the top of atmosphere with introduction of expert knowledge. *Remote Sensing of Environment, pp. 257-271.*, 2003.
- [10] S. Sawadogo, J. Brajard, A. Niang, C. Lathuiliere, M. Crépon, and S. Thiria. Analysis of the senegalo-mauritanian upwelling by processing satellite remote sensing observations with topological maps. 2009.

ARTICLE A SOUMETTRE!

Annexe C

Conférence Internationale

**European Geosciences Union General
Assembly 2011 Vienna Helsinki-He 03-08
April 2011
Résumé de la présentation**



Spatio-temporal variability of the Senegalo-Mauritanian upwelling from satellite observations

Salam Sawadogo (1), Ousmane Farikou (2), Awa Niang (3), Julien Brajard (4), Michel Crépon (4), and Sylvie Thiria (4)

(1) L.T.I. / Ecole Polytechnique de Thiès, Thiès, Senegal (s_sawadogo@yahoo.fr), (2) L.T.I. / Ecole Supérieure Polytechnique, UCAD, Dakar, IUSTA (Abéché) Senegal (farikou@yahoo.fr), (3) L.T.I. / Ecole Supérieure Polytechnique, UCAD, Sénégal (awaniang@ucad.sn), (4) IPSL/LOCEAN, Université. Paris 6, France (Julien.Brajard@locean-ipsl.upmc.fr).

The Senegalo-Mauritanian upwelling is a very productive upwelling occurring along the West coast of Africa. It extends from 26°N down to 10°N. It has been well documented during the past decades from in-situ and satellite observations. The analysis of satellite-derived chlorophyll concentration (Lathuilière et al, 2008) showed that the upwelling can be split into two regions, one north of 21°N where the seasonality is very weak, the other south of that longitude, where the upwelling presents a strong seasonality.

The aim of the present study is to find some objective criteria for decomposing the physical biochemical phenomenon into different spatio-temporal phases leading to coherent interpretation of that phenomenon. In the present study, interest is focused on the south part of the upwelling, between 10°N-21°N mainly facing the coast of Senegal and Mauritania, which presents a strong seasonal variability.

Its seasonal and inter-annual variability south of 20°N was analyzed by processing ocean color data from 1998 to 2007 (Chl-a concentration) provided by the SeaWiFS satellite radiometer and sea surface temperature by the NOAA/AVHRR instrument.

We used a classification methodology based on these two parameters consisting in a neural network topological map and a hierarchical ascendant classification.

We showed that six classes can explain most of the variability of this region in temperature and Chl-a concentration, and among these classes, one allowed us to explain the variability of the upwelling. Its extent is maximum in February-March, minimum in August-September. The interannual variability is linked to that of the wind. The classes can be considered as statistical indices allowing us to investigate the variability of the upwelling.

Lathuilière C., V. Echevin, M. Levy (2008)- Seasonal and intraseasonal surface chlorophyll-a variability along the northwest African coast- J. Geophys. Res, 113, C05007, doi:10.1029/2007JC004433

Annexe D

Présentation orale sous forme de poster (EGU)

Présentation orale sous forme de poster



Introduction

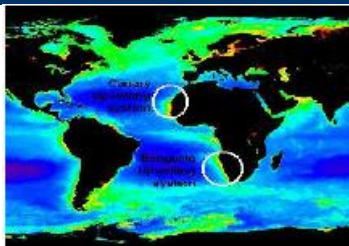
The Senegalo-Mauritanian upwelling is a very productive upwelling occurring along the West coast of Africa.

The Senegalo-Mauritanian upwelling is one of the major systems of coastal upwelling along the western edges of continents (California, Peru-Chile, Benguela)

The analysis of surface temperature and chlorophyll-a concentrations shows that the Senegalo-Mauritanian upwelling can be split into two regions (Lathuilière et al, 2008):

- > One north of 21° N, where the seasonality is very weak.
- > The other south of 21° N where the upwelling presents a strong seasonality.

The aim of the present study is to find some objective criteria for decomposing the physical/biochemical phenomenon into different spatio-temporal phases leading to coherent interpretation of that phenomenon.



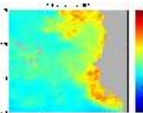
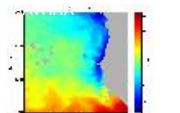
Studied area

In the present study, interest is focused on the south part of the upwelling region, between 9° N to 21° N and 14° W to 26° W.

Data sets

Sea Surface Temperature SST
AVHRR MODIS satellite Radiometer

CHL-a concentration
SeaWiFS satellite Radiometer

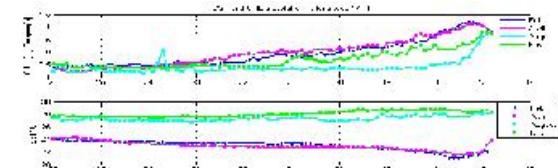


- Spatial resolution : 0.1° - Temporal resolution : monthly -
Image : 121x121 pixels
- Period: January 1998 to December 2007

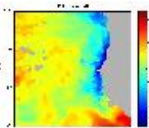
The SST data are provided by the NOAA / AVHRR instrument.

CHL-a data are provided by the SeaWiFS satellite radiometer

Longitudinal profiles of temperature and chlorophyll-a



SST index



Since the upwelling signature is mainly characterized by the offshore gradient of SST, we thus decided to describe it by a SST gradient index.

This index (denoted ISST) is defined as the difference, at a given latitude, between the mean SST in the band 22° - 24° W and the SST at the observed point at the given latitude.

We compute the climatology of SST as an average for each month of valid pixels (other than land and cloud). We obtain 12 matrix of size 121x121 (climatology from January to December)

Idem for ISST and CHL-a.

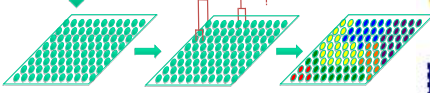
To resume, there are two types of data :

Data base	Size	Type	Description
F1	3x121x121x12x10	(ISST, CHL-a) + SST	Original Data
F2	3x121x121x12	(MISST, MCHL-a) + MSST	Monthly climatology Monthly averages over 10 years

Methodology

Classification using a self-organizing map (SOM) + Hierarchical ascending clustering (HAC)

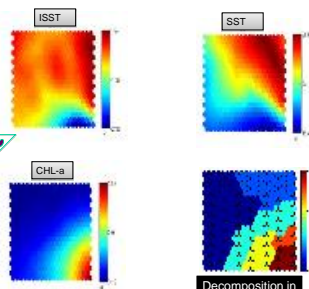
dataset F2 (ISST, SST, chl-a)



1) Unsupervised classification of the pixels into 400 groups with a self-organizing maps

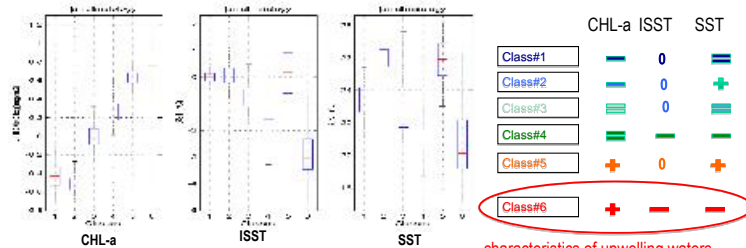
2) hierarchical ascending clustering

3) classification of F2 in 6 classes.



Result-Analysis

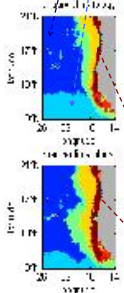
"Box Plots" showing the concentrations of Chl-a (in Log scale), ISST and the SST of the different classes in January.



characteristics of upwelling waters

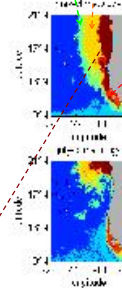
Class #1 and class #2

are associated with offshore surface waters non affected by the coastal upwelling. Class#2 corresponds to the North Equatorial Counter Current (NECC) that moves northward in summer.



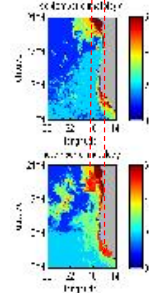
Class #3 and class #4

Class #3 and class #4 are intermediary classes characterized by median chlorophyll-a values and seem to be described by chlorophyll-a patches out of the upwelling region.



Class #5

Class #5 is atypical. It is characterized by waters rich in chlorophyll with no SST gradient and hot waters.

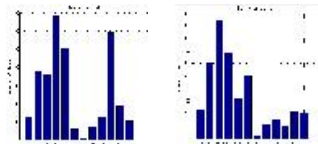


Class#6

Represents coastal upwelling waters. This class is concentrated along the coast and extends a few tens of kilometers offshore

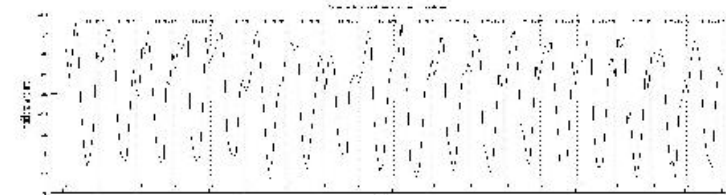
Seasonal variability of the spatial extent of class 6

Class 6 has a strong seasonal variability with maximum in March, April, May, a second maximum in October-November and a minimum during the rainy season (July, August) when the long shore component of the wind weakens.



Seasonal variability of the upwelling

Seasonal variations of the upwelling are related to those of the wind. Note the strong seasonal variability of the wind with a maximum in March when the extension of the class 6 is the maximum.



Conclusion

> A topological map associated with a hierarchical ascending clustering provides an efficient index to investigate the variability of the Senegalo-Mauritanian upwelling by combining SST and Chl-a observed by satellite remote sensors through a statistical relationship.

> The decomposition into classes allowed us to better understand the dynamics of the Senegalo-Mauritanian upwelling, in particular its spatial variability depending on the season.

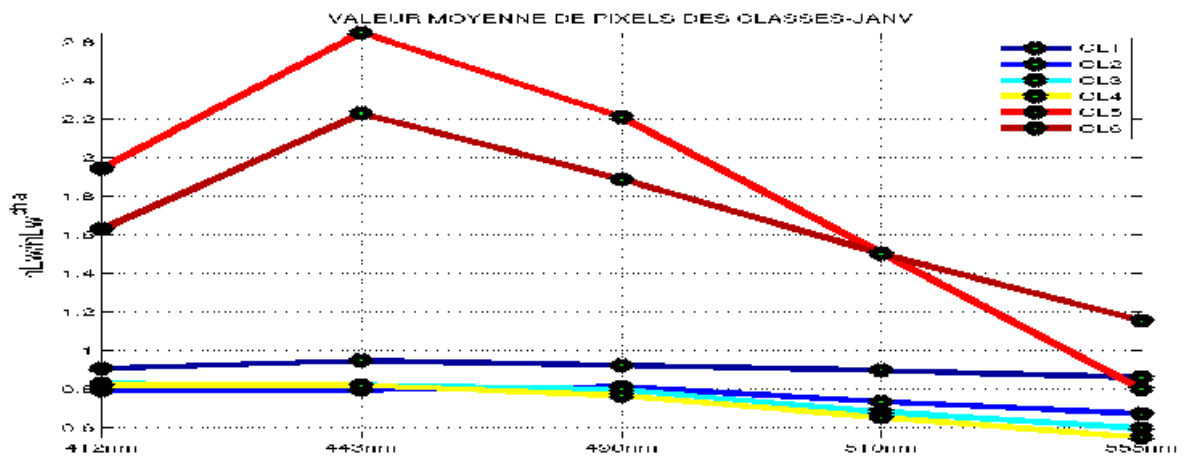
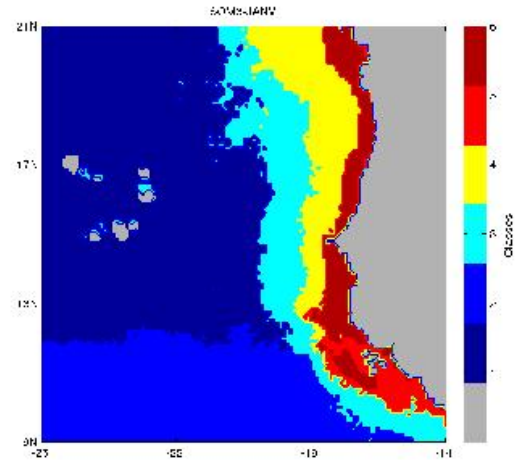
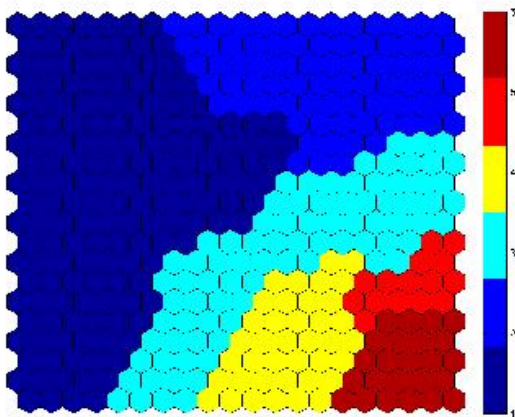
> We have been able to quantify the spatial extent of the upwelling in terms of pixel numbers and estimate its seasonal variability which is linked to this of trade winds.

Annexe E

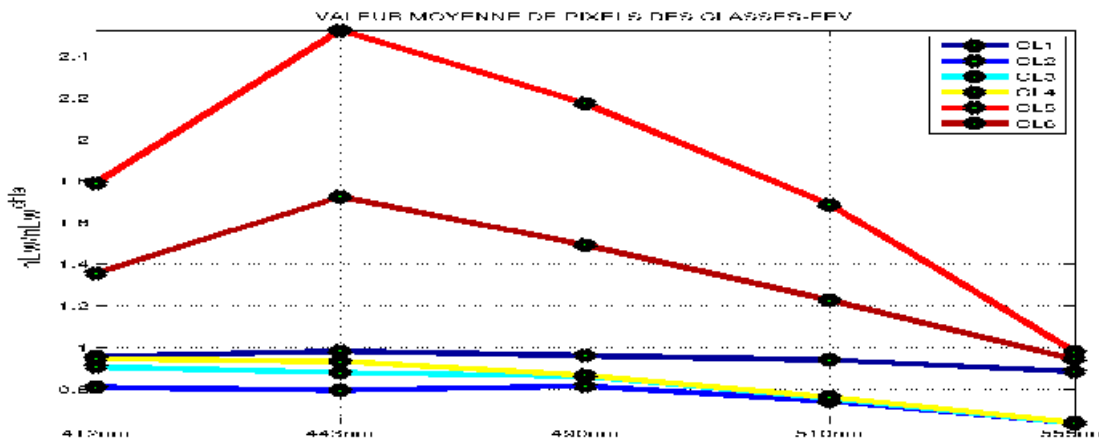
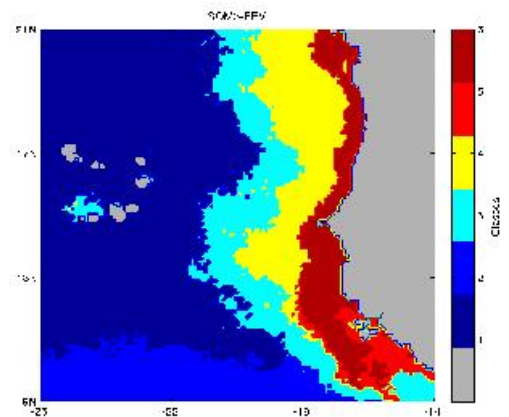
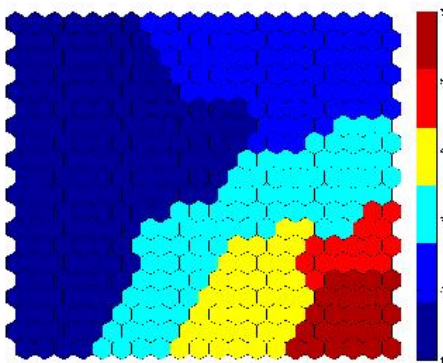
Spectres climatologiques

Les spectres climatologiques

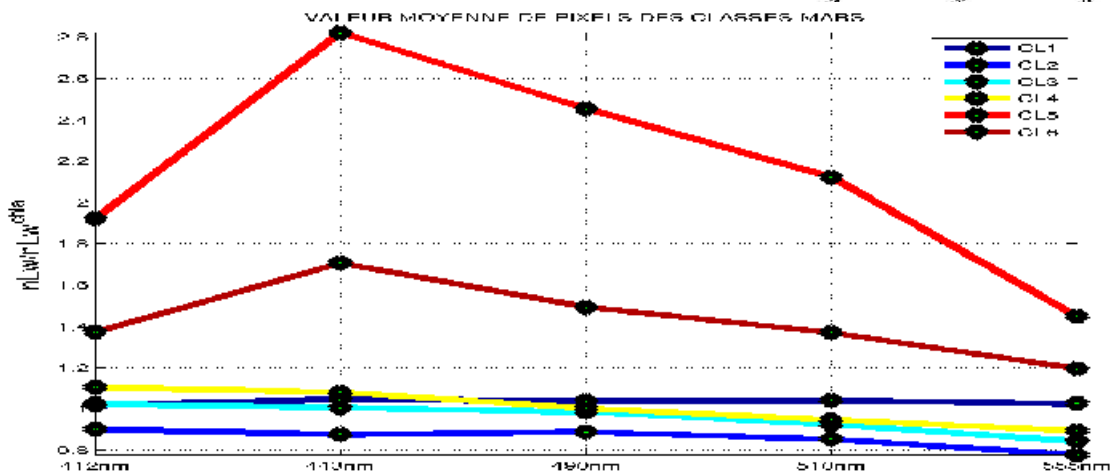
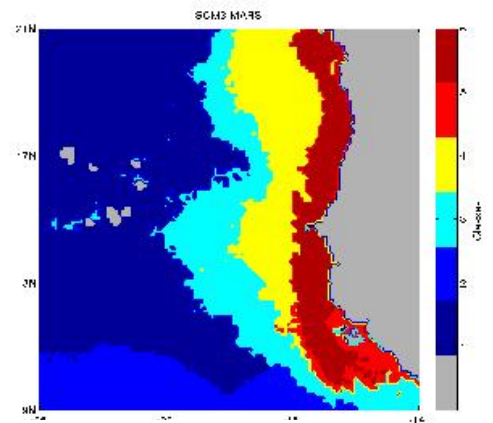
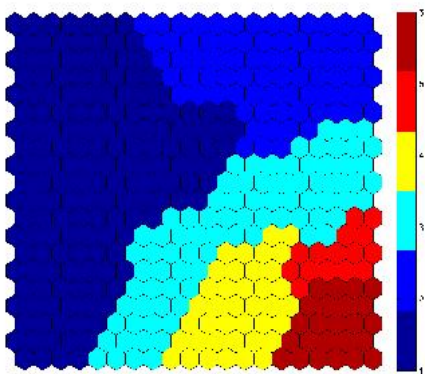
SOM3->Carte geog->Spectres de classes – janv climato



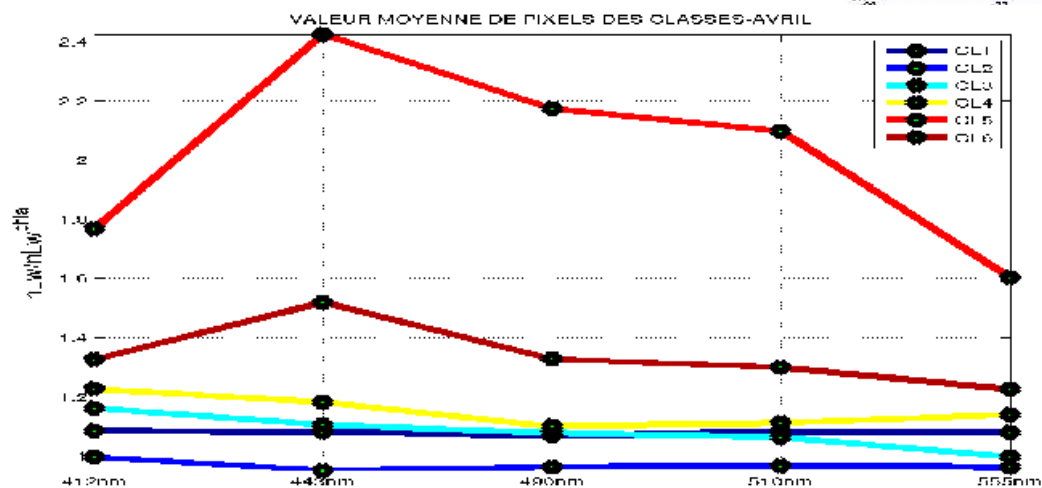
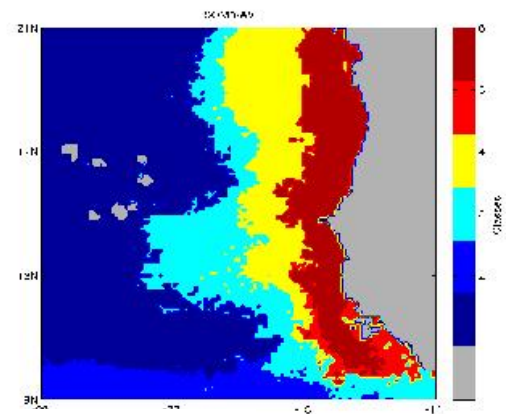
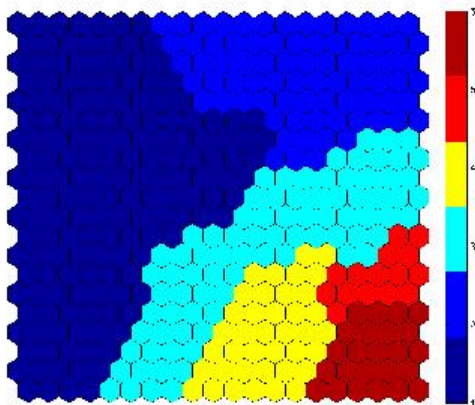
SOM3->Carte geog->Spectres de classes – fev climato



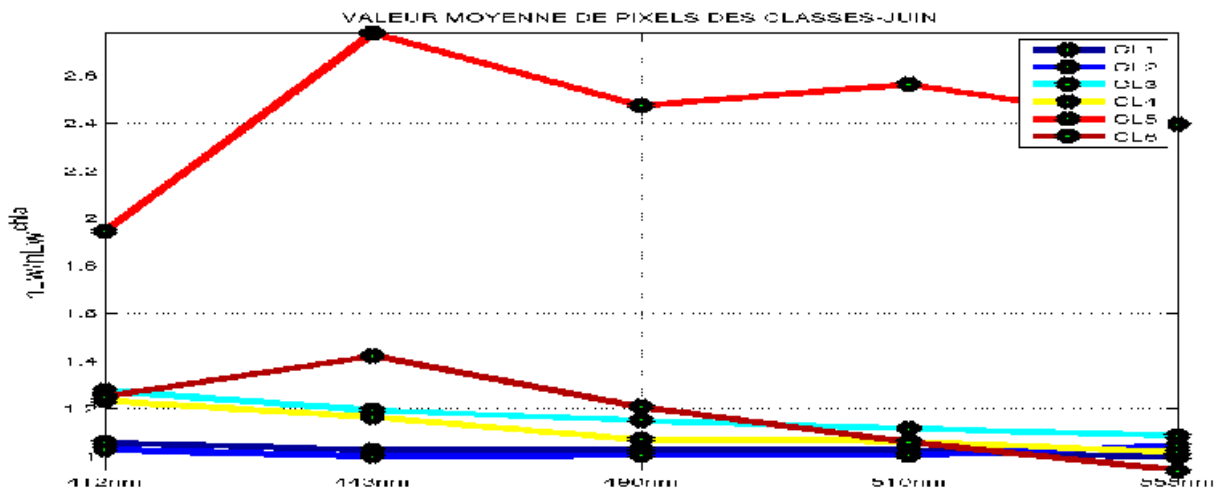
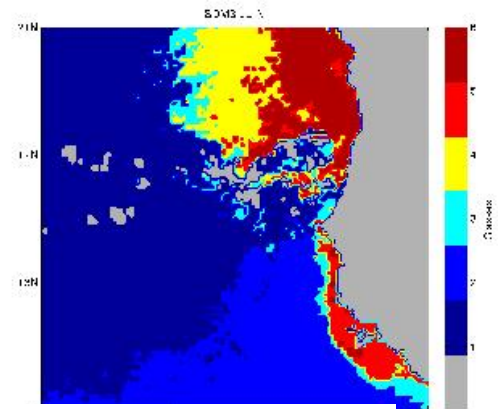
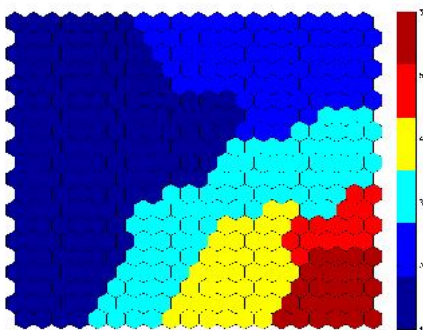
SOM3->Carte geog->Spectres de classes – mars climato



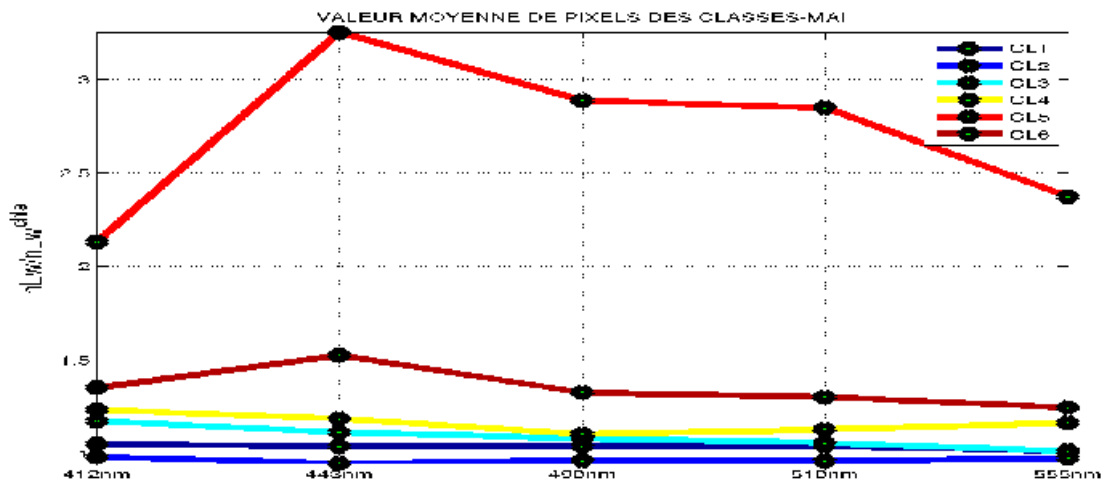
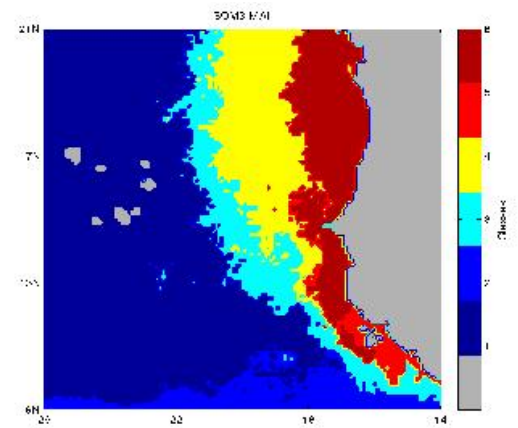
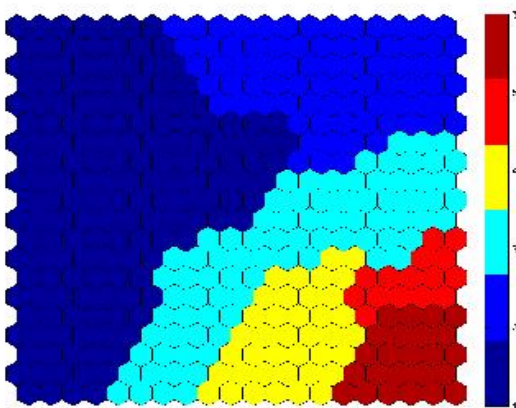
SOM3->Carte geog->Spectres de classes – avril climato



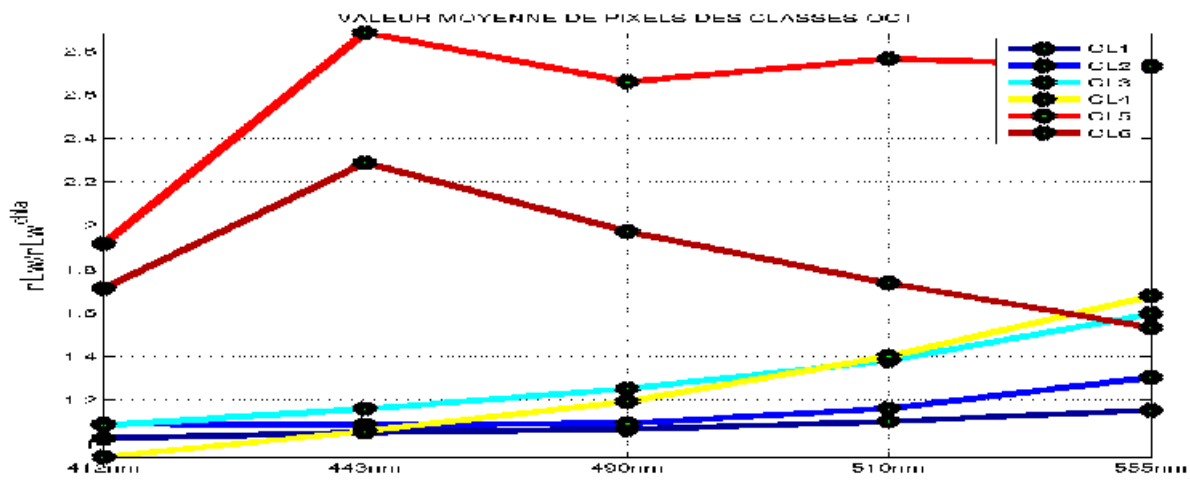
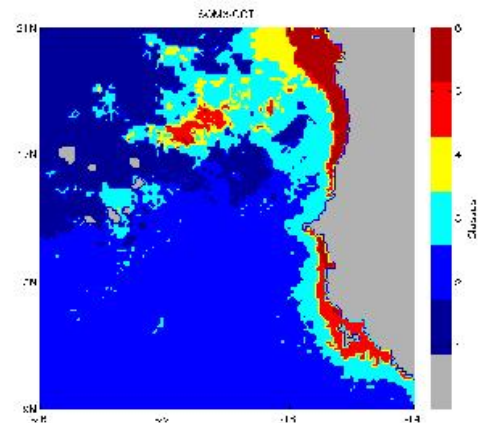
SOM3->Carte geog->Spectres de classes – juin climato



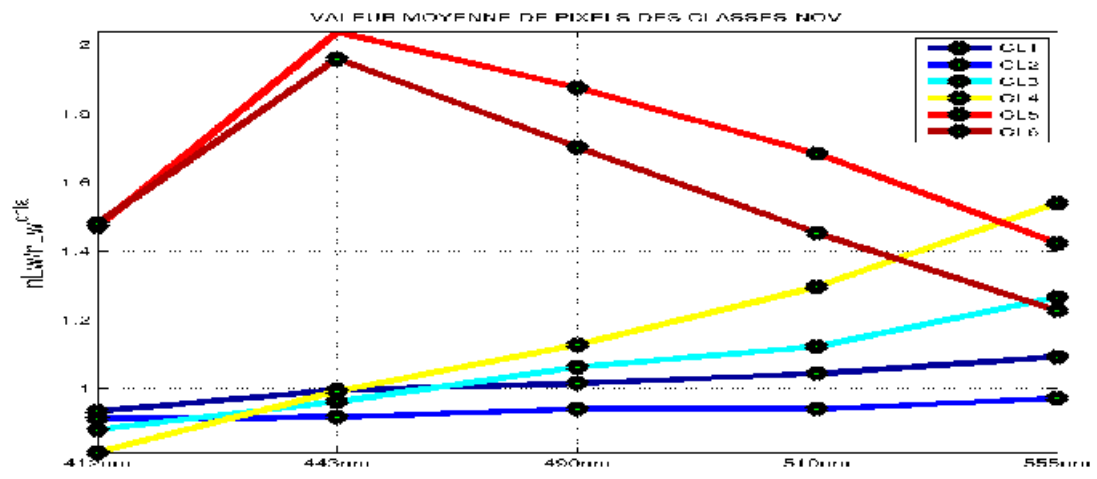
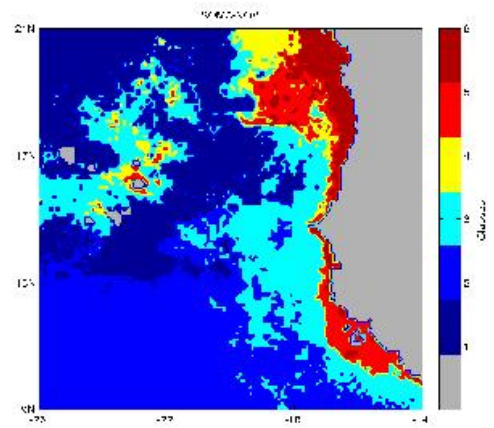
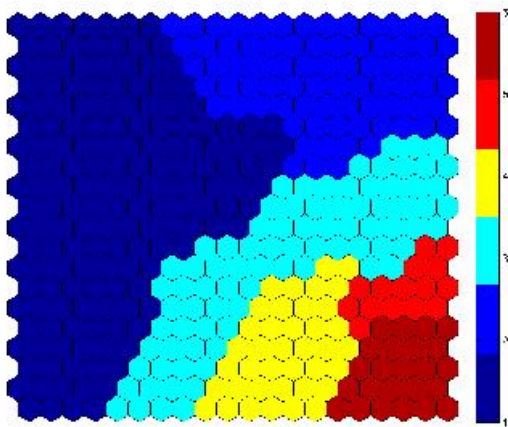
SOM3->Carte geog->Spectres de classes – mai climato



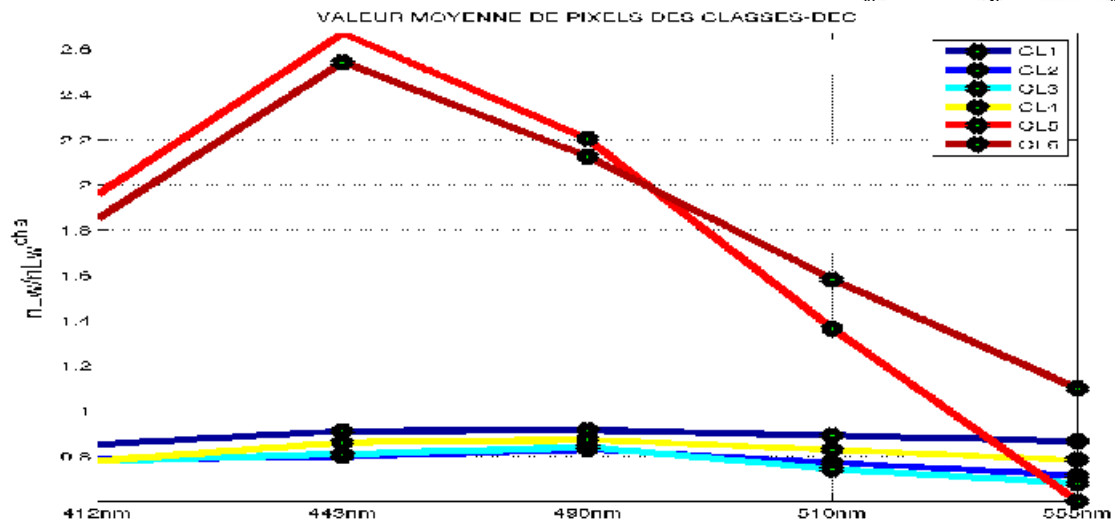
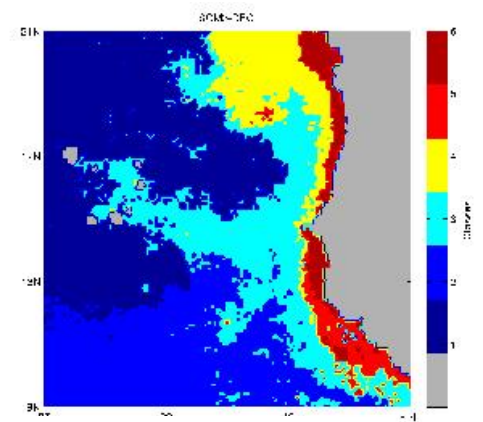
SOM3->Carte geog->Spectres de classes – oct climato



SOM3->Carte geog->Spectres de classes – nov climato



SOM3->Carte geog->Spectres de classes – dec climato



Annexe F

Statistique SOM2(Chla-a,ISST)

Statistiques de la classification de
SOM2(Chla-a,ISST)

Tableau 1 – Résultats de l'apprentissage MLP et projection du reseau optimal sur les données tests.

nLw($\lambda=412nm$)					
Neurones	Iterations	RMS_All	RMS_App	RMS_Val	RMS_test
3	100	0,1715	0,1714	0,1736	0,1697
3	200	0,1712	0,1712	0,1710	0,1715
3	400	0,1711	0,1708	0,1703	0,1730
3	600	0,1712	0,1713	0,1727	0,1689
3	800	0,1711	0,1713	0,1710	0,1704
3	1000	0,1712	0,1714	0,1708	0,1704

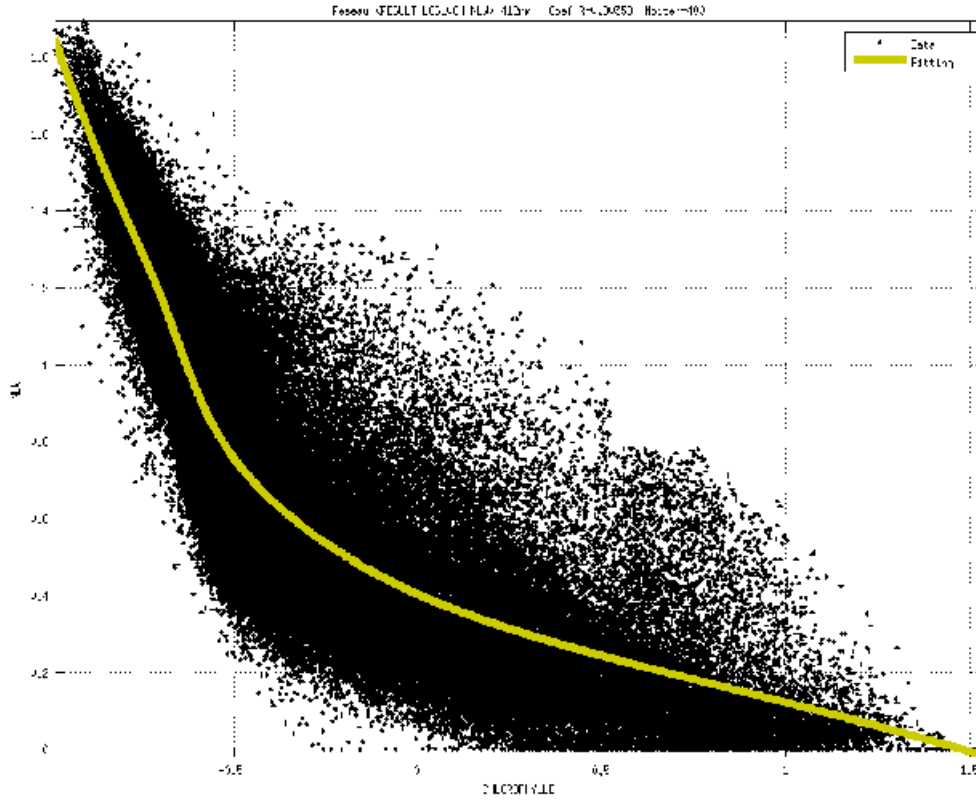
Figure 1 – Projection du reseau optimal sur les données tests. ($\lambda = 412nm$)

Tableau 2 – Résultats de l'apprentissage MLP et projection du reseau optimal sur les données tests.

nLw($\lambda=443nm$)					
Neurones	Iterations	RMS_All	RMS_App	RMS_Val	RMS_test
2	100	0,1394	0,1714	0,1396	0,1385
2	200	0,1393	0,1390	0,1398	0,1404
2	400	0,1394	0,1390	0,1404	0,1400
2	600	0,1393	0,1395	0,1401	0,1380
2	800	0,1394	0,1395	0,1385	0,1395
2	1000	0,1393	0,1391	0,1402	0,1394

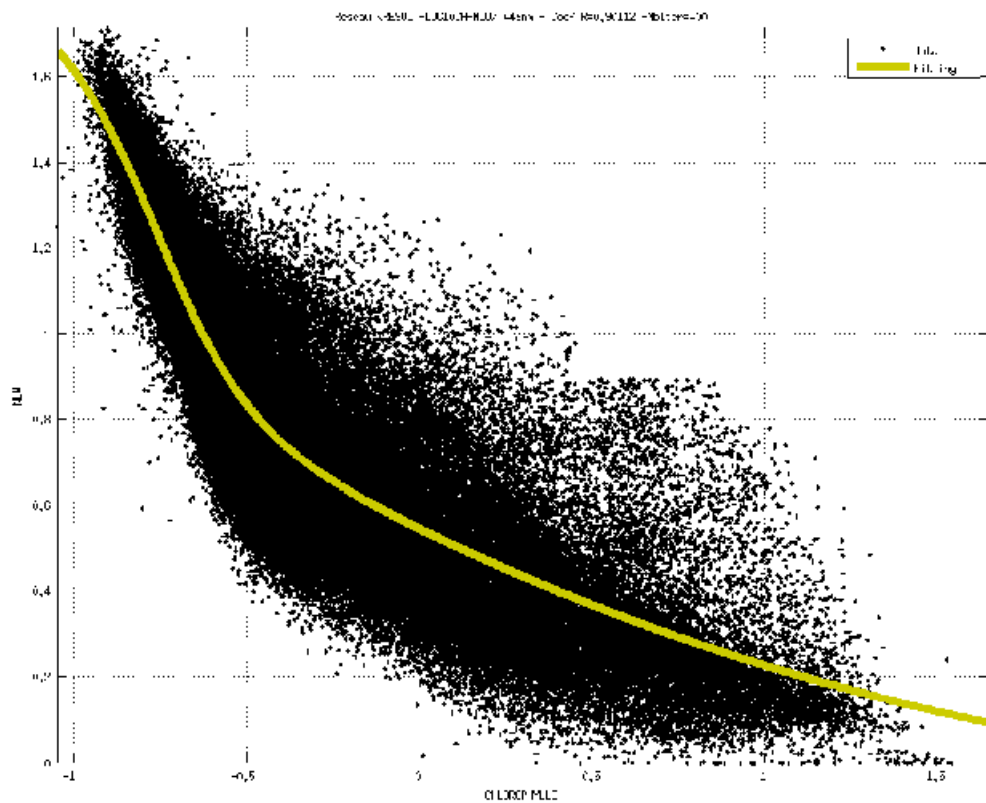
Figure 2 – Projection du reseau optimal sur les données tests. ($\lambda = 443nm$)

Tableau 3 – Résultats de l'apprentissage MLP et projection du reseau optimal sur les données tests.

nLw($\lambda=490nm$)					
Neurones	Iterations	RMS_All	RMS_App	RMS_Val	RMS_test
2	100	0,0969	0,0969	0,0970	0,0967
2	200	0,0969	0,0968	0,0972	0,0972
2	400	0,0969	0,0968	0,0966	0,0975
2	600	0,0969	0,0967	0,0974	0,0973
2	800	0,0969	0,0969	0,0973	0,0965
2	1000	0,0971	0,0970	0,0973	0,0975

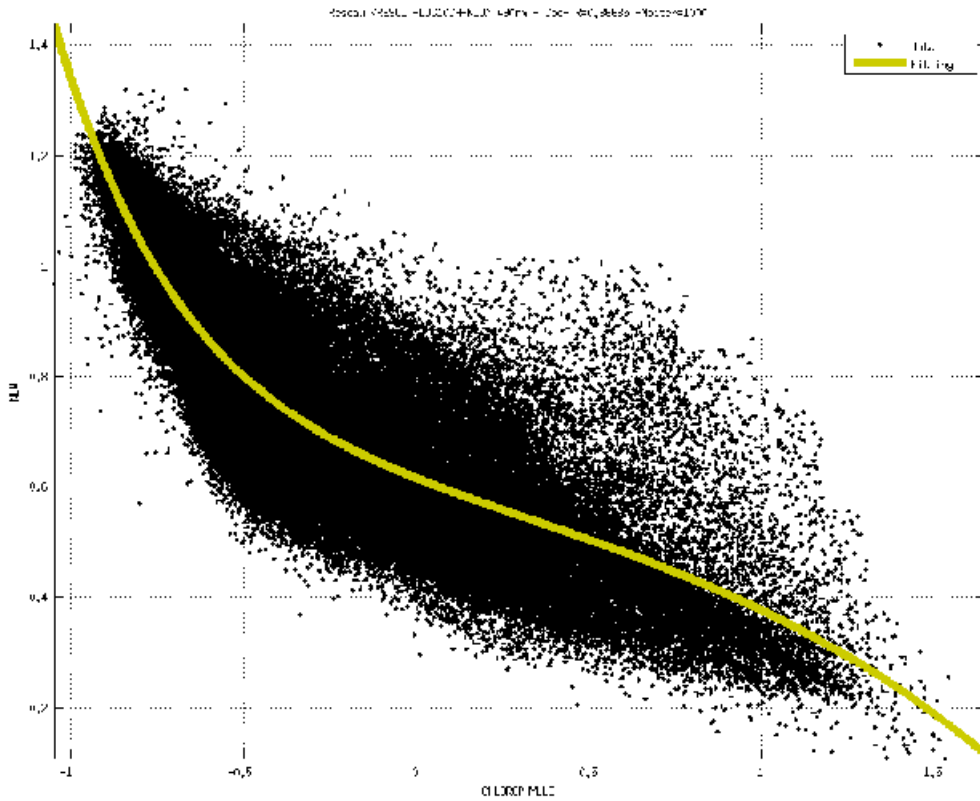
Figure 3 – Projection du reseau optimal sur les données tests. ($\lambda = 490nm$)

Tableau 4 – Résultats de l'apprentissage MLP et projection du reseau optimal sur les données tests.

nLw($\lambda=510nm$)					
Neurones	Iterations	RMS_All	RMS_App	RMS_Val	RMS_test
2	100	0,0705	0,0703	0,0712	0,0706
2	200	0,0705	0,0705	0,0703	0,0707
2	400	0,0705	0,0707	0,0706	0,0697
2	600	0,0705	0,0704	0,0703	0,0710
2	800	0,0705	0,0705	0,0705	0,0705
2	1000	0,0705	0,0704	0,0706	0,0707

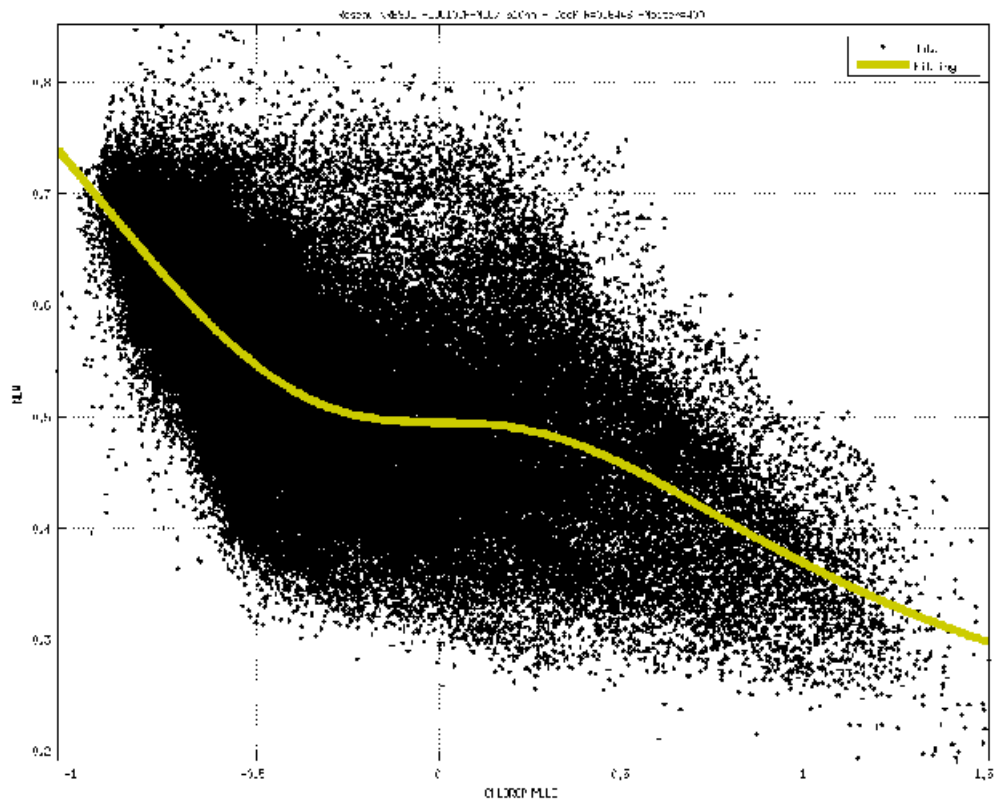
Figure 4 – Projection du reseau optimal sur les données tests. ($\lambda = 510nm$)

Tableau 5 – Résultats de l'apprentissage MLP et projection du reseau optimal sur les données tests.

nLw($\lambda=555nm$)					
Neurones	Iterations	RMS_All	RMS_App	RMS_Val	RMS_test
2	100	0,0534	0,0534	0,0531	0,0533
2	200	0,0536	0,0534	0,0544	0,0535
2	400	0,0534	0,0535	0,0534	0,0528
2	600	0,0536	0,0535	0,0536	0,0540
2	800	0,0536	0,0536	0,0538	0,0532
2	1000	0,0534	0,0531	0,0541	0,0541

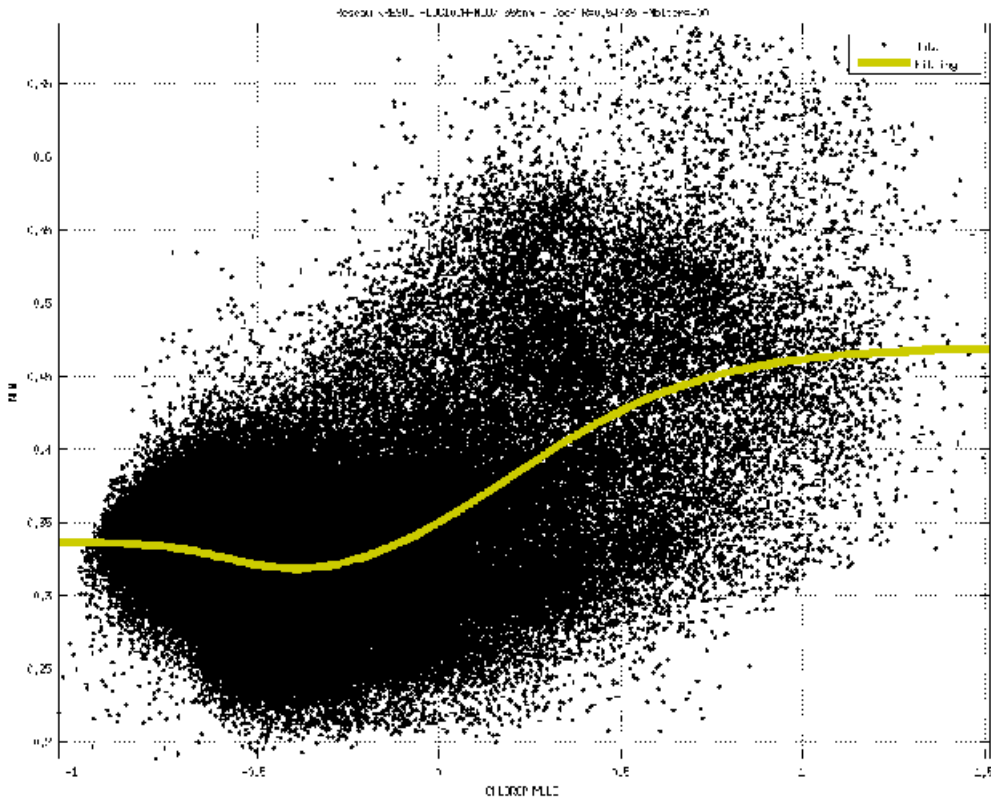
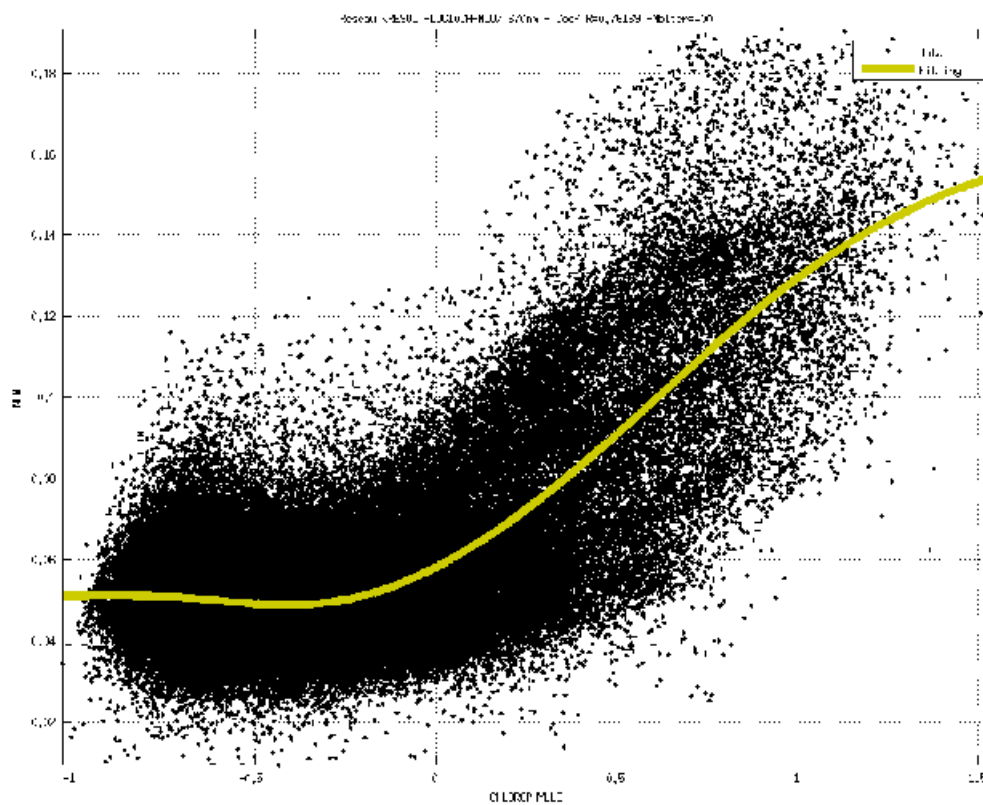
Figure 5 – Projection du reseau optimal sur les données tests. ($\lambda = 555nm$)

Tableau 6 – Résultats de l'apprentissage MLP et projection du réseau optimal sur les données tests.

nLw($\lambda=670nm$)					
Neurones	Iterations	RMS_All	RMS_App	RMS_Val	RMS_test
2	100	0,0144	0,0143	0,0144	0,0146
2	200	0,0144	0,0144	0,0144	0,0142
2	400	0,0144	0,0144	0,0143	0,0145
2	600	0,0144	0,0144	0,0143	0,0143
2	800	0,0144	0,0144	0,0144	0,0144
2	1000	0,0144	0,0144	0,0144	0,0142

Figure 6 – ($\lambda = 670nm$)

Liste des Figures

1.1	Concentration moyenne en chlorophylle-a en surface ($mg.m^{-3}$) sur la période 09/1997-08/2000 à partir des données satellite SeaWiFs (NASA Goddard Space Center Flight). Les quatre principaux EBUS sont indiqués : le Courant de Californie, le Courant de Humboldt, le Courant des Canaries et le Courant de Benguela. Les cartes regionales sont mises à la disposition par l'IRD/AOOS	12
1.2	Situation géographique de la zone d'étude : bathymétrie (en isoligne, en m), plateau continental (en vert)	14
1.3	Circulation atmosphérique méridienne montrant les cellules de Hadley et les alizés (Trades)	15
1.4	Circulation tropicale du vent en surface (a) en juillet (b) en janvier. (d'après Leroux, 1996)	16
1.5	Schéma de circulation de surface pour la côte Nord-Ouest Africaine (a) en hiver boréal, (b) au printemps, (c) en été et (d) en automne (Mittelstaedt, 1991)	17
1.6	Schéma de l'upwelling et de sa dépendance du rotationnel du vent au voisinage de la côte.	19
1.7	Images satellite SEAWIFS donnant la concentration en chlorophylle des eaux de surface. Données mensuelles (climatologiques) pour 1997 à 2007	20
1.8	Carte d'index d'upwelling, $10^{\circ}N$ et $35^{\circ}N$ pour la période 1969-1976	21
2.1	Exemple de sur-apprentissage d'un réseau approximant une fonction $\mathbf{y} = \mathbf{G}(\mathbf{x})$ sur un ensemble de données réelles générées par cette loi.	26
2.2	Comportement des erreurs en apprentissage et en validation au cours d'un apprentissage .	26
2.3	neurone artificiel	27
2.4	Traitement de données par réseau de neurones	28
2.5	Perceptron	29
2.6	Représentation matricielle du perceptron	29
2.7	Exemple de perceptron	30
2.8	Classification à partir du perceptron	30
2.9	Accélération de la convergence	35
2.10	Principe général de la modélisation : une observation z est associée à un indice c choisi parmi p l'aide de χ ; cet indice permet de définir le référent w_c (Dreyfus et al., 2002). . .	37
2.11	Architecture d'une carte topologique en $2 - D$. Le réseau est constitué de deux couches : une couche d'entrée qui sert à la présentation des observations et une couche d'adaptation (pour laquelle il faut définir un voisinage, figure 2.12) formée d'un treillis régulier à 2 dimensions dont chaque nœud est occupé par un neurone, qui est lui-même connecté à tous les éléments de la couche d'entrée. Chaque neurone c est affecté d'un référent w_c (Dreyfus et al., 2002).	37

2.12	Représentation de la topologie discrète d'une carte à deux dimensions constituée de $10 * 10$ neurones ; chaque point de la figure représente un neurone c . La distance δ entre deux neurones est définie sur le maillage. Elle permet de définir le voisinage d'ordre d (ici $d = 1, 2$ ou 3), qui représente l'ensemble des neurones dont la distance au neurone c est inférieure ou égale à d (Dreyfus et al., 2002).	38
2.13	Conservation de la topologie des données : deux neurones voisins dans C (distance discrète) doivent représenter deux ensembles d'observations voisines dans D (distance euclidienne) (Lachkar and Gruber, 2012).	39
2.14	Principe général de la modélisation par carte auto-organisatrice : une observation z est associée à un indice c choisi sur la carte (parmi p neurones) à l'aide de la fonction χ ; cet indice permet de définir le référent w_c (Dreyfus et al., 2002).	39
2.15	Adaptation du neurone gagnant et de son voisinage : plus un neurone est proche du neurone gagnant et plus il sera influencé (Lachkar and Gruber, 2012).	40
3.1	Principe de l'algorithme OC4V4. En abscisse on a le rapport de réflectance marine (en \log_{10}), en ordonnée la concentration en Chl-a en \log_{10})	49
4.1	Haut :-SST de Janvier 1999 et de Janvier 2007 ; Bas :-SST de Septembre 1999 et de Septembre 2007	54
4.2	Image de la Chl-a en Janvier 1999 et en Septembre 2006. On remarque une forte variation saisonnière et aussi une forte variation d'une année à autre.	55
4.3	Profils longitudinaux de température (haut) et de chlorophylle (bas) pour les mois de Février, Avril, Aout et Novembre à $17^\circ N$. On remarque la zone d'upwelling marquée par des températures froides et une forte concentration en Chl-a près de la côte et la forte variabilité saisonnière.	56
4.4	Mois de janvier-novembre (climatologie)	59
4.5	Cartes de distribution issues de la carte auto-organisatrice de Kohonen. Les cellules bleues foncées représentent des valeurs faibles, alors que les cellules rouges foncées (couleur chocolat) représentent des valeurs élevées. L'échelle d indique les valeurs transformées de l'unité de variables.	60
4.6	Résultat pour les six classes de SOM2 formés à partir de la carte auto-organisatrice de Kohonen (résultats bruts de l'analyse).	61
4.7	projection sur la carte géographique (climatologie :Janvier-Avril)	63
4.8	projection sur la carte géographique (climatologie :Mai-Aout)	63
4.9	projection sur la carte géographique (climatologie :Septembre-Décembre)	64
4.10	Box-plot représentant les statistiques descriptives des caractéristiques géophysiques de chacun de classes six. Les valeurs sont réparties de manière logarithmique sur l'axe des ordonnées pour les concentrations en chlorophylle-a.	64
4.11	Variation saisonnière des SST et ISST médianes des différentes classes	65
4.12	Variation saisonnière des concentrations médianes en Chl-a des différentes classes	66
4.13	Extension spatiale (nombre de pixels) des classes 5 et 6	67

4.14	Composante <i>N-S</i> du vent à Dakar. On remarque la forte variabilité saisonnière avec un maximum au mois de mars.	67
4.15	Carte topologique SOM3 apprise avec les 3 variables et décomposée en 6 classes (en haut à droite) et projection des 3 variables (SST, ISST et Chloro-a) sur la carte. On remarque que la topologie de la nouvelle carte SOM dénotée SOM3 est similaire à la	68
4.16	Boxplots pour Janvier et Aout de SOM3. Les structures des classes sont analogues à celles de SOM2	69
4.17	Répartition mensuelle des six classes de SOM3. On observe la quasi disparition de la classe 6 qui est celle de l'upwelling en Juillet Aout Septembre pendant la saison des pluies	70
4.18	séries temporelles des ISST	71
4.19	séries temporelles des SST	71
4.20	séries temporelles des Chl-a	72
4.21	SLA en Janvier 2003 et Septembre 2004	73
5.1	Boxplots montrant la distribution des classes, année 2002.	75
5.2	Données brutes mois de janvier	76
5.3	Données brutes mois de janvier (APRES EGALISATION)	77
5.4	Données filtrées mois de janvier (APRES EGALISATION)	78
5.5	tests du réseau [$nLw(\lambda=412nm)$]	80
5.6	spectres normalisés [$nLw^*(\lambda=412nm \text{ à } 555nm)$] des classes 1 à 6.	81
5.7	procédure de la normalisation des spectres [$nLw^*(\lambda=412nm \text{ à } 555nm)$] des classes 1 à 6.	82
6.1	processus de calcul de $nLw^*(\lambda)$	86
6.2	Cartes de distribution issues de la carte auto-organisatrice de Kohonen. Les cellules bleues foncées représentent des valeurs faibles, alors que les cellules rouges foncées(couleur chocolat) représentent des valeurs élevées. L'échelle d indique les valeurs transformées de l'unité $nLw^*(\lambda)$	87
6.3	Résultat pour les dix classes de $nLw^*(\lambda)$ formés à partir de la carte auto-organisatrice de Kohonen.	88
6.4	Représentation des dix spectres de $nLw^*(\lambda)$	89
6.5	Profondeur moyenne optique associée à chaque facteur de réflectance captée par chaque neurone de la carte SOM. La barre de couleur sur la droite est l'échelle la valeur AOT.	90
6.6	Pattern de AOT pour les dates : 16, 17, 24, 26, 28, 31 Janvier 2004	90
6.7	Carte de concentration en chlorophylle-a pour les date de : 16, 17, 24, 26, 28, 31 Janvier 2004	91
6.8	Patterns de spectres pour le 16, 17, 24, 26, 28, 31 Janvier 2003 pour lesquels une couverture AOT est très faible.	92
6.9	Patterns des spectres pour les journées du 16, 17, 24, 26, 28, 31 Janvier 2003 pour lesquels nous avons amélioré la classe 1 (D4, haptophytes), classe 5 (D1, diatomées), classe 4 (D3, SLC). Nous avons fusionné les classes 2, 3, 5 et 7 qui ne peuvent être déterminées maintenant en D2 et les classes 8, 9 et 10 sont associées aux eaux du cas-2 en D5.	93

Liste des Tableaux

3.1	Les bandes de SeaWiFS et leurs utilités.	50
4.1	Fichiers de base de données.	58
4.2	Statistiques descriptives des caractéristiques (Chl-a) de chacun des six classes.	62
4.3	Statistiques descriptives des caractéristiques (ISST) de chacun des six classes.	62
4.4	Statistiques descriptives des caractéristiques (Chl-a) de chacun des six classes.	62
4.5	Statistiques descriptives des caractéristiques (ISST) de chacun des six classes.	62
4.6	Classification générale en six classes (caractérisation des variables géophysiques)	65
5.1	Statistiques descriptives des caractéristiques (Chl-a) de chacune des six classes.	79

Bibliographie

- Allen, J. (1973). Upwelling and coastal jets in a continuously stratified ocean. *J. Phys.*
- Alvain, S., Moulin, C., Dandonneau, Y., and Bréon, F. (2005). Remote sensing of phytoplankton groups in case 1 waters from global seawifs imagery. *"Deep Sea Research Part I Oceanographic Research Papers 52, 11 1989-2004"*.
- Aristegui, J., Alvarez-Salgado, X., Barton, E., Figueiras, F., Hernandez-Leon, S., Roy, C., and Santos, A. (2004). Oceanography and fisheries of the canary current/iberian region of the eastern north atlantic, chapter 23 in the sea, vol. 14. *edited by A.R. Robinson and K.H. Brink, pp. 877-931, John Wiley and Sons, Inc., New York.*
- Babin, M., Stramski, D., Ferrari, G. M., Bricaud, H. C. A., Obolensky, G., and Hoepffner, N. (2003). Variations in the light absorption coefficients of phytoplankton, nonalgal particles, and dissolved organic matter in coastal waters around europe. *J. Geophys.Res.,108(C7), 3211. doi :10.1029/2001JC000882.*
- Badran, F., Yacoub, M., and Thiria, S. (2005). Self-organizing maps and unsupervised classification. *G. Dreyfus, Springer, Berlin Heidelberg New York.*
- Bakun, Andrew, S., C., and Nelson (1991). The seasonal cycle of wind-stress curl in subtropical eastern boundary current regions. *J. Phys. Oceanogr., 21, 1815–1834.*
- Barth, J. (1989). Stability of a coastal upwelling front 2. model results and comparison with observations. *J. Geophys. Res.*
- Barton, E. D. (1998). Eastern boundary of the north atlantic. *A.R. Robinson and K.H. Brink, John Wiley and Sons, Inc., New York.*
- Beletsky, D., O'Connor, W., Schwab, D., and Dietrich, D. (1997). Numerical simulation of internal kelvin waves and coastal upwelling fronts-. *J. Phys. Oceanogr, pp.1197-1215.*
- Bishop, C. (1995). Neural networks for pattern recognition. *Oxford : Oxford University Press.*
- Bonn, F. and Rochon, G. (1992). Précis de télédétection ; vol. 1 : Principes et méthodes. *Presse de l'Université du Québec et AUPELF, Sainte-Foy, 485 p.*
- Borme, P., Benrejeb, M., and Haggrège, J. (2007). Les réseaux de neurones. présentation et application. *Edition Technosup.*
- Brajard, J. (2006). Méthodologie neuronale pour l'inversion des signaux satellitaires de couleur de l'océan. traitement des aérosols absorbants et restitution de la concentration en chlorophylle-a. *Thèse de doctorat, Université Paris 6.*

- Bricaud, A., A. Morel, and André, J. (1987). Spatial temporal variability of algal biomass in the mauritanian upwelling zone, as estimated from czcs data. *Adv. Space Res.*
- Campbell, J., Blaisdell, J., and Daryi, M. (1995). Level-3 seawifs data products : Spatial and temporal binning algorithms. *NASA Tech. Memo, 104566, vol.36.*
- Caroli, E. (1999). Analyse numérique de la morphologie des marges continentales et implication de l'évolution du niveau marin sur l'immersion des plate formes continentales est-africaines. *Thèse de Doctotat, Ecole Normale Supérieure de Lyon, Bordeaux.*
- Crepon, M., Richez, C., and Chartier, M. (1984). Effect of coastline geometry on upwellings. *J. Phys. Oceanogr., pp. 1365-1382.*
- Cury, P. and Roy, C. (1991). *Orstom Editions, 38-66.*
- Demarcq, H. and Faure, V. (2000). Coastal upwelling and associated retention indices from satellite sst. application to octopus vulgaris recruitment. *Oceano. Acta, 23 , p.391-407.*
- Diouf, D., Niang, A., Thiria, S., Brajard, J., Crepon, M., and Sawadogo, S. (2011). Onitoring aerosol optical thickness and chlorophyll-a to the west africa coast using a neural-variational method. *Journal des Sciences Pour l'Ingénieur, Vol. 13, pp. 46-51. ISSN : 0851-4453.*
- Dreyfus, G., Martinez, J., Samuelides, M., Gordon, M., Badran, F., Thiria, S., and Herault, L. (2002). Réseaux de neurones, méthodologie et applications. *Eyrolles, collection Algorithmes.*
- Dubruvin (1991). La zze mauritanienne et son environnement géographique géomorphologique et hydroclimatique. *In chavance, P.M Girardin.*
- Echevin, V., Puillat, I., Grados, C., and Dewitte., B. (2004). Seasonal and mesoscale variability in the peru upwelling system from in situ data during the years 2000 to 2004. *Poster, Proceedings.,(Guayana, Vol. 68 (I), No2, 167-173).*
- Ekman, V. W. (1905). On the influence of the earth's rotation on ocean currents, volume 11. *Astr. o. Fysik (Stockholm). 53 pp.*
- Farikou, O., Sawadogo, S., Niang, A., Brajard, J., Crépon, M., and Thiria, S. (2011). Variabilité spatio-temporelle de l'upwelling sénégal-mauritanien à partir d'observations satellitaires. *Journal des Sciences Pour l'Ingénieur, Vol. 13, pp. 60-71. ISSN : 0851-4453.*
- Fischer, G., Donner, B., Ratmeyer, V., Davenport, R., and Wefer., G. (1996). Distinct year to year particle flux variations off cape blanc during 1988-1991 : Relation to oodeduced sea surface temperatures and trade winds. *Journal of Marine Research, 54, 73-98.*
- Gabric, A. J. L., Camp, L. V., Nykjaer, L., Eifler, W., and Schrimpf, W. (1993). Offshore export of shelf production in the cape blanc (mauritania) giant filament as derived from coastal zone color scanner imagery. *Journal of Geophysical Research, 98, 4697- 4712.*
- Giraud, X. (2001). Reconstitution paléocéanographique du signal $\delta^{15}N$: modélisation couplée physique et biogéochimique d'un upwelling côtier. *Thèse de Doctorat, Université de Bordeaux I.*
- Giraud, X., Bertrand, P., Garçon, V., and Dadou, I. (2000). Modeling $\delta^{15}N$ evolution : First palaeoceanographic applications in a coastal upwelling system. *Journal of Marine Research, 58, 609-630.*

- Gordon, H., Brown, O., Evans, R., Brown, J., Smith, R., Baker, K., and Clark, D. (1988). A semi-analytical radiance model of ocean color. *Journal of Geophysical Research*, vol. 93, p. 10909-10924.
- Gordon, H. R. and Wang, M. (1994). Retrieval of water-leaving radiances and aerosol optical thickness over the oceans with seawifs : A preliminary algorithm. *Appl. Opt.*, 33(3), 443-453.
- Halpern, D. (2002). Offshore ekman transport and ekman pumping off peru during the 1997-1998 el niño. *Geophys. Res. Lett.*, 29, 19-1, 19-3.r
- Hooker, S., Firestone, E., and Acker, J. Eds. *NASA Goddard Space Flight Center, Greenbelt, Maryland, 73 pp., plus color plates.*
- Hopfield, J. (1982). Neural networks and physical systems with emergent computational abilities. *Proceedings of the National Academy of Sciences*, 79 :2554.
- Jacques, G. and Tréguer, P. (1986). Ecosystèmes pélagiques marins. *Masson, 243 pp.*
- Jain, A. and Dubes, R. (1998). Algorithms for clustering data. *Englewood. Cliffs, New Jersey, USA : Prentice Hall Advanced Reference Series, 320 pp.*
- Kohonen, T. (1982). Self-organized formation of topologically correct feature maps. *Biological Cybernetics* 43, pp. 59–69.
- Kohonen, T. (2001). Self organizing maps. *Springer, Berlin, Heidelberg, New York, Tokyo, 459pp.*
- Lachkar, Z. and Gruber, N. (2012). A comparative study of biological production in eastern boundary upwelling systems using an artificial neural network. *Biogeosciences*, 9, 293–308, 2012.
- Lange, C., Romero, O., Wefer, G., and Gabric, A. (1998). Offshore influence of coastal upwelling off mauritania, nw africa, as recorded by diatoms. *in - Deep-Sea Res. Part I*, pp. 985-1013.
- Lathuiliere, C., Echevin, V., and Levy, M. (2008). Seasonal and intraseasonal surface chlorophyll-a variability along the northwest african coast. *J. Geophys. Res.*, 113, C05007, doi :10.1029/2007JC004433.
- Leroux, M. (1996). La dynamique du temps et du climat. *Enseignement des Sciences de la Terre. Masson, Paris, 310 pp.*
- Maigret, J. (1972). Campagne expérimentale de pêche des sardinelles et autres espèces pélagiques (juillet, 1970- octobre 1971). tome i. *Observation concernant l'océanographie et la biologie des espèces, Soc. Centr. Equip. Territ. Intern. ; mars 1972, 143.*
- Marie, C. (2007). Contribution à l'étude de l'upwelling du pérou chili. *Thèse de Doctotat, Université Toulouse III-Paul Sabatier.*
- McClain, C. R., Yeh, E., and Fu, G. (1992). An analysis of gac sampling algorithms : A case study. nasa technical memorandum 104,566, vol. 4, s.b hooker and e.r firestones. Eds., *NASA Goddard Space Flight Center, Greenbelt, Maryland, 22 pp., plus color plates.*
- McCulloch, W. S. and Pitts, W. H. (1943). A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics*, 5 :115-133.
- Mittelstaedt, E. (1991). The ocean boundary along the northwest african coast : Circulation and oceanographic properties at the sea surface. *Prog. Oceanogr.*, 26, 307– 355.

- Mobley, C. (1999). Estimation of the remote-sensing reflectance from above-surface measurements. *Applied Optics*, vol. 38, p. 7442 - 7455.
- Morel, A. (1988). Optical modeling of the upper ocean in relation to its biogenous matter content (case 1 waters). *Journal of Geophysical Research*, 106, 7163-7180.
- Morel, A. and Maritorena (2001). Bio-optical properties of oceans waters : A reappraisal. *Journal of Geophysical Research*, 106, 7163-7180.
- Mueller, J. and L. (2000). Seawifs algorithm for the diffuse attenuation coefficient $k(490)$ using water-leaving radiances at 490 and 555 nm. seawifs postlaunch calibration and validation analyses, part 3. *SeaWIFS Postlaunch Technical Report Series*, vol. 11, 2000.
- Myint, S. and Walker, N. (2002). Quantification of surface suspended sediments along a river dominated coast with noaa avhrr and seawifs measurements. *Louisiana, USA. Int. J. Remote Sensing*, 2002, vol. 23, No 16, p. 3229-3249.
- Niang, A., Badran, F., Moulin, C., Crepon, M., and 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*, pp. 82-94.
- Niang, A., Gross, L., Thiria, S., Badran, F., and Moulin, C. (2003). Automatic neural classification of ocean colour reflectance spectra at the top of atmosphere with introduction of expert knowledge. *Remote Sensing of Environment*, pp. 257-271.
- O'Brien, J. and Hurlburt, H. E. (1972). A numerical model of coastal upwelling - j. phys. *Oceanogr*, pp. 14-26.
- O'Reilly, J., S.Maritorena, Mitchell, B., Siegel, D., Carder, K., Garver, S., and McClain, K. M. (1998). Ocean color chlorophyll algorithms for seawifs. *Journal or geophysical research*, 103 :24,937-24,953.
- Postel, E. (1955). Contribution à l'étude de la biologie de quelques scambridale de l'atlantique tropicooriental. *Ann. Stn. Océanogr. Salammbô 10 : 167pp.*
- Pradhana, Y., Lavendera, S. J., Hardman-Mountfordb, N., and Aikenb, J. (2006). Seasonal and inter-annual variability of chlorophyll-a concentration in the mauritanian upwelling : Observation of an anomalous event during 1998–1999. *Deep-Sea Research II 53 (2006) 1548–1559.*
- Rebert, J. P. (1983). Hydrographie et dynamique du plateau continental sénégalais. *Doc. Sci. Centr. Rech océanogr. Dakar Thiaroye*, 89,99p.
- Renault, L. (2008). Impact des jets côtiers atmosphériques sur l'upwelling du système de courants de humboldt. *Thèse de doctorat, Université Toulouse III-Paul Sabatier.*
- Roy, C. (1991). Les upwellings : Le cadre physique des pêcheries côtières ouest africaines. *Pêcherie ouest Africaine, Variabilité, instabilité et échangeement.*
- Roy, C. (1992). Réponse des stocks de poissons pélagiques à la dynamique des upwellings en afrique de l'ouest : Analyse et modélisation. *Orstom édition, Thèse 146p.*
- Roy, C. and Cury, P. (1988). Migration saisonnière du thiof (*epinepnhis aeneus*) au sénégal : influence des upwellings sénégalais et mauritaniens. *Océanologica Acta*, 11, 25-36.

- Rumelhart, D., Hinton, G., and Williams, R. (1986). Learning internal representations by error propagation. In D.E. Rumelhart and J.L. McClelland (Eds.), *Parallel distributed processing* (pp. 318-362). Cambridge, MA : MIT Press.
- Sanyo (2002). Etude pour le plan d'aménagement des ressources halieutique en république islamique de la mauritanie techno marine, inc. sanyo overseas agro-fisheries consultants co. ltd. *Chapitre 1-9 : pages variables.*
- Sarangi, R., Chauhan, K., and Nayak, P. (2002). Vertical diffuse attenuation coefficient (kd) based optical classification of irs-p3 mos-b satellite ocean colour 52 data. *Proc. Indian Acad. Sci. (Earth Planet. Sci.), 111, No 3, September 2002, pp. 237-245.*
- Sawadogo, S., Brajard, J., Niang, A., Lathuiliere, C., Crépon, M., and Thiria, S. (2009). Analysis of the senegalo-mauritanian upwelling by processing satellite remote sensing observations with topological maps. *The 2009 International Joint Conference on Neural Networks.*
- Speth, P. and Detlefsen, H. (1982.). Meteorological influences on upwelling off northwest africa. 180, *Conseil International pour l'Exploration de la Mer.*
- Thiria, S., Mejia, C., Badran, F., and Crepon, M. (1993). A neural network approach for modelling non linear transfer functions : Application for wind retrieval from spaceborne scatterometer data. *J. Geophys. Res. 98, 22827-22841.*
- Toure, D. and Deme, I. G. (1991). variations spatio-temporelles de la biomasse phytoplanctonique sur le plateau continental sénégalais. *Cury Philippe (ED.). Pêcheries ouest africaines : variabilité, instabilité et changement. Paris : ORSTOM, (1991). p. 90-102. Groupes de Travail, Dakar ; Casablanca (SEN ; MAR), 1988/12/12-17 ; 1989/06/12-17.*
- Wang, C. and Carton, S. X. J. (2004). global survey of ocean-atmosphere and climate variability. *Earth climate : The Ocean-Atmosphere Interaction, C. Wang, S.P. Xie, and J.A. Carton (eds.), Geophys. Monograph, 147, AGU, Washington D.C., 1-19.*
- Widrow, B. and Hoff, M. E. (1960). Adaptative switching circuit, wescon convention record, part 4, new york. *Institute of Radio Engineers, 96-104.*
- Yamagata, T. and Lizuka, S. (1995). Simulation of tropical thermal domes in the atlantic : A seasonal cycle. *J. Phys. Oceanogr., 25 (9), 2129-2140.*
- Yang, H. and Gordon, H. (1997). Remote sensing of ocean color :assessment of water-leaving radiance bi-directional effects on atmospheric diffuse transmittance. *Applied Optics, vol. 36, p. 7887-7897.*