

Université du Québec à Rimouski (UQAR)

et

Université Pierre et Marie-Curie (Paris VI)

THÈSE DE DOCTORAT

spécialité: océanographie

présentée par

Martin DESRUISSEAUX

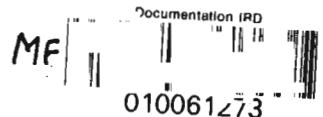
Pertinence des données altimétriques en halieutique appliquées à la pêche thonière

Jury: Jean-Claude BRÊTHES
Jean-François DUMAIS
Paul NIVAL
Michel PETIT
Antonio RAMOS
Jean-Michel STRETTA

31 mai 2006



1er septembre 2004



Résumé

Les thons se situent près du sommet de la pyramide alimentaire des océans pélagiques, et leur nourriture est concentrée dans des zones restreintes. Pour la trouver, ces poissons doivent parcourir de grandes distances. Ils tendent à se concentrer dans des régions océaniques où les conditions environnementales sont à l'intérieur de leurs limites physiologiques et la probabilité de trouver de la nourriture plus élevée. Les auteurs dans la littérature s'accordent à dire que les conditions de température et d'oxygène influencent fortement les mouvements, la distribution et la capturabilité des thons. Il est admis que les thons tendent à se concentrer le long de discontinuités thermiques telles que les fronts océaniques. La littérature rapporte plusieurs expériences utilisant les images de température de surface de la mer, ou celles de la concentration en chlorophylle-*a*, pour en déduire les positions où la présence de thon est plus probable.

Les études rapportées dans la littérature explorent souvent les relations entre les concentrations de poissons et un paramètre environnemental pris individuellement. D'autres études ont montré l'intérêt d'utiliser simultanément plusieurs paramètres. Cette thèse cherche à estimer la pertinence du paramètre altimétrique (anomalie de la hauteur de l'eau) par rapport aux autres paramètres.

La région d'étude (le sud-ouest de l'océan Indien) nous fournit une abondance exceptionnelle de données. Nous disposons d'environ 150 000 données de pêches à la senne sur une période allant de 1991 à 2001 inclusivement. Ces données ont été fournies à l'IRD (Institut de Recherche pour le Développement) par les pêcheurs en échange d'images de température de surface de résolution LAC (*Local Area Coverage*) captées par l'antenne de l'IRD sur l'île de la Réunion. Outre ces images de température, nous disposons dans cette même région d'images de concentrations en chlorophylle-*a*, d'anomalie de la hauteur de l'eau et de pompage d'Ekman.

Pour cette thèse, nous définissons un *descripteur du paysage océanique* comme une combinaison d'un paramètre environnemental (par exemple la température), une opération (par exemple le gradient) et une position spatio-temporelle relative (par exemple la valeur 15 jours avant la pêche). Nous avons ensuite exploré des corrélations statistiques entre ces descripteurs et la présence ou l'absence de thons en bancs libres observés par les pêcheurs. Nous n'avons pas pris en compte les quantités pêchées afin d'atténuer les variations dues aux différences d'habiletés des pêcheurs. Nous n'avons pas non plus distingué les différentes espèces de thon ; nous n'avons pris en compte que leur classement en deux catégories selon que les individus avaient plus ou moins environ 10 kg.

La problématique halieutique de cette thèse, à savoir mettre en évidence les relations entre les données de pêches et les données environnementales, nécessite le développement d'outils d'informations géographiques. Ce domaine informatique est moins mature que celui des bases de données par exemple. Le domaine géospatial fait l'objet de standards internationaux émergeant, grâce entre autres aux efforts du consortium OpenGIS®. Les travaux de cette thèse s'appuient largement sur ces standards et ont été effectués en partie sous forme de participations à des projets libres (*Open Source*), notamment Geotools 2.

Nos résultats suggèrent que les concentrations en chlorophylle-*a* et l'anomalie de la hauteur de l'eau, lorsque utilisées séparément, sont d'une efficacité similaire pour expliquer la présence ou l'absence de thon. Leur utilisation conjointe dans un modèle linéaire accroît encore leur efficacité. En revanche, l'apport de la température de surface à ces modèles ne semble pas significatif. Cette dernière observation est à prendre avec précaution puisque le rôle de la température peut être moins direct que les descripteurs que nous avons testés (par exemple les thons se trouvent généralement à proximité d'un gradient thermique plutôt que directement dans celui-ci).

Nos résultats montrent aussi que les corrélations entre la présence de thons et les descripteurs du paysage océanique apparaissent significativement non-linéaires. Les réseaux de neurones artificiels de type *sigma-pi* semblent alors appropriés pour modéliser ces corrélations.

Remerciements

Je remercie l'unité ESPACE de l'IRD : Frédéric Huyn pour m'avoir accueilli dans l'unité qu'il dirige et Michel Petit pour m'avoir proposé ce sujet de thèse, ainsi que Jean-Michel Stretta pour avoir accepté le rôle de directeur de thèse du côté de l'université Pierre et Marie Curie.

Je tiens à remercier tout particulièrement Jean-Claude Brêthes, mon directeur de thèse du côté de l'université du Québec à Rimouski (UQAR). Je ne crois pas exagérer en disant que sans son support décisif, cette thèse n'aurait probablement jamais été menée à son terme.

Lors du démarrage de la thèse, Laurent Dagorn m'a été d'une aide précieuse à la fois par ses interventions durant les comités de pilotages (interventions qui ont influencées la discussion sur les perspectives de ce travail), et par ses références bibliographiques. Jean-Michel Stretta m'a été d'un grand secours lors de l'examen doctoral de l'hiver 2000, et Michel Petit d'un support constant durant les cinq années. Il m'a aussi été d'une aide considérable lors du bouclage final de la partie halieutique de la thèse.

Je remercie Guillaume de Magny (compagnon doctorant) et le professeur Anatoli Terekhin (professeur de l'université Lomonosov de Moscou) qui ont effectué les analyses préliminaires qui ont ouvert la voie au chapitre statistiques de cette thèse. Je remercie aussi Morgan Mangeas et Touraïvane, de l'unité ESPACE en Nouvelle-Calédonie, pour leur aide sur les chapitres statistique et informatique respectivement. Étant donné la place de l'analyse croisée des données pour cette thèse (un domaine qui n'est pas ma spécialité), les conseils et les clarifications apportés par Morgan ont été précieux. Du côté de Montpellier, Frédéric Huynh, Isabelle Pelissier et Monique Martiny ont fait tout ce qui était possible pour me donner les meilleures conditions de travail.

Je remercie tous les membres du jury, dont Jean-François Dumais, Antonio Ramos et Paul Nival, d'avoir accepté d'évaluer ma thèse, et pour plusieurs de s'être déplacé d'autant loin pour assister à la soutenance. Antonio m'a donné plusieurs conseils éclairés, et sa discussion a été des plus stimulantes. Paul Nival a eu la gentillesse de m'accepter dans son école doctorale, même lorsque je dépassais les échéanciers. Alain Nierga du côté de Paris VI, et Ghislaine Férand-Pouget, Isabelle Pelissier et Monique Martiny du côté de l'IRD-Montpellier, ont considérablement facilité (voir rendu possibles) les démarches administratives.

Je remercie Renaud Pianet qui nous a fournit les données des senneurs français, et Mme Pilar Pallares de l'IEO qui nous a fournit les données de pêche espagnols. Ces remerciements s'étendent à l'ensemble des patrons de pêche et des équipes de terrain de l'IEO, l'IRD et de la SFA (*Seychelles Fishing Authority*) qui ont communiqué, collecté, vérifié et traité ces données.

Mes remerciements les plus amicaux s'adressent à l'ensemble de l'équipe IRD de la Maison de la Télédétection de Montpellier (et parfois d'ailleurs) pour leur accueil et leur support sans lequel je n'aurais pas tenu jusqu'au bout : Alice Servenay, Guillaume de Magny, Christophe Boschet, Florence Lahet, Vincent Heurteaux, Isabelle Pelissier, Monique Martiny, Frédéric Huynh, Michel Petit, Antonio Ramos, Laurent Demagistri, Rémi Ève, Pascale Couprie, Marc Lointier, Nicolas Devaux, Cécile Adell, Marc Despinoy, Aurélie Moulins et bien d'autres.

Table des matières

Chapitre 1

Océanographie des pêches thonières.....	9
1.1 Les enjeux de la pêche thonière.....	10
1.1.1 Les espèces concernées.....	11
1.1.2 Le gisement thonier de l'océan Indien.....	13
1.1.3 Le système pêche et la télédétection.....	15
1.2 Apport de la télédétection dans le processus décisionnel.....	19
1.2.1 Recherche directe d'une ressource halieutique.....	19
1.2.2 Recherche indirecte d'une ressource halieutique.....	19
1.2.3 Suivi de la maturation des masses d'eaux.....	20
1.2.4 Description du paysage océanique.....	21
1.3 Les données disponibles sur les mouvements des thonidés.....	22
1.3.1 Données de marquages.....	22
1.3.2 Données de pêches.....	22
1.4 Synopsis d'écologie thonière.....	24
1.4.1 Adaptations physiologiques et comportementales aux facteurs abiotiques.....	24
1.4.2 Adaptations comportementales aux facteurs biotiques.....	25
1.4.3 Relations entre paramètres environnementaux et distribution.....	26
1.4.4 Dispositifs concentrateurs de poissons (DCP).....	26
1.4.5 Comportement gréginaire.....	27
1.4.6 Reproduction.....	28
1.5 Synopsis des courants dans l'océan Indien.....	29
1.5.1 Composantes du courant net.....	31
1.5.2 Les remontées d'eaux.....	31

Chapitre 2

Présentation des sources de donnée utilisées.....	32
2.1 Les données halieutiques.....	33
2.2 Les données satellitaires.....	35
2.2.1 La station SEAS (Surveillance de l'Environnement Assistée par Satellite).....	36
2.2.2 La topographie de la surface de l'eau.....	36

Chapitre 3

Élaboration des outils d'investigation.....	42
3.1 Les services logiciels.....	43
3.1.1 Problématique.....	43
3.1.2 Interopérabilité.....	44
3.1.3 Apport du consortium OpenGIS®.....	44
3.1.4 Quelques spécifications.....	45
3.1.5 Apport des implémentations libres.....	46
3.1.6 Les projets Seagis, Geotools et GeoAPI.....	47
3.2 Architecture de l'afficheur d'informations géographiques.....	49
3.2.1 Abstraction d'un tableau de coordonnées (PointArray).....	49
3.2.2 Encapsulation d'un tableau de coordonnées (LineString).....	51
3.2.3 Liste chaînée de séquences de segments de droites (Polyline / Polygon).....	51
3.3 Systèmes de référence des coordonnées.....	56
3.3.1 Conversions et transformations de coordonnées.....	58
3.3.2 Création des systèmes de référence des coordonnées.....	59
3.3.3 Systèmes et transformations de coordonnées distribuées.....	60
3.4 Matrices de données (images).....	63
3.4.1 Conversion des valeurs de pixels vers les valeurs géophysiques.....	64
3.4.2 Opérations sur les matrices de données.....	65
3.4.3 Limites des spécifications du consortium OpenGIS.....	69
3.4.4 Extensions aux spécifications du consortium OpenGIS.....	70

3.5 Les bases de données.....	74
3.5.1 Termes employés pour désigner les données.....	75
3.5.2 Structure de la base de métadonnées satellitaires.....	75
3.5.3 Structure de la base de données des captures.....	78
3.5.4 Interfaces de programmation accédant aux bases de données.....	83
3.5.5 Enregistrements des activités dans un journal.....	87
Chapitre 4	
Corrélations et modèles.....	88
4.1 Variables utilisées et hypothèses.....	89
4.1.1 Variables à expliquer.....	90
4.1.2 Variables explicatives.....	91
4.1.3 Pertinence des interpolations des données environnementales.....	91
4.2 Distribution des données.....	95
4.2.1 Distribution géographique des données de prospection.....	96
4.2.2 Distribution des données environnementales aux positions de prospection.....	96
4.3 Corrélations entre paramètres environnementaux et captures.....	100
4.4 Analyses factorielles discriminantes.....	105
4.4.1 Choix des descripteurs du paysage océanique (variables explicatives).....	105
4.4.2 Choix des données de pêches (variables expliquées).....	107
4.4.3 Résultats des modèles linéaires pour les thons > 10 kg.....	107
4.5 Modélisation non-linéaire.....	133
4.5.1 Les réseaux de neurones artificiels.....	133
4.5.2 Apprentissage supervisé du réseau.....	135
4.5.3 Application d'un réseau de neurones artificiels aux données de pêches.....	136
Chapitre 5	
Résumé et conclusion.....	139
5.1 Résumé des résultats statistiques.....	140
5.2 Travaux futurs.....	144
5.2.1 Nouveaux descripteurs du paysage océanique à considérer.....	144
5.2.2 Réseaux bayésiens.....	144

Index des tables

Tableau 1: Espèces de thons rencontrées dans les océans tropicaux et tempérés.....	11
Tableau 2: Quelques expériences et applications de la télédétection en halieutique.....	17
Tableau 3: Paramètres de l'ellipsoïde WGS 1984 et de celle de TOPEX/POSÉIDON.....	37
Tableau 4: Systèmes de référence des coordonnées à différentes étapes de l'affichage.....	53
Tableau 5: Signification des valeurs de pixels des images de température de la Réunion.....	65
Tableau 6: Principaux attributs de SampleDimension donnant les informations du tableau 5.....	65
Tableau 7: Opérateurs de convolutions utilisés pour les calculs de magnitude du gradient.....	68
Tableau 8: Opérateur similaire à Sobel étendu à une taille de 9×9	68
Tableau 9: Exemple de table de correspondances pour les images de température.....	71
Tableau 10: Interfaces d'OpenGIS reflétées dans la base de données.....	76
Tableau 11: Principaux paramètres dans la base de méta-données satellitaires.....	77
Tableau 12: Quelques codes de la FAO (extrait de la table Espèces).....	79
Tableau 13: Paramètres environnementaux dans la base de données des captures.....	80
Tableau 14: Chaînes d'opérations pour chaque colonne de la table Environnements.....	81
Tableau 15: Associations des bancs notées dans la base de données.....	82
Tableau 16: Codes des espèces et catégories de taille utilisés comme variables expliquées.....	90
Tableau 17: Exemples de descripteurs retenus pour les modèles linéaires.....	106
Tableau 18: Transformations de variables appliquées sur les gradients.....	106
Tableau 19: Coefficients du modèle linéaire utilisant la concentration en chlorophylle-a.....	114
Tableau 20: Coefficients du modèle linéaire utilisant la température GAC.....	117
Tableau 21: Coefficients du modèle linéaire utilisant la température LAC.....	121
Tableau 22: Coefficients du modèle linéaire utilisant l'anomalie de la hauteur de l'eau.....	123
Tableau 23: Coefficients d'un modèle linéaire utilisant les données de CHL et de SLA.....	125
Tableau 24: Coefficients d'un modèle linéaire utilisant l'ensemble des paramètres.....	126
Tableau 25: Coefficients d'un modèle linéaire utilisant la température GAC et la SLA.....	127
Tableau 26: Coefficients d'un modèle alternatif utilisant la température GAC et la SLA.....	127
Tableau 27: Coefficients d'un modèle linéaire utilisant l'ensemble des paramètres.....	129
Tableau 28: Coefficients du modèle linéaire utilisant l'ensemble des paramètres.....	130
Tableau 29 : Sommaire des corrélations de Spearman de la figure 35.....	140
Tableau 30: Sommaire des taux de succès des modèles.....	141

Index des illustrations

Figure 1: Distribution des principales espèces de thons (source: FAO2).....	11
Figure 2: Prises annuelles mondiales des espèces commerciales de thons.....	13
Figure 3: Prises annuelles dans l'océan Indien des espèces commerciales de thons.....	13
Figure 4: Prises annuelles des espèces commerciales de thons par océan.....	14
Figure 5: Circulation dans l'océan Indien lors de la mousson d'été.....	30
Figure 6: Circulation dans l'océan Indien lors de la mousson d'hiver.....	30
Figure 7: Principe des mesures altimétriques (source: AVISO).....	37
Figure 8: Schéma du satellite topex/poseidon (source: AVISO).....	38
Figure 9: Trace au sol du satellite topex/poseidon (source: AVISO).....	39
Figure 10: Géométrie d'un pulse d'altimètre sur la surface de la mer.....	40
Figure 11: Services logiciels requis (de gauche à droite), des standards jusqu'aux applications.....	43
Figure 12: Spécifications d'implémentation d'OpenGIS utilisées dans cette étude (version 1.0).....	45
Figure 13: Positions directes et relatives des points de contrôles d'un polygone.....	50
Figure 14: Hiérarchie simplifiée des principales classes de systèmes de référence des coordonnées.....	57
Figure 15: Système de référence des coordonnées 3D (x,y,t) utilisé.....	58
Figure 16: Exemple de chaîne des opérations.....	59
Figure 17: Flux de données à travers les fabriques de systèmes de référence des coordonnées.....	60
Figure 18: Diagramme simplifié de GridCoverage et de ses dépendances dans Geotools 2.....	63
Figure 19: Aperçu des outils et de leurs inter-connections.....	74
Figure 20: Structure de la base de données d'images.....	76
Figure 21: Table des formats, avec bandes (SampleDimensions) et catégories.....	78
Figure 22: Structure partielle de la base de données des captures.....	78
Figure 23: Exemple d'enregistrements de la table Environnements.....	81
Figure 24: Agencement des drapeaux visité, présence et libre dans les tables physico-halieutes.....	82
Figure 25: Illustration du séquençage des classes CoverageDataBase → CoverageEntry.....	83
Figure 26: Captures par espèces et proportions de coups de sennes non-nuls.....	90
Figure 27: Déplacements maximums des bateaux par rapport à la position de la veille.....	91
Figure 28: Températures GAC en fonction des températures LAC, avec et sans interpolations.....	93
Figure 29: Comparaison des valeurs avec et sans interpolation.....	93
Figure 30: Améliorations des corrélations de Spearman apportés par les interpolations.....	94
Figure 31: Proportions des coups de sennes contenant les catégories étudiées.....	95
Figure 32: Distribution du tonnage des calées.....	95
Figure 33: Nombre de données de prospection par demi-degré d'angle.....	96
Figure 34: Distribution des paramètres environnementaux des bancs libres le jour des prospections.....	98
Figure 35: Corrélations de Spearman entre les captures et différents paramètres environnementaux.....	102
Figure 36: Comparaison des corrélations de Spearman entre températures lac et gac.....	104
Figure 37: Proportion des présences en fonction des modèles linéaires.....	107
Figure 38: Distribution géographique du taux de présence par demi-degré d'angle (bancs libres).....	109
Figure 39: Distribution géographique du taux de présence par mois (bancs libres et associés).....	111
Figure 40: Variations spatiales du taux de présence et des paramètres environnementaux.....	112
Figure 41: Distribution des prédictions du modèle fondé sur la chlorophylle-a.....	115
Figure 42: Exemple de carte de potentiel générée à partir des concentrations en chlorophylle-a.....	115
Figure 43: Distribution des prédictions du modèle basé sur la température de surface GAC.....	117
Figure 44: Version modifiée du modèle de la figure 43 après remplacement de ∇SST_{-15} par ∇SST	118
Figure 45: Exemple de carte de potentiel générée à partir des températures de surface.....	119
Figure 46: Version modifiée du modèle de la figure 43 avec ajout des termes ϕ et $\phi \times SST_{-15}$	120
Figure 47: Distribution des prédictions du modèle fondé sur la température de surface.....	121
Figure 48: Exemple de carte de potentiel générée à partir des températures de surface.....	122
Figure 49: Distribution des prédictions du modèle fondé sur l'anomalie de hauteur de l'eau.....	123
Figure 50: Exemple de carte de potentiel générée à partir des anomalies de hauteur de l'eau.....	124
Figure 51: Exemple de carte de potentiel générée à partir de la température GAC et la SLA.....	128
Figure 52: Distribution des prédictions du modèle fondé sur l'ensemble des paramètres.....	131
Figure 53: Exemple de carte de potentiel générée à partir de l'ensemble des paramètres.....	132
Figure 54: Exemple de réseau de neurones artificiels de type feed-forward à deux couches.....	134
Figure 55: Illustration de deux sigmoïdes couramment utilisées comme fonction d'activation ψ	134
Figure 56: Exemple de réseau sigma-pi.....	135
Figure 57: Exemple de réseau de neurones expérimenté sur les données de pêches.....	137
Figure 58: Distribution des prédictions du modèle basé sur un réseau de neurone.....	138

Index des équations

Équation 1: Calcul de gradient à partir d'opérateurs de convolution.....	67
Équation 2: Évaluation d'un descripteur du paysage océanique.....	91
Équation 3: Modélisation (pas nécessairement linéaire) de la probabilité de présence.....	133
Équation 4: Sortie du réseau de neurones sigma-pi.....	135
Équation 5: Correction à apporter aux poids d'une connexion entre deux neurones.....	136

Index des méthodes

Méthode 1: Transformation de coordonnées sur une machine locale (API de Geotools 2).....	61
Méthode 2: Transformation de coordonnées sur une machine locale ou distante (API d'OpenGIS).....	62
Méthode 3: Évaluation de valeurs à une coordonnée géographique (API de Geotools 2).....	64

Préface

Le projet SEAGERH cherche à améliorer les connaissances de la ressource halieutique et la dynamique de l'environnement marin. La démarche est fondée sur une collaboration entre différents acteurs (pêcheurs, gestionnaires et scientifiques) du secteur de la pêche des grands pélagiques. Tout d'abord, ce procédé caractérise l'environnement océanographique par différents produits issus de l'imagerie satellitaire : la température de l'eau de surface, la concentration en chlorophylle-*a*, les anomalies de la hauteur de l'eau et le pompage d'Ekman. Ces données satellitaires, dont la couverture est mondiale et homogène, servent de briques d'un système d'observation continue des structures océanographiques où évolue la ressource halieutique.

Les pêcheurs ayant intégré le projet SEAGERH doivent accepter la règle suivante : réception des produits environnementaux à bord en contrepartie de la fourniture en temps quasi réel des données de captures. Des données issues de cet échange ont servi dans ce travail à estimer l'impact de l'environnement sur le comportement des thons (chapitre). Le projet tente de promouvoir une manière de gérer les pêcheries fondée sur une meilleure connaissance de l'environnement marin. Il constitue une étape supplémentaire pour l'aide à la mise en oeuvre de nouvelles réglementations des pêches. Ce projet est soutenu par le Ministère de la Recherche (France – Réseau RTE contrat #31).

Le travail présenté dans cette thèse s'intègre dans le projet SEAGERH en proposant d'évaluer le potentiel des données altimétriques (anomalie des hauteurs dynamiques et courants géostrophiques) pour l'halieutique. Plus spécifiquement, il s'agit :

- de tester la cohérence des paramètres issus de l'altimétrie avec les autres paramètres (température, concentration en chlorophylle-*a*) pour décrire le paysage océanique dans le cas de l'océan Indien ;
- d'évaluer la pertinence relative de ces produits issus de l'altimétrie pour caractériser les zones de concentration des bancs de pélagiques ;
- de participer au développement d'un outil d'information halieutique qui replace, de façon synoptique, la ressource dans son contexte spatial, environnemental et temporel. Cet outil se veut pertinent à la fois pour la recherche et pour la gestion / décision.

Dans le cadre de cette thèse, ces trois points justifient l'investissement nécessaire dans trois domaines : la mesure satellitaire des océans, l'océanographie opérationnelle des pêches et le développement de briques de SIG (Système d'information géographique) que l'on peut déployer sous forme de services.

Le chapitre 3 élaboré les outils informatiques mis en œuvre. On peut le voir comme une mini-thèse dans la thèse, avec sa propre problématique beaucoup plus large que ce qui aurait été strictement nécessaire pour répondre à la question halieutique. Il y sera abondamment question de standards internationaux, notamment des spécifications du consortium OpenGIS®. Ce chapitre ne cherchera pas à faire une présentation générale de ces standards ; ces présentations existent déjà par ailleurs. Les chapitres 3.3 et 3.4 en particulier feront plutôt une présentation orientée vers l'usage que nous allons faire de ces standards et vers la façon dont nous les avons implémentés dans le projet libre Geotools 2. Ce chapitre 3 n'a toutefois pas la structure qu'aurait eue une thèse en informatique dans la mesure où il mélange des explications formelles avec des détails techniques. Son organisation se rapproche davantage de celle d'un tutoriel introduisant le lecteur à l'utilisation des outils développés pour cette thèse au sein du projet Geotools 2.

Chapitre 1

Océanographie des pêches thonières

1.1 Les enjeux de la pêche thonière

La pêcherie thonière se situe au troisième rang du commerce international de la mer, derrière la pêcherie crevettière et celle des poissons démersaux. Plus de 50% des captures mondiales sont effectuées par le Japon et les États-Unis. Ces deux pays consomment à eux seuls environ 75% des captures mondiales des principales espèces de thonidés. Le Japon capture annuellement 200 000 tonnes de thon de qualité sashimi et en importe 270 000 tonnes, dont 50 000 en provenance d'une pêche non réglementée effectuée sous des pavillons de complaisance.

La majorité des spécialistes de la dynamique des populations des grands pélagiques pensent que, comme pour 70% des pêcheries mondiales (Garcia et Newton, 1997), les stocks de thonidés sont proches de la surexploitation (Deriso et Bayliff, 1991). Cette situation est généralement expliquée par la demande croissante de poissons et par la forte augmentation de la capacité de pêche des flottilles. Cependant le monde des pêcheries thonières est compliqué et il est actuellement difficile de déterminer quel est l'état réel des stocks. Pour assurer le renouvellement des stocks, il ne suffit pas d'épargner quelques géniteurs. La pression démographique des espèces concurrentes joue aussi : une espèce concurrente peut occuper la niche écologique de l'espèce exploitée. De plus, on a constaté que chez les espèces à tendance grégaire, l'émission des produits génitaux ne se déclenche que lorsqu'un nombre élevé d'individus sont groupés. Enfin, les très jeunes individus ont de meilleures chances de survie face aux prédateurs s'ils sont assez nombreux pour déborder ceux-ci : les prédateurs disposent de ressources surabondantes pendant le laps de temps nécessaire à la croissance des proies au-delà d'une taille critique (Collignon, 1991). Les prélèvements possibles sont donc limités par la quantité qui peut être extraite sans hypothéquer le renouvellement des stocks. Mais l'application de ce principe se complique dans le cas des pêches multi-spécifiques : chaque unité capture habituellement plusieurs espèces de thons, dont quelques unes peuvent être localement surexploitées et d'autres non. La gestion des pêches thonières est rendue encore plus difficile du fait qu'en parcourant de grandes distances, les thons traversent régulièrement plusieurs juridictions. La gestion requiert alors une coopération internationale.

On prévoit que la population mondiale continuera à croître jusqu'à 10 milliards de personnes au milieu du 21^e siècle. En réponse à la demande croissante, les captures de thons augmentent régulièrement depuis 50 ans. Les flottilles ont atteint une telle taille que plusieurs navires opèrent bien en-dessous de leur niveau de rentabilité optimal. Dans ces conditions, il est difficile pour les gouvernements d'assurer une exploitation rationnelle et durable de la ressource. Dans son étude sur la gestion des capacités de pêche de la flotte thonière mondiale, Joseph (2003) conclut ainsi :

« Bringing new vessels into the tuna fisheries of the world has been unrestricted and been considered everyone's right, and this right is enshrined in international law. It is time to change this policy, and time to limit the number of vessels authorized to fish for tuna. The high seas can no longer be considered a frontier in which its natural resources are there for the taking. The current legal and political basis ensuring the right of every person to fish on the high seas must be re-examined and brought in line with current reality. This will require bold new approaches as to the management of the high-seas resources. The time we have to work on this is limited, and action must be swift if we are to ensure that tuna populations are maintained at levels of abundance that can support maximum yields on a sustained basis, and to guarantee to future generations the option to enjoy the benefits of these resources. »

L'écologie des thonidés est déjà complexe en soi ; gérer de surcroît les capacités de pêche thonière est une gageure. En effet, la pêche thonière est constituée d'une grande variété (à la fois de taille et de type) de bateaux et d'engins. Cette variété rend difficile les comparaisons de données de pêche provenant de différents pêcheurs. De plus, cette pêcherie implique environ 70 pays opérant dans leurs zones économiques, à l'étranger ou en zone internationale. Même si les stocks et le marché concernés sont globalement les mêmes, les objectifs de ces pays diffèrent. Par exemple certains veulent favoriser l'emploi, d'autres le retour économique. Les pavillons de complaisance ou l'interaction des investisseurs étrangers qui arment des unités de pêche sous le pavillon du pays côtier ajoutent à la confusion. Cette complexité rend délicates l'application et l'interprétation des modèles économiques, même sophistiqués (Gréboval et Munro, 1999).

Ces éléments suggèrent que les méthodes habituelles de gestion des pêcheries (les quotas ainsi que les restrictions saisonnières, par zone ou par engin, etc.) manquent d'efficacité. Des approches plus novatrices peuvent être nécessaires pour arriver à un meilleur contrôle de la capacité de pêche des flottilles. Le projet SEAGERH (voir préface) en est une tentative. Avant d'aborder les méthodes et outils employés dans le cadre

de ce projet, les pages qui suivent présenteront quelques éléments d'halieutique et d'océanographie utiles à la compréhension de notre travail et de sa place dans la recherche actuelle. Il ne s'agira toutefois pas d'un exposé exhaustif des connaissances sur le sujet.

1.1.1 Les espèces concernées

La quasi totalité des espèces de thonidés intéressent, à des degrés divers, toutes les activités de pêches : sportives, artisanales ou industrielles. Quasiment toutes sont aussi recensées à l'annexe I de la Convention de 1982 en tant qu'espèces de grands migrateurs. Les systématiciens classent ces « thonidés » dans le sous-ordre des *scombridés* qui englobe les thons, les marlins et voiliers (makaires), les espadons et les espèces de petits thonidés des zones tropicales et tempérées de toutes les régions du monde. Le thon (*Thunnini*) comprend 13 espèces et 4 genres: *Thunnus*, *Euthynnus*, *Katsuwonus* et *Auxis*. De loin, c'est le genre *Thunnus* qui est le plus recherché par la pêche industrielle. Mis à part le thon rouge du nord (*Thunnus thynnus*) qui est absent de l'océan Indien, on trouve toutes les autres espèces dans tous les océans tropicaux et tempérés (Atlantique, Pacifique, Indien). Ces espèces sont énumérées dans le tableau 1. La figure 1 les illustre, ainsi que leur distribution. On distingue les thons tropicaux des thons tempérés pour des raisons de préferendum thermique différent (section 1.4.1) et d'exploitation par différentes pêcheries.

Nom français	Nom anglais	Nom latin	Préferendum	Légende
Thon rouge du sud	Bluefin tuna	<i>Thunnus maccoyii</i> (Castelnau, 1872)	10 – 22 °C	■
Thon obèse	Bigeye tuna	<i>Thunnus obesus</i> (Lowe, 1839)	11 – 25 °C	■ ■
Albacore	Yellowfin tuna	<i>Thunnus albacares</i> (Bonnaterre, 1788)	23 – 32 °C	■ ■ ■
Germon	Albacore tuna	<i>Thunnus alalunga</i> (Bonnaterre, 1788)	15 – 22 °C	■ ■ ■ ■
Listao ¹	Skipjack	<i>Katsuwonus pelamis</i> (Linnaeus, 1758)	20 – 32 °C	■ ■ ■ ■ ■

Tableau 1: Espèces de thons rencontrées dans les océans tropicaux et tempérés

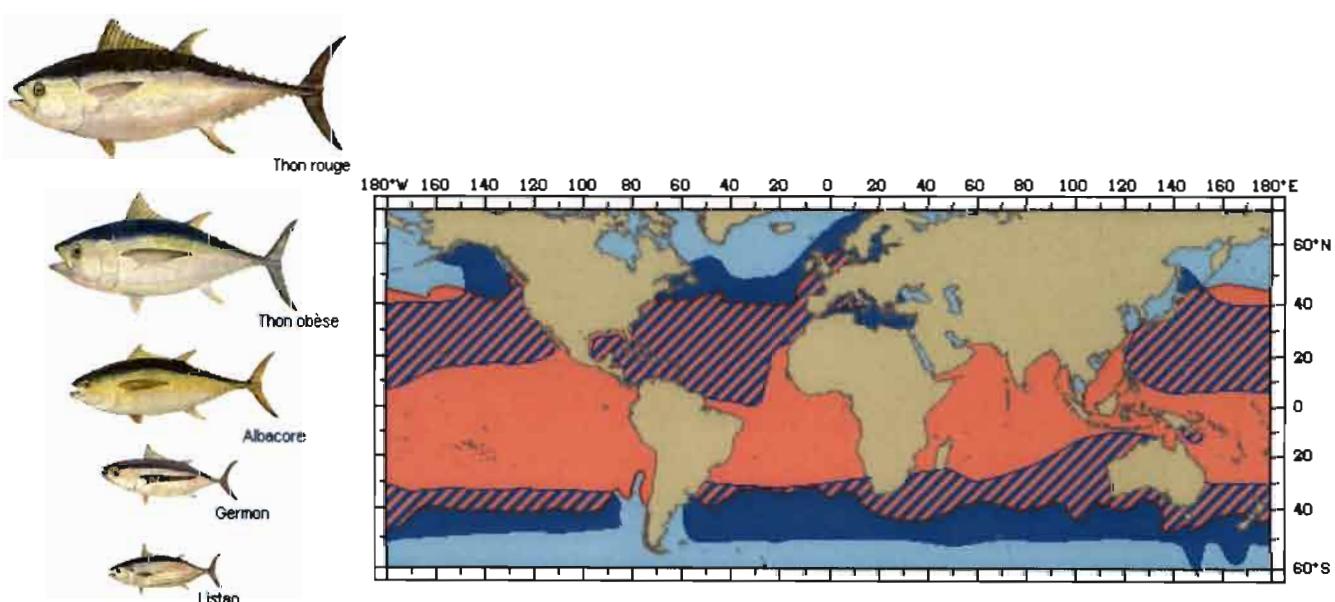


Figure 1: Distribution des principales espèces de thons (source: FAO²)

Les distributions des thons tropicaux et tempérés sont représentées en rose et en bleu respectivement. La correspondance entre ces couleurs et les espèces est donnée par la dernière colonne du tableau 1.

Les thons tropicaux incluent le listao et l'albacore, tandis que les thons tempérés incluent le thon rouge et le germon. Le thon obèse s'observe dans les deux catégories. Notre travail se concentrant sur l'océan Indien, les principales espèces concernées sont :

¹ Aussi appelée « bonite à ventre rayé ».

- **Thon obète**

De taille moyenne, le thon obète ressemble beaucoup à l'albacore et est parfois confondu avec lui par les pêcheurs et les enquêteurs. Les juvéniles forment souvent des bancs près de la surface sous des objets flottants, où ils se mélangent parfois avec des albacores et des listaos. L'association avec les objets flottants devient moins fréquente au fur et à mesure que la taille des thons augmente. On les retrouve alors plus souvent sous la thermocline, en partie grâce à une couche de graisse sous cutanée qui le rend plus adapté que les autres thons tropicaux au froid. Cette couche de graisse en fait aussi un poisson recherché pour le marché japonais lucratif du *sashimi*. La plupart des thons obèses pêchés sont de grands individus (plus de 20 kg) capturés par des palangres en profondeur. Des individus plus petits sont aussi capturés près de la surface par les senneurs (CTOI, 2003). Les prises palangrières baissent de plusieurs dizaines de milliers de tonnes par an depuis 1985, tandis que celles des senneurs augmentent fortement en partie grâce à la pêche sous objets flottants. Cette dernière nécessite une surveillance particulière, car elle semble représenter un danger pour le recrutement à la fois du thon obète et de l'albacore.

- **Albacore**

L'albacore est la cible favorite des senneurs industriels. Il représente 30% des captures mondiales en tonnage, soit la deuxième espèce la plus pêchée après le listao. Dans l'océan Indien, la longueur à la fourche des individus capturés va de 30 cm à 170 cm, pour une taille à maturité estimée à 110 cm. Dans les bancs d'individus de grande taille capturés dans l'océan Indien, les mâles sont majoritaires et de taille moyenne apparemment plus élevée (150 cm) que dans les autres océans. Les individus plus petits (les juvéniles) forment des bancs mélangés aux listaos et aux juvéniles de thon obète. Ces bancs de juvéniles sont principalement rencontrés dans les eaux tropicales de surface, tandis que les poissons plus grands exploitent des eaux plus profondes. Les albacores d'âge intermédiaire ne sont que rarement capturés par les pêcheries industrielles (CTOI, 2003). La capture de petits individus est grandement favorisée par l'utilisation de dispositifs concentrateurs de poisson (section 1.4.4 à la page 26) ainsi que des bouées et des balises. Les halieutes s'accordent à dire que ce type d'activité peut influencer fortement le recrutement et, par là, le taux optimal des prises (Anganuzzi et al., 1996).

- **Listao**

Les listaos forment des bancs mélangés en association avec d'autres thons de même taille, tels que les juvéniles d'albacore. Cette espèce a une durée de vie courte, une mortalité naturelle élevée (« *turn over* » important), une forte fécondité et une reproduction précoce. Étant donné que la taille de première maturité est de 41 à 43 cm pour les mâles et les femelles, l'essentiel des captures consiste en individus qui se sont déjà reproduits (CTOI, 2003). Son habitat préféré se trouve plutôt dans la couche mélangée près de la surface. Il constitue ainsi une cible privilégiée pour les pêcheries de surface, notamment la senne coulissante. Depuis plusieurs années, cette espèce est présente en quantité dans toutes les pêcheries commerciales effectuées entre 40°S et 45°N. Elle constitue à elle seule environ 50% des prises mondiales de thon. Les prises sont passées de 400 000 tonnes en 1970 à 2 millions de tonnes en 1999 (figure 2), dont 350 000 tonnes dans l'océan Indien. Dans cet océan, les prises de listao ont augmenté comme dans le Pacifique et l'Atlantique, mais ne constituent pas pour autant la plus grande proportion. Durant la décennie 1990 à 2000, les prises d'albacore et de listao y étaient équivalentes (figure 3).

Les figures 2 et 3 représentent les captures mondiales de thon par espèce commerciale, pour l'ensemble de la planète et pour l'océan Indien respectivement. Les codes de la FAO² utilisés dans ces figures sont :

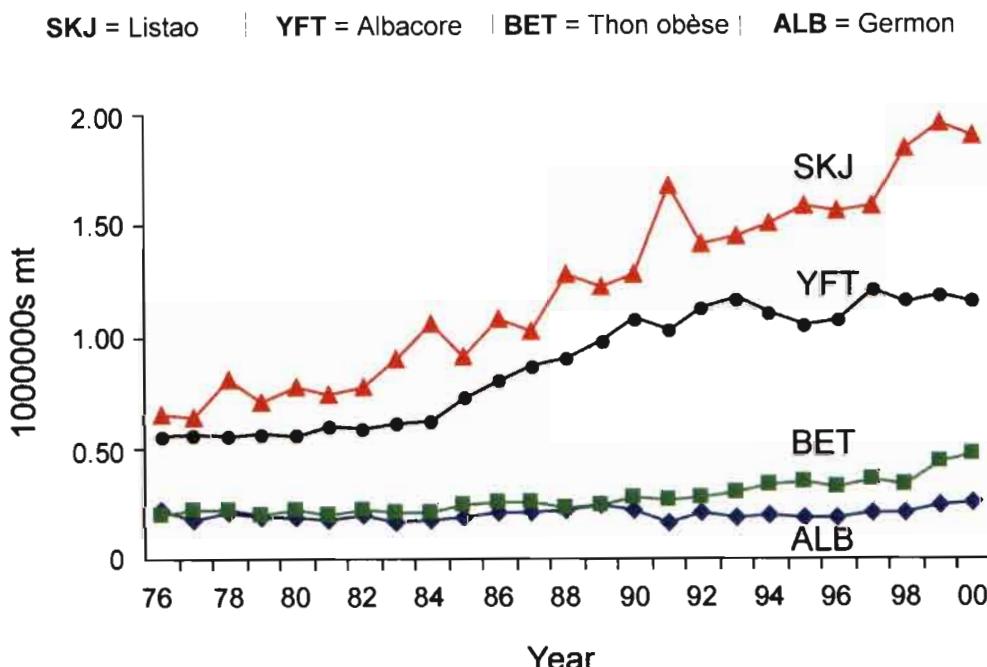


Figure 2: Prises annuelles mondiales des espèces commerciales de thons
Source: FAO (Joseph, 2003)

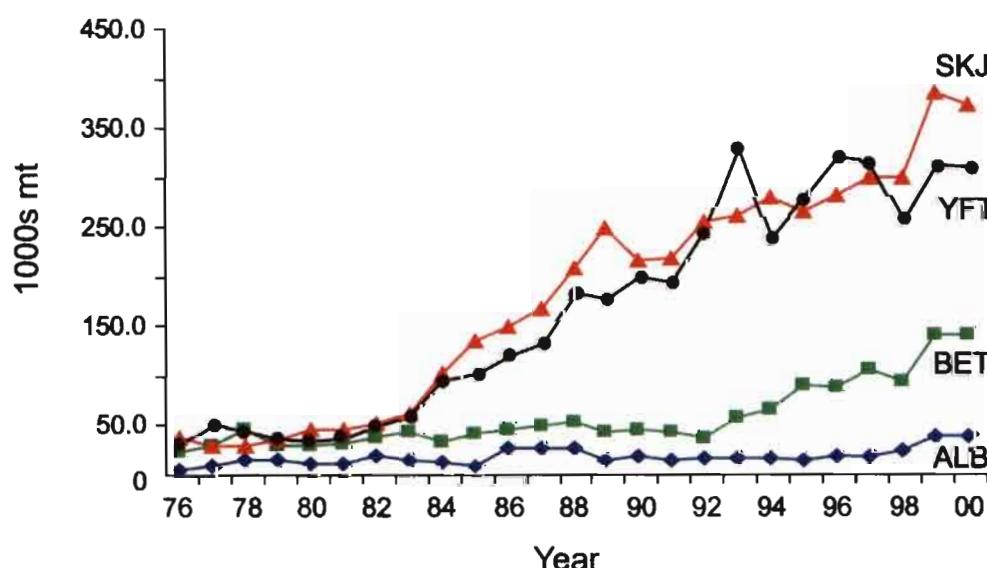


Figure 3: Prises annuelles dans l'océan Indien des espèces commerciales de thons
Source: FAO (Joseph, 2003)

1.1.2 Le gisement thonier de l'océan Indien

Au début des années 1980, la flottille des thoniers senneurs européens (essentiellement espagnols et français), confrontée à divers problèmes d'accès à la ressource en Atlantique tropical (Petit, 1984), s'est redéployée dans l'océan Indien occidental. Avant ce redéploiement, la production de l'ensemble de l'océan Indien avoisinait 8% de la production mondiale (figure 4). Elle provenait essentiellement des pêcheries artisanales des Maldives et du Sri Lanka ainsi que, pour une plus faible part, des palangriers asiatiques.

² FAO: Food and Agricultural Organization.

Depuis 1993, ce taux est passé à 20% de la production annuelle mondiale. On attribue 75 % de cette production à la flottille européenne de senneurs ainsi qu'à la pêcherie des Maldives et du Sri Lanka. Parmi les espèces capturées, l'albacore et le listao représentent 80% du tonnage.

La production de l'océan Indien est comparable à celle de l'océan Atlantique (figure 4). Les captures moyennes sont de l'ordre de 500 000 tonnes, avec la même prédominance (80%) de l'albacore et du listao. En revanche, le Pacifique Ouest produit à lui seul plus d'un million de tonnes, principalement avec les flottilles des senneurs américains et japonais. Ces dernières se sont redéployées depuis le Pacifique Est et Nord, elles aussi au début des années 1980 (Petit, 1984). Près des deux tiers des prises mondiales sont effectuées dans le Pacifique.

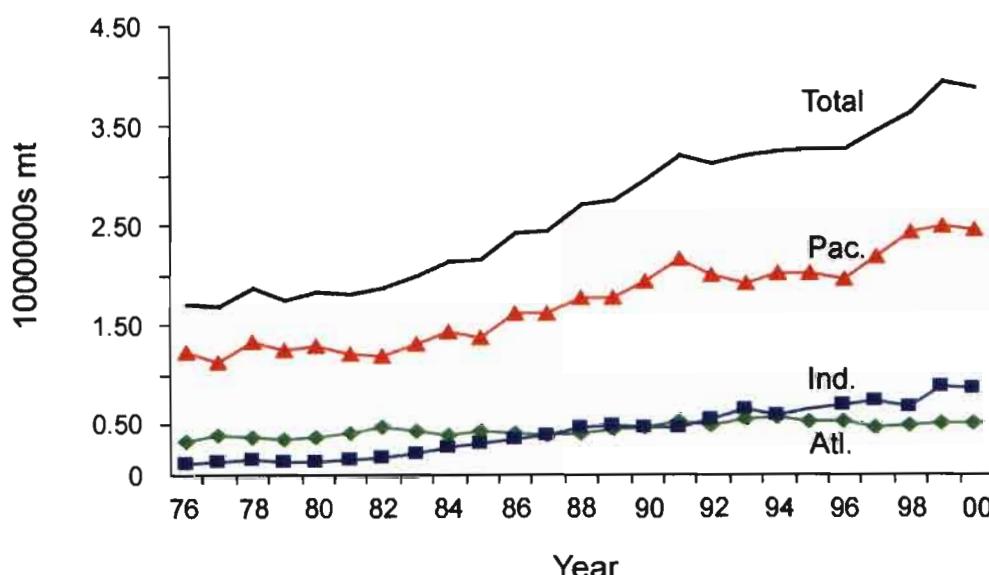


Figure 4: Prises annuelles des espèces commerciales de thons par océan
Source: FAO (Joseph, 2003)

Pour les espèces de poissons répertoriées dans les statistiques de pêche de la FAO (dont les thonidés ne représentent qu'une fraction), le niveau trophique moyen des espèces capturées a décliné de 1950 à 1994 (Pauly et al., 1998). Ces auteurs ont observé une transition des captures à partir d'espèces de fond piscivores à vie longue, vers des espèces pélagiques invertébrées et planctonivores (ou planctonophage) à vie plus courte. Cet effet est plus prononcé dans l'hémisphère nord et mène à un accroissement temporaire des captures, suivi d'une stagnation et d'un déclin. Dans le cas des pêches d'albacore dans l'océan Indien, les poids moyens annuels des individus capturés ont d'abord décliné, sont restés stables entre 1970 et 1990, puis ont décliné de nouveau après 1993 (CTOI, 2003). Dans son rapport de 2003, le comité scientifique de la CTOI³ considère que « *la tendance actuelle à accroître la pression de pêche sur les albacores juvéniles par les senneurs pêchant sur objets flottants aura probablement, si elle se poursuit, des effets délétères sur le stock* » (CTOI, 2003). Cependant, la commission note que les captures d'albacores par la pêche à la senne ont atteint des niveaux records en 2003, alors même que l'effort de pêche était d'un niveau comparable à celui de 2000. Les causes de cette augmentation ne sont pas encore connues, mais une amélioration de l'efficacité de la pêche semble plus probable qu'une augmentation des stocks (toujours d'après les spécialistes de la commission).

De part ses caractéristiques biologiques mentionnées à la page 12, le listao est considéré comme une ressource plus résistante et plus difficile à sur-pêcher que l'albacore et le thon obèse (CTOI, 2003). Mais bien qu'il n'y ait pas de préoccupation immédiate pour cette espèce dans l'océan Indien (la situation est plus préoccupante en Atlantique), la commission souligne que les captures ne peuvent continuer de croître au rythme actuel de façon indéfinie.

³ CTOI: Commission des thons de l'océan Indien.

1.1.3 Le système pêche et la télédétection

Le travail présenté dans cette thèse découle d'une certaine conception du « système pêche » (Sharp, 1981). Dans ce modèle, la société humaine et les espèces marines exploitables appartiennent à des mondes différents et mènent des existences parallèles jusqu'au moment où l'homme fait un effort pour rencontrer ces espèces. Cet effort est l'activité de pêche, qui exige (en particulier pour les pélagiques) la découverte des bancs. Durant cette étape d'effort de recherche, l'homme n'a pas encore fait intrusion dans le monde des espèces marines, et nous supposerons que ces dernières ignorent l'homme⁴. Le banc trouvé, l'interaction entre les deux mondes commence lorsque l'homme réagit à cette découverte. S'il y a capture, le poisson est alors absorbé dans la société humaine. Sinon l'interaction cesse, bien que le point et l'instant de contact soient mémorisés de part et d'autre.

Le « système pêche » ainsi pensé, il apparaît que la répartition des espèces commerciales doit être évaluée à partir des bancs détectés plutôt que pêchés. La télédétection est une étape décisive, qui n'a pas attendu l'ère des satellites pour jouer son rôle (surtout dans le domaine pélagique). Le pêcheur lui-même l'utilise par son savoir-faire : il recherche les bancs de poissons soit directement à vue, ou soit en associant la concentration des bancs avec un ensemble de paramètres, qu'ils soient biologiques (par exemple une présence d'oiseaux) ou physiques (marées, courants, etc.). Depuis l'antiquité, pour signaler le passage de bancs de thons en Méditerranée, des guetteurs se sont postés sur des hauteurs dominant la mer dans des tours ou « thunoscopes » (Farrugio, 1981). Les capteurs modernes élargissent le domaine spectral des observations océaniques, ainsi que leur couverture spatio-temporelle. Ces nouvelles informations collaborent à la définition d'indicateurs, comme nous le verrons plus loin.

L'observation aérienne d'animaux aquatiques est quasiment née avec l'aviation (Petit, 1991). Par exemple des avions ont servi à localiser des bancs pour les pêcheurs de sardines, maquereaux et mullets au printemps 1918 à Audierne (France), ainsi qu'au repérage des phoques entre 1923 et 1929 près de Terre-Neuve (Canada). En 1925, un ballon tiré par un remorqueur a été expérimenté pour l'observation des thons près de cap Zébib (Tunisie). Dans un article de mars 1925 intitulé « *les hydravions au service des pêcheries américaines* », la « *Pêche maritime* » rapporte :

« Le rendement annuel des pêches sur la côte Pacifique des États-unis, qui oscille entre 18 et 20 millions de dollars, a été augmenté en 1923 et 1924, de 25% par suite de l'utilisation des hydravions. [...] Pour le thon et les autres espèces migratrices qui apparaissent dans les eaux de San-Diego et du Mexique au début de l'été et les fréquentent jusqu'en septembre, les pêcheurs s'efforcent de les dépister le jour en suivant leur proie et la nuit en observant la phosphorescence de la mer. Mais en 1922 et 1923, les pêcheurs furent souvent en défaut, le passage fut très irrégulier et le poisson se tenait beaucoup plus au large que d'habitude. Le rendement eût été très faible sans le secours des avions. Mais, grâce à eux, et malgré les conditions particulièrement défavorables, les quantités pêchées furent équivalentes à celles des meilleures années antérieures. [...] Sans qu'il soit nécessaire d'insister, on conçoit quelle sécurité usiniers et pêcheurs trouvent dans une telle organisation. Le temps, l'argent et la peine sont ainsi largement économisés pour le plus grand profit de tous⁵. »

— « *La pêche maritime* », mars 1925

Ces 25 dernières années ont vu l'apparition de nouveaux capteurs embarqués à bord de vecteurs aériens ou spatiaux. Après les vecteurs aériens, ce sont donc les vecteurs spatiaux (satellites à défilement ou géostationnaires) qui ont été testés avec des capteurs passifs ou des capteurs actifs. La littérature est riche de ces expérimentations avec différents vecteurs et surtout différents types de capteurs. Le chapitre 1.2 donne un aperçu de quelques expérimentations, en cherchant surtout à faire ressortir l'intérêt des idées sous-jacentes. Dans certains cas, ces idées pourraient être de nouveau développées à partir d'un appareillage plus moderne.

Si les expériences ponctuelles sont nombreuses, en revanche leur mise en applications opérationnelles l'est bien moins. A cela, deux explications peuvent être avancées : le coût du passage de l'expérimental à l'opérationnel, et le fait que ces applications opérationnelles sont rarement l'objet d'une publication scientifique. Cette dernière raison est motivée par un désir de conserver une avance technologique dans un contexte concurrentiel, mais ce raisonnement est devenu moins évident depuis l'explosion d'internet (voir par exemple la section 3.1.5 à la page 46). Cependant, plusieurs synthèses ont été publiées durant le dernier

⁴ Rappelons qu'il s'agit d'un modèle simplifié. Dans la pratique, il arrive que des bancs de thons s'associent au bateau (section 1.4.4). Il est aussi plausible que le bruit du bateau affecte le comportement de certaines espèces marines.

⁵ À l'époque, l'état des stocks ne suscitait évidemment pas l'inquiétude d'aujourd'hui.

quart de siècle. C'est ainsi que sous l'égide de la FAO, Tomczak (1977) puis Butler *et al.* (1988) ont inventorié vingt-deux applications des techniques aérospatiales aux pêcheries, dont certaines ne sont restées que ponctuelles et expérimentales. Toujours sous l'égide de la FAO, deux autres synthèses sur la télédétection satellitaire utilisée dans les pêches pélagiques côtières et océaniques ont été écrites par Yamanaka *et al.* (1988) et par Le Gall (1989).

Dans le cas particulier de l'aéroporté, on peut citer deux rapports : le premier (Marsac *et al.*, 1987) ne traite que de la pêche thonière et des travaux de l'IRD. Le second (Hunter et Churnside, 1995) présente les travaux d'une vingtaine de chercheurs, de centres de recherches et d'industries, qui se sont réunis sous l'égide de la NOAA⁶. Ce groupe a identifié trois thèmes de recherches : (1) l'amélioration des indices existant en matière d'abondance des petits pélagiques et du thon rouge, (2) le développement de mesure de la biomasse pour les petits pélagiques et (3) le développement d'une halieutique opérationnelle aéroportée pour l'étude et le suivi de la distribution et des mouvements des saumons océaniques et des thonidés.

⁶ NOAA: National Oceanic and Atmospheric Administration

Tableau 2: Quelques expériences et applications de la télédétection en halieutique**Synthèses**

Année	Auteurs	Espèce cible	Paramètres	Capteurs	Vecteurs
1987	Marsac <i>et al.</i>	Thonidés	Température de surface	Radiomètre infra-rouge	Avion
1988	Butler <i>et al.</i>		Divers	Divers	Avion / satellite
1988	Yamanaka <i>et al.</i>	Pélagiques côtiers et hauturiers	Divers	Divers	Satellite
1989	Le Gall	Thonidés	Divers	Divers	Avion / satellite
1995	Hunter et Churnside	Pélagiques	Divers	Divers	Avion
1997	Tomczak		Divers	Divers	Avion / satellite

Recherche directe et indirecte d'une ressource halieutique / tactique de pêche

Année	Auteurs	Espèce cible	Paramètres	Capteurs	Vecteurs
1970	Roithmayr	Maquereau espagnol	Bioluminescence	Télévision	Avion
1976	Cram et Hampton	Pilchard	Bioluminescence	Télévision	Avion
1977	Stretta	Thonidés	Température de surface	Radiomètre infra-rouge	Avion
1979	Cram	Pilchard	Chlorophylle	CZCS (NIMBUS-7)	Satellite
1981	Squire et Krumboltz		Détection directe	LIDAR	Avion
1984	Laurs <i>et al.</i>	Thonidés	Turbidité + chlorophylle	CZCS (NIMBUS-7)	Satellite
1984	Leming et Stuntz	Crevette	Hypoxie	CZCS (NIMBUS-7)	Satellite
1986	Montgomery <i>et al.</i>	Thonidés	Température de surface + couleur	CZCS (NIMBUS-7)	Satellite
1988	Lohoefner <i>et al.</i>	Anchois, menhaden		Photographies	Avion
1988	Shinomi et Tameishi	Saury	Température de surface	AVHRR (NOAA)	Satellite
1989	Herron <i>et al.</i>	Butterfish	Température de surface	AVHRR (NOAA)	Satellite
1989	Nakashima <i>et al.</i>	Capelin	Couleur	CASI	Avion
1990	Le Gall	Huitres (aquaculture)	Détection directe	Photo, vidéo	Avion
1990	Nakashima	Capelin		Photographie	Avion
1990	Petit <i>et al.</i>	Thonidés	Détection directe	SAR	Avion
1991	Barbieri <i>et al.</i>	Germon, espadon	Température de surface	AVHRR (NOAA)	Satellite
1991	Power et May jr	Albacore (pêche palangrière)	Température de surface	AVHRR (NOAA)	Satellite
1992	Borstad <i>et al.</i>	Hareng	Couleur	CASI	Avion
1992	Ramos	Thonidés	Température de surface + couleur	AVHRR/CZCS	Satellite
1993	Nakashima et Borstad	Capelin	Couleur	CASI	Avion
1993	Santiago <i>et al.</i>	Germon	Température de surface	AVHRR (NOAA)	Satellite
1994	Oliver <i>et al.</i>	Thonidés	Détection directe	LIDAR	Avion
1995	Hunter et Churnside	Pélagiques	Détection directe	LIDAR	Avion

Évaluation de l'effort de pêche et estimation de l'abondance

Année	Auteurs	Espèce cible	Paramètres	Capteurs	Vecteurs
1989	Pringle et Duggan	Homard	Détection directe	Photographies	Avion
1990	Bour	Troca, Faviidés	Cartes thématiques	SPOT multispectral	Satellite
1990	Petit <i>et al.</i>	Thonidés	Détection directe	SAR	Avion
1991	Petit	Thonidés	Température de surface + détection directe	Radiomètre infra-rouge	Avion
1992	Lo <i>et al.</i>	Anchois, maquereau, sardine	Détection directe	Photographies	Avion
1997	Lutcavage <i>et al.</i>	Thon rouge	Détection directe	Photographies	Avion

Recherche de zones potentielles / prévisions de pêche

Année	Auteurs	Espèce cible	Paramètres	Capteurs	Vecteurs
1989	Petit <i>et al.</i>	Thonidés	Guyot	SPOT multispectral	Satellite
1991	Stretta	Thonidés	Température de surface	METEOSAT	Satellite

1.2 Apport de la télédétection dans le processus décisionnel

Les applications de la télédétection en halieutique peuvent se classer suivant deux catégories, selon qu'elles recherchent la ressource directement ou indirectement. Indépendamment du mode de recherche, on distingue aussi deux finalités : une recherche pour la capture ou une évaluation d'abondance à des fins de gestion.

1.2.1 Recherche directe d'une ressource halieutique

Nous avons mentionné dans la section 1.1.3 que le repérage visuel des thons n'est pas récent. Aujourd'hui, de nombreux capteurs aéroportés viennent compléter la vision d'observateurs embarqués. Un des capteurs passifs les plus classiques est la pellicule photographique, utilisée dans le visible ou le proche infra-rouge. Des photographies diurnes ont été utilisées pour estimer l'abondance de bancs d'anchois et de menhaden (Lohoechner et al., 1988) ou de capelan (Nakashima, 1990). Par des nuits sans lune, des amplificateurs de lumière ont permis de visualiser des bancs de poissons à la surface de la mer qui excitent les organismes bioluminescents du phytoplancton (Roithmayr, 1970; Roithmayr et Wittmann, 1973; Cram et Hampton, 1976). D'autres capteurs testés incluent le spectromètre CASI⁷ (Nakashima et Borstad, 1993; Borstad et al., 1982) et le LIDAR (Squire et Krumboltz, 1981; Oliver et al., 1994). Ce dernier est un capteur actif émettant une lumière cohérente à partir d'un laser embarqué. Ce signal est modifié en fonction de la nature de la cible et, après retour vers l'avion, est enregistré et analysé.

La présence du thon se manifeste à la surface de l'eau de différentes façons, que détaillent Petit et Stretta (1989). Le banc peut être sous la surface et se déplacer dans une même direction, créant ainsi des frémissements et des irrégularités à la surface de l'eau. Dans ce cas, aucune partie du corps des poissons ne crève la surface ; il n'y a donc pas d'éclaboussures. Le banc peut aussi être en grande activité de prédation. L'eau est alors agitée en surface et très blanche ; elle semble bouillir. Ces modifications de la surface de l'eau sont susceptibles d'être repérées par un radar de type SAR (*Synthetic Aperture Radar*). Une expérience menée en Méditerranée a permis de discriminer les bancs de thons, les cétacés (baleines et dauphins), les bateaux et leurs sennes à l'aide d'un radar E-SAR à partir d'un avion évoluant à 2000 mètres d'altitude.

Aujourd'hui, les satellites d'observation de la terre sont capables d'une résolution spatiale suffisamment fine pour détecter des bancs par des méthodes similaires à celles qui ont été testées à bord des avions. Toutefois, la fréquence d'apparition des bancs en surface rend improbable leur détection directe par satellite, étant donné que la fréquence d'acquisition de ces derniers est faible : les deux événements se produisent rarement en même temps. Même en supposant cet enregistrement réalisé, il peut être difficile de discriminer la signature du banc dans le bruit spectral ambiant.

1.2.2 Recherche indirecte d'une ressource halieutique

De nombreuses expériences et applications opérationnelles de télédétection utilisent la température de surface de la mer (Tomczak, 1977; Butler et al., 1988; Laurs, 1989a; Hunter et Churnside, 1995). Par exemple la radiométrie aérienne avec un sondeur de type BARNES a permis de guider des bateaux de pêche dans des zones de fronts thermiques (Marsac et al., 1987). L'avion recense également toutes sortes d'apparences à la surface de la mer (bancs de poissons, objets flottants, cétacés, oiseaux, unités de pêches, etc.). À partir de bonnes connaissances océanographiques de la dynamique d'une zone frontale, il a été possible dans certains cas de diffuser des prévisions à court terme (1 à 3 jours) sur l'évolution du système (front thermique-thonidés). Ces tentatives ont été faites dès 1976 entre le cap Lopez et l'île de Sao Tomé dans l'Atlantique tropical oriental (Stretta, 1977). Dans ces exemples où la télédétection aérospatiale est utilisée pour améliorer la tactique de pêche, l'information issue d'un capteur est « consommée » dans la journée voire les heures qui suivent l'acquisition.

Au Chili, des données des températures de surface mesurées par satellites ont été utilisées pour définir des aires probables de concentration d'espadon et de germon (Barbieri et al., 1991). Les informations satellites étaient analysées à terre puis transmises aux pêcheurs. Il est intéressant de noter que le groupe cible était la pêche artisanale (bateaux de 7 à 15 mètres) plutôt qu'industrielle, et qu'en amont, toute une infrastructure a été mise en place pour la réussite de ce programme (cours de formation aux pêcheurs artisanaux, création d'un climat de confiance, suivi, etc.). Comme nous le verrons dans la section 2.2.1, c'est ce type de programme en partenariat qui a été mis en route à La Réunion à partir de 1992 (Petit et al., 1994).

⁷ CASI: *Compact Airborn Spectrographic Imager*.

D'autres études ont montré l'intérêt d'utiliser simultanément plusieurs paramètres. Par exemple l'utilisation combinée de données thermiques de surface et de données de la couleur de l'eau ont permis de définir les limites des zones de pêche aux germons (Laurs *et al.*, 1984; Montgomery *et al.*, 1986). Ces derniers tendent à se regrouper dans les eaux chaudes et claires de la zone frontale pour éviter les eaux plus froides à haute productivité mais turbides (Laurs, 1989b).

On peut citer encore des recherches de zones favorables aux concentrations de saury (*Cololabis saira*) par l'analyse des températures de surface et des gradients thermiques obtenus par satellite (Shinomi et Tameishi, 1988) ou la recherche des rassemblements de butterfish (*Perprilus burti*), espèce très liée à la proximité et à la persistance des systèmes frontaux (Herron *et al.*, 1989). Les satellites météorologiques géostationnaires et à défilement tels ceux de la série TIROS (NOAA) ont été très souvent mis à contribution et la liste des publications avec ce même profil peut s'allonger considérablement, chaque halieute y allant de la corrélation entre son espèce cible (souvent pélagique) et le paramètre de l'environnement (la température de surface, fréquemment). Les conclusions peuvent être contradictoires. Par exemple, Power et May (1991) rapportent une absence de corrélation entre les prises par unité d'effort de pêche palangrière de l'albacore dans le golfe du Mexique et les données thermiques de surface (température et gradients thermiques horizontaux) mesurées par satellite. Leur étude porte sur plus de 6 000 informations de pêche et 109 images satellites. Parmi les raisons évoquées pour expliquer cette absence de corrélation, on avance que la température de surface est très variable et ne donne pas nécessairement d'indications concernant les eaux profondes où l'albacore est pêché. En revanche, Santiago *et al.* (1993) ont trouvé que dans le golfe de Gascogne, 90% des pêches de germon se font entre 17 et 22°C.

Ces contradictions apparentes découlent en partie de la difficulté de décrire les niches écologiques marines et leur dynamique. La synthèse des nombreux travaux réalisés sur l'environnement du thon n'est pas aisée et quelquefois même conflictuelle. Ces études peuvent se répartir en trois groupes (Petit, 1991) :

- Les études de corrélations entre un facteur (comme la température de la mer) et les prises. Les résultats se ramènent habituellement à quelques courbes en cloche. Ces études monofactorielles peuvent être utiles pour délimiter l'habitat et connaître la plus grande extension de l'espèce, mais l'extrapolation à d'autres régions ou d'autres saisons est hasardeuse.
- Les juxtapositions des études précédentes (par exemple : températures de surface et oxygène). Ces juxtapositions montrent qu'un facteur ne peut pas être pris isolément. La règle du facteur limitant doit plutôt être remplacée par la notion de synergie entre les différents paramètres. Cela signifie que le comportement du thon dépend, à un niveau donné d'un des paramètres (par exemple la température de surface), du niveau des autres paramètres (par exemple l'oxygène, la salinité, *etc.*). Le comportement du poisson tendra à le ramener vers des conditions les plus proches possibles de sa fenêtre environnementale optimale (Petit et Hazane, 1983).
- Les études qui tentent d'intégrer l'ensemble des paramètres pour décrire cette fenêtre environnementale optimale plutôt que l'habitat lui-même.

Ces travaux donnent quasiment toujours des résultats positifs entre la variation des paramètres et l'abondance relative de thons. Toutefois, elles ne font pas toujours la distinction entre trois niveaux de probabilités : celle que le thon soit présent dans un environnement habitable (dont les paramètres sont à l'intérieur des limites physiologiques de l'animal) ; celle qu'il se concentre en bancs dans un milieu où les paramètres correspondent à sa fenêtre environnementale optimale ; et la probabilité de capturabilité ou vulnérabilité des bancs. Cette dernière dépend des conditions optimales requises pour une bonne efficacité de l'engin de pêche lors de la capture (Petit, 1991).

1.2.3 Suivi de la maturation des masses d'eaux

Si l'on admet que le comportement du thon est déterminé par sa recherche de nourriture, alors la localisation des bancs est corrélée à la localisation des zones à forte densité de nourriture. Ces dernières se trouvent dans des zones où la production primaire est suffisamment importante pour pouvoir supporter une forte production secondaire. Dans l'océan tropical, les zones de forte production primaire sont souvent associées aux zones de remontées de la thermocline. Si cette dernière atteint la surface, elle se traduira par un refroidissement en surface. C'est le cas par exemple des remontées d'eaux côtières. Si la thermocline remonte dans la couche euphotique sans atteindre la surface, elle passera inaperçue sur les images de température de surface mais peut être associée à une anomalie de la hauteur de l'eau négative. C'est le cas par exemple en présence de vents cycloniques.

Une diminution de la température de surface peut donc signaler le début d'un processus d'enrichissement de la masse d'eau. Un scénario thermique idéalisé proposé par Stretta (1991) consiste en une maturation de la masse d'eau pendant quatre semaines (par exemple avec une élévation de la température de surface de 1°C par semaine) suivie d'une stabilisation thermique pendant une à deux semaines. Cette période est considérée propice à la production d'animaux fourrage pour les thons. À l'issue de cette évolution thermique, un modèle peut donc considérer la zone comme favorable à leur présence.

Cette approche utilise l'évolution dans le temps de la température de surface comme la signature de la richesse d'une masse d'eau. Elle fait intervenir le concept du passé hydrologique. Cette vision dynamique représente un progrès par rapport à la vision statique des corrélations entre les prises de thonidés et les températures de surface relevées le jour de la pêche (Stretta, 1991).

La section précédente mentionnait l'intérêt d'examiner plusieurs paramètres pour délimiter les zones de présences probables de thons. Cette section a présenté la pertinence de s'intéresser au passé hydrologique des masses d'eaux. La présente thèse tentera de combiner les deux approches.

1.2.4 Description du paysage océanique

Dans le domaine terrestre, le « paysage » est un concept qui a donné lieu à de nombreuses études, controverses et polémiques, notamment entre les différentes écoles de géographes. Dans une récente synthèse sur le sujet, Fotsing (1998) remarque que « toute tentative de définition semble revenir au même point de départ : le paysage est une donnée visible, c'est ce qui se voit » et cite Deffontaines *et al.* (1987) : « Le paysage est une portion d'espace perceptible à un observateur, où s'individualise une combinaison de faits visibles et invisibles et d'interrelations, dont on ne perçoit, à un moment donné que le résultat global ».

Contrairement au domaine terrestre, ce qui était visible pour l'individu du domaine maritime était des plus uniforme, limité à la surface de l'eau. Seuls les « gens de mer » avaient aiguisé, adaptation oblige, leur moyen de perception et développé un sens marin. Les exemples qui précédent illustrent comment les capteurs de télédétection servent de traducteurs, d'amplificateurs d'une réalité à la fois de fine résolution et d'étendue considérable pour la perception individuelle. Ils aident à révéler le paysage océanique (Petit, 1995).

Pour cette étude, nous nous intéressons à l'environnement que perçoit le thon, et qui influe par là son comportement. Nous supposons que le thon perçoit un résultat global, et qu'il évolue donc dans un paysage océanique. Nous ne mesurons qu'une fraction des composantes de ce paysage ; ces mesures forment des « descripteurs du paysage océanique » qui seront définies dans la section 4.1.2 (page 91).

1.3 Les données disponibles sur les mouvements des thonidés

1.3.1 Données de marquages

Pour obtenir des informations sur les déplacements des thons, la méthode conventionnelle consiste à capturer quelques individus, attacher une marque sur chacun d'eux et les relâcher. Lorsque ces individus sont recapturés (généralement par des pêcheurs), on dispose d'un point de départ et d'un point d'arrivée de ces thons, mais sans information sur le trajet suivi entre ces deux points. Une méthode plus récente consiste à remplacer la marque classique par un instrument électronique (marque archive) mesurant des paramètres comme la luminosité ambiante et la pression du milieu ambiant. Ces paramètres permettent d'estimer le trajet suivi par le thon, mais ces données ne sont récupérées que si le poisson est recapturé. Même en cas de recapture des poissons, la récupération des marques archives est compliquée par le caractère multinational des pêches. Ces inconvénients peuvent être résolus par l'utilisation de marques « *pop-up* », qui se détachent du poisson à un instant préprogrammé, remontent à la surface et transmettent leurs données à des satellites ARGOS (Block et al., 1998). De telles marques archives sur le thon rouge de l'Atlantique ont confirmé que cette espèce pouvait migrer de l'ouest de l'Atlantique vers l'est, puis retourner à l'ouest dans la même année, et qu'elle peut aussi plonger jusqu'à 1000 mètres (Block et al., 2001).

1.3.2 Données de pêches

Les marques archives coûtent cher et leurs données sont rares. À cause du manque de données, les études des populations de thons (leurs abondances, distributions et déplacements) sont souvent effectuées à partir des données de pêches. Ces études utilisent des captures par unités d'effort en supposant une capturabilité⁸ constante et une distribution aléatoire de l'effort de pêche. En effet, si les pêches s'effectuaient à des positions aléatoires, alors il n'y aurait pas de corrélation entre les valeurs même si la population de thons sous-jacente a une structure spatiale. Dans ce cas, la moyenne et l'écart type de l'abondance auraient pu être estimés directement à partir des données (Petigas, 1996).

Malheureusement pour les gestionnaires, la distribution de l'effort de pêche n'est pas aléatoire. Elle suit la tendance générale des distributions de poissons à l'échelle des saisons, mais suit moins bien les variations aux échelles de quelques jours (Rose et Leggett, 1993). La capturabilité dépend de la distribution de l'effort de pêche relativement à celle du poisson (Swain et Sinclair, 1994), et l'écart-type de l'abondance ne peut pas être estimé sans faire intervenir un modèle de la distribution des bancs de thons. Le caractère non-aléatoire des distributions des thons et de l'effort de pêche, ainsi que les autres facteurs qui modifient la capturabilité (par exemple les changements de la technologie, ou les variations de la tactique de pêche) sont simplement ignorés (Brill, 1994a).

Les captures par unités d'efforts pourraient être un meilleur indicateur de l'abondance si on y combine des données océanographiques ainsi que des informations sur les capacités physiologiques des thons (Brill, 1994a). Les données environnementales sont déjà exploitées par des modèles de prédictions de la production primaire. Ces prédictions sont utilisées pour établir une échelle de probabilité de présence des poissons.

Plusieurs études ont tenté de délimiter l'habitat des thons à partir de statistiques de pêches et de conditions océaniques moyennées dans l'espace et dans le temps (Sund et al., 1981). Mais les analyses statistiques basées sur des captures moyennes et des données environnementales moyennes ne dévoilent pas toujours des corrélations (Brill, 1994a).

On distingue principalement deux méthodes de pêches, l'une active et visant les thons près de la surface, et l'autre passive et pouvant capturer des thons à de plus grandes profondeurs. Le type de pêche influe donc sur l'interprétation des données :

- **Pêche à la senne**

La pêche à la senne est une pêche de surface qui exploite la tendance des thons à se regrouper en bancs. Le bateau encercle un banc avec un filet vertical d'environ 1,5 kilomètre de longueur et de près de 150 mètres de profondeur. Le filet est ensuite refermé en tirant un câble qui passe dans un anneau attaché au bas du filet. Ce type de pêche est utilisé principalement dans les régions océaniques où la

⁸ Proportion de la population capturée par unité d'effort.

couche mélangée est peu profonde : une thermocline plus haute incite, le plus souvent, les thons à rester plus près de la surface.

Les thoniers senneurs sont des bateaux de 30 à 120 mètres pouvant stocker au maximum 4000 tonnes de poissons congelés. La flottille industrielle européenne de l'océan Indien correspond plutôt à des navires de 75 mètres en moyenne, pouvant embarquer 1200 tonnes de thon. Ces navires font des marées de plusieurs semaines. Leur cible favorite est traditionnellement l'albacore et le listao. Mais dans la dernière décennie, les prises de thon obèse ont considérablement augmenté avec la technique de capture sous objets flottants.

- **Pêche à la palangre**

Pour ce type de pêche, le bateau installe une ligne d'une longueur allant de quelques kilomètres jusqu'à 130 kilomètres. De cette ligne principale attachée à des flotteurs partent des lignes secondaires où quelques milliers d'hameçons sont suspendus. Des appâts naturels ou artificiels sont accrochés à ces hameçons. Le pêcheur peut choisir la profondeur de sa ligne, et capturer ainsi des thons à des profondeurs plus grandes que la pêche à la senne. La palangre est plus efficace si elle est posée à la profondeur préférée du thon, ou à la profondeur du plancher de l'habitat (profondeur limite des conditions acceptables de température et d'oxygène pour le thon). En moyenne les hameçons sont situés aux alentours de 150 à 200 mètres, où la température de l'eau est plus favorable aux thons de grande taille. La pose peut prendre plusieurs heures et la ligne est ainsi maintenue en place pendant 10 ou 20 heures. Pendant ce laps de temps, le pêcheur n'a pas de contrôle sur les espèces capturées (système passif). Cependant, en plus de la profondeur de la ligne, le pêcheur peut choisir le type deurre et cibler ainsi ses captures de poissons selon leur comportement, leur habitat ou leur taille (thons, espadons et marlins, requins, *etc.*). Traditionnellement, la cible préférée des pêcheurs est le thon obèse à destination du marché du sashimi. Les prises accessoires restent toutefois plus nombreuses que dans le cas de la pêche à la senne.

Les gros palangriers sont généralement asiatiques (Japon, Corée, Chine, Taïwan) et diffèrent sensiblement des petites unités artisanales des pays de la COI (Commission de l'océan Indien : Comores, Madagascar, Maurice, Réunion, Seychelles). Ces derniers ciblent actuellement plutôt l'espodon et le germon (marché frais) avec des lignes de 4 à 5 kilomètres. Ils opèrent beaucoup plus près de leurs côtes, contrairement aux palangriers asiatiques dont la pêcherie est mondiale.

- **Autres**

Il existe d'autres types de pêche aux thonidés (canne, ligne, filet, *etc.*), mais ils sont marginaux pour l'océan Indien. Nous ne les aborderons donc pas dans ce travail.

L'ensemble des captures palangrières avoisine 15% des prises mondiales de thonidés, contre plus de 60% pour la pêche à la senne. Dans l'océan Indien (surtout dans sa partie occidentale), la majorité des captures d'albacores sont effectuées par la pêche à la senne. Cette pêcherie s'effectue soit sur objets flottants (section 1.4.4) ou sur des bancs libres. Dans le premier cas, les captures sont composées principalement de petits albacores associés avec des listaos et des thons obèses juvéniles. Dans le cas des bancs libres, les individus capturés sont plus gros. Entre 1995 et 2000, les pêcheries européennes des senneurs sur des objets flottants représentaient 50 à 66% des calées (CTOI, 2003).

Les données de pêches utilisées dans le cadre de cette thèse seront présentées dans le chapitre 2.1.

1.4 Synopsis d'écologie thonière

Ces poissons pélagiques⁹ vivent principalement dans la couche mélangée au-dessus de la thermocline, dont la profondeur varie entre 10 et 150 mètres selon la région de l'océan et la saison. Le thon se nourrit d'autres poissons, de calmar et de crustacés. Ils ont une stratégie de nutrition opportuniste (Roger, 1994) et ne montrent pas d'exclusivité pour des proies d'une espèce particulière. Toutefois, à l'échelle régionale et à un instant donné, une petite quantité d'espèces peuvent représenter la quasi-totalité de l'alimentation des thons d'un groupe d'âge donné. Dans l'océan Indien, on trouve en majorité du calmar dans les estomacs de thons (Roger, 1994). On observe aussi des marques de coups de becs de calmar sur les thons pêchés, confirmant la présence de calmar dans les zones qu'ils prospectent. Cette nourriture, intermédiaire entre le zooplancton et le necton, se trouve dans des zones productives telles que les remontées d'eaux côtières, dômes et autres zones de divergence (Stretta, 1991). Les thonidés semblent se déplacer de l'une à l'autre de ces zones et parcouruent de très grandes distances ; un thon peut traverser un océan en quelques mois. On ne parle pas ici de migration, mais plutôt d'errance aléatoire. On réservera le terme « migration » aux déplacements orientés et liés à des cycles temporels (Collignon, 1991).

Un thon de première maturité consomme entre 7 et 15% de son poids par jour. Or, le thon se situe au sommet ou près du sommet de la pyramide alimentaire des océans pélagiques, et sa nourriture est concentrée dans des zones restreintes (distribution agrégative). Ces poissons doivent parcourir de grandes distances pour la trouver. Pour survivre, ils doivent être des prédateurs efficaces. Aussi leur physiologie est-elle particulièrement adaptée à la chasse de leurs proies. Les thons disposent d'une vision binoculaire (stéréoscopique) ainsi que des détecteurs chimiques (olfaction) et « physiques » (ligne latérale) sensibles. Surtout, ils disposent d'une relative homéothermie qui leur permet d'aller vite et loin. Mais cette physiologie bien adaptée coûte cher en énergie, donc en nourriture. C'est un « paradoxe écologique » dont la résolution est loin d'être complètement élucidée par les physiologistes et biologistes (Korsmeyer et Dewar, 2001). Enfin, le comportement gréginaire des thonidés (c'est-à-dire leur tendance à former des bancs) est aussi un atout pour prospecter plus et dépenser moins (Petit, 1991).

1.4.1 Adaptations physiologiques et comportementales aux facteurs abiotiques

Le thon est un poisson à régulation thermique interne. Il est proche de l'homéothermie, c'est-à-dire qu'il maintient une température interne plus élevée que celle de l'eau ambiante même lorsque son activité est faible (nage lente par exemple). Ce métabolisme nécessite lui-même l'utilisation d'une plus grande proportion de l'oxygène dissous dans l'eau. Le thon a perdu ou réduit plusieurs des structures nécessaires pour pomper l'eau à travers ses branchies. Il compense en nageant la bouche ouverte, ce qui force un jet d'eau à passer au travers de ses branchies. La concentration minimale requise d'oxygène dissous dépend de la vitesse de nage du thon, et n'est pas bien connue pour toutes les espèces. On a observé que le listao augmente sa vitesse de nage lorsque la concentration en oxygène descend en-dessous de 4 ml/l à 25°C, alors que l'albacore ne montre aucune différence de comportement à des concentrations aussi basses que 2,5 mg/l (Dizon, 1977). Le thon obèse tolère des concentrations aussi basses que 0,6 ml/l (Lowe et al., 2000).

On croit qu'une température interne plus élevée permet au thon de maintenir un plus haut taux de glycolyse¹⁰, et donc de bénéficier plus rapidement de l'énergie chimique nécessaire à un effort soudain. Le taux métabolique peut augmenter en situation d'effort pour atteindre des valeurs exceptionnellement élevées (Brill, 1996). On peut aussi présumer que la digestion se fait plus rapidement lorsque la température des viscères est plus élevée. L'animal est ainsi plus rapidement prêt à se nourrir. Le principal avantage du taux métabolique élevé serait ainsi non pas de nager en soutenant une vitesse élevée (ou du moins pas uniquement), mais plutôt d'accélérer la digestion et de permettre une récupération rapide après un exercice épuisant (Brill, 1996). Comme chez certains requins, les performances métaboliques du thon sont liées à une capacité peu ordinaire de fonctionnement des muscles blancs en aérobic comme en anaérobic (Dickson *et al.*, 1993).

La régulation thermique du thon est imparfaite. Un de ses principaux problèmes physiologiques est d'éviter la surchauffe. Le thon thermorégule sa température par un comportement qui profite de la proximité

⁹ Qui vivent en pleines eaux, libres de tous contacts avec le fond et qui ne dépendent pas significativement du benthos pour leur nourriture (Collignon, 1991).

¹⁰ Première étape d'une série de réactions chimiques de la respiration et de la fermentation conduisant à la dégradation de glucose en acide pyruvique ou en acide lactique. Cette conversion fournit à la cellule de l'énergie sous forme de deux molécules d'ATP.

d'eaux chaudes et d'eaux plus froides de part et d'autre d'une thermocline. Dans la région de Hawaï, le thon obèse nage le jour dans des eaux entre 14 et 17°C à environ 250 mètres de profondeur, à partir desquelles il effectue périodiquement des excursions rapides dans les eaux de surfaces plus chaudes (Holland et al., 1990). Durant ces excursions, le thon peut subir des augmentations de température allant jusqu'à 12°C par minute.

Chez tous les poissons, le sang quitte les branchies à la température de l'eau ambiante. Dans le cas des thons, ce sang artériel est réchauffé en passant à travers un réseau dense de veines transportant le sang chaud en provenance des muscles. Ce système de rétention de la chaleur, avantageux en-dessous de la thermocline, peut devenir contre-productif lorsque le thon nage rapidement dans les eaux de surface. Les mesures de la température du corps et de la profondeur de deux thons obèses en océan ouvert (près d'Hawaï) ont montré que ces derniers ont la capacité de changer rapidement (en l'espace de 45 secondes) le taux de transfert de chaleur par un facteur 100 (Holland et al., 1992).

Les thons se réchauffent donc plus rapidement durant leurs remontées qu'ils ne se refroidissent durant leurs descentes. Le taux de transfert de la chaleur peut aussi varier lors des déplacements horizontaux. En faisant nager des thons dans des tunnels soumis à un gradient de température, des expériences ont montré que la température des muscles de l'albacore diminue plus rapidement lorsque la température de l'eau passe de 32 à 28 °C que lorsqu'elle passe de 26 à 22 °C (Dewar et al., 1994). Un gradient horizontal de 4°C n'a donc pas le même impact pour des températures initiales différentes.

Cette capacité physiologique de thermorégulation permet aux thons obèses d'étendre leurs niches écologiques vers les eaux froides en dessous de la thermocline. Des marques électroniques insérées dans le ventre de thons obèses dans le Pacifique sud indiquent que cet animal peut plonger à des profondeurs supérieures à 1000 mètres (Leroy, 2003). Les températures minimales requises par les différentes espèces de scombridés ne sont pas bien connues. Dans le cas du marlin rayé, ce minimum semble être non pas une valeur fixe, mais plutôt 8°C de moins que la température de la couche mélangée (Brill et al., 1993).

Différentes espèces, ou différentes tailles de la même espèce, n'exploiteront pas le même habitat vertical du fait des différences dans leurs capacités physiologiques. Par exemple durant le jour, le listao et l'albacore occupent des eaux moins profondes que le thon obèse (Holland et al., 1990). Les premiers passent la majorité de leur temps dans la couche mélangée ou juste au-dessus de la thermocline, et ne plongent qu'occasionnellement dans les eaux plus froides. A l'opposé, le thon obèse reste bien en-dessous de la thermocline (dans des eaux de 15°C) pendant le jour, et ne fait que de brèves excursions dans les eaux plus chaudes de la couche mélangée. Ces différences dans les déplacements verticaux ne s'expliquent toutefois pas que par des capacités de thermorégulation différentes, puisque l'albacore et le thon obèse ont des capacités à peu près équivalentes (Brill et al., 1999).

1.4.2 Adaptations comportementales aux facteurs biotiques

Dans une revue de la littérature, Brill (1994a) note que les auteurs s'accordent à dire que les conditions de température et d'oxygène influencent fortement les mouvements, distributions et capturabilités des thons. Mais des observations simultanées des mouvements de thons et de leurs proies montrent qu'une partie seulement des déplacements des thons peuvent s'expliquer par des facteurs abiotiques limitants. À de petites échelles spatio-temporelles et à l'intérieur de l'habitat délimité par des facteurs abiotiques, la structure tridimensionnelle de l'environnement biotique et sa dynamique (densité des proies, migrations quotidiennes...) sont des facteurs clés des déplacements horizontaux et verticaux des thons (Josse et al., 1998).

La couche diffusante (*deep scattering layer*, abrégée DSL) se trouve partout dans toutes les mers du monde. Elle se compose entre autres de céphalopodes (*Japetella diaphana*) ainsi que d'espèces de poissons (*Argyropelecus aculeata*, *Myctophum selenops*, *Scopelarchus analis*, *Diaphus mollis*) dont les thons se nourrissent parfois (Dagorn et al., 2000d). La nuit, les organismes de la DSL remontent près de la surface (Longhurst, 1976). Au lever du soleil, ils redescendent vers les couches plus profondes (en-dessous de 400 mètres en Polynésie française) et y restent tout le jour. On croit que la DSL est à peu près uniforme, mais elle n'est pas homogène lorsqu'elle remonte la nuit ; elle dépend de la structure océanographique.

Une hypothèse intéressante serait que les thons descendent dans la DSL lorsqu'ils ne trouvent pas de nourriture conséquente en surface, afin d'y chercher le minimum nécessaire à leur survie. La migration nycthémérale typique du thon obèse suit celle de la DSL : le thon passe la nuit près de la surface (dans les premiers 100 mètres de profondeur) et le jour à de plus grandes profondeurs, entre 350 et 450 mètres. Ces migrations verticales s'effectuent au lever du jour et à la tombée de la nuit. À ces migrations s'ajoutent,

durant le jour, des excursions fréquentes du thon obèse dans les eaux de surfaces plus chaudes, pour des raisons de régulations thermiques (section 1.4.1) et peut-être aussi pour récupérer d'une dette en oxygène. Ces excursions se font environ toutes les 2½ heures pour les thons obèses en Polynésie Française, et toutes les heures pour les thons obèses plus petits (environ 12 kg) près des îles Hawaïennes (Dagorn et al., 2000d).

L'albacore et le thon obèse remontent la nuit. Le germon serait la seule espèce qui ne remonte pas la nuit, mais cette hypothèse est difficile à confirmer étant donné que le germon est aussi la seule espèce pour laquelle on ne dispose pas de données de marquage.

La variété de comportements observée dans les mouvements verticaux est souvent associée aux capacités physiologiques inégales des différents individus ou espèces de thon (section 1.4.1). Toutefois un modèle numérique simple, basé uniquement sur une distribution des proies similaire à celle de la DSL, suggère que cette variété peut aussi s'expliquer par une émergence de stratégies co-adaptées à l'exploitation du même environnement (Dagorn et al., 2000a). Cette variété apparaît sans qu'il ne soit nécessaire d'introduire des contraintes physiologiques au modèle.

1.4.3 Relations entre paramètres environnementaux et distribution

On considère que les paramètres environnementaux importants pour le thon sont la température de surface, la concentration en oxygène dissous et la salinité (Pesce et Gentile, 2003). Les thons tendent à se concentrer le long de discontinuités thermiques comme les fronts océaniques (Sund et al., 1981; Sharp, 2001). Toutefois, la tendance de l'albacore à nager peu de temps dans la couche de surface (0 – 20 mètres) soulève un doute sur la pertinence des cartes de température de surface pour guider les bateaux de pêches vers des zones de pêches potentiellement favorables, sauf si ces zones de pêches sont marquées par un gradient de température fort (Cayré, 1991).

Dans les annales du tournoi de pêche des poissons porte-épée¹¹ à Hawaï, les captures de l'année 1995 sont significativement différentes de celles des autres années. Cette année là, un tourbillon cyclonique a dominé l'océanographie de la côte de Kona durant toute la semaine du tournoi. Or il est connu que les tourbillons cycloniques dans les eaux sub-tropicales favorisent la remontée de la nutricline vers la couche euphotique. Cette remontée favorise elle-même une augmentation locale de la production biologique. Une couche de surface bien mélangée et de forts courants induits par la présence du tourbillon ont caractérisé les eaux où ont eu lieu les plus grandes captures du makaire bleu du Pacifique (*Makaira mazara*), ce qui suggère une réponse biologique directe (par exemple des limitations physiologiques) ou indirecte (par exemple la disponibilité des proies) du makaire bleu à son environnement (Seki et al., 2002).

1.4.4 Dispositifs concentrateurs de poissons (DCP)

Les mouvements horizontaux des thons tropicaux sont connus pour être affectés par des structures physiques bien définies comme les monts sous-marins, pentes récifales (« outer reef drop-offs ») et les dispositifs concentrateurs de poissons (DCP, ou FAD¹² en anglais). Près de la moitié des captures mondiales de thon (par poids) sont faites sur des bancs associés à des objets flottants, des mammifères ou des monts sous-marins (Fréon et Dagorn, 2000). La colonisation d'objets flottants par les thons peut s'effectuer en quelques jours (Bard et al., 1985). Les objets noirs ainsi que les objets de taille moyenne semblent plus attractifs que les très gros ou très petits objets. Les autres caractéristiques (volume, couleur) ne semblent pas avoir d'effet significatif (Hall et al., 1999). D'ailleurs, lors de campagnes de marquages à Hawaï et en Polynésie Française, on a observé que le listao s'associait parfois à l'objet flottant et parfois au bateau, avec des mouvements de va-et-vient entre les deux. Le caractère attractif des objets flottants, peu importe leur nature, semble donc fort pour ces poissons (Fréon et Dagorn, 2000). Les données de marquages de thons capturés près d'un FAD révèlent des comportements que l'on peut classer en 4 groupes (Dagorn et al., 2000b) :

- Les individus qui quittent le FAD et ne montrent aucune tendance à y retourner pendant la durée du marquage.
- Les individus qui restent à quelques centaines de mètres du FAD pendant toute la durée du marquage. Les thons peuvent rester plusieurs jours au même point, « emprisonnés » par un objet flottant, sans

¹¹ En anglais: *billyfish*.

¹² FAD: *Fish Aggregating Device*.

quantité suffisante de nourriture au point de perdre du poids (Suzuki, communication personnelle citée dans Dagorn et Fréon, 1999).

- Les individus qui passent la journée près du FAD le quittent la nuit et reviennent au même FAD ou à un FAD adjacent le jour suivant.
- Les poissons qui visitent un FAD y restent quelque temps (pas nécessairement la journée entière) et reprennent leur chemin.

À partir de marquages ultrasoniques et de modèles numériques, on estime le rayon d'attraction d'un dispositif FAD à environ 7-13 km (Dagorn et al., 2000d; Dagorn et al., 2000c). Une demi-douzaine d'hypothèses ont été proposées pour expliquer ces associations des thons avec les objets flottants. Dans une discussion de ces hypothèses, Fréon et Dagorn (2000) considèrent les deux suivantes comme les plus crédibles :

- **Indicateurs de régions plus productives**

Les objets flottants pourraient être des indicateurs de régions plus productives. Par exemple ces indicateurs peuvent provenir de régions plus riches (embouchures des rivières, mangrove...) et s'être déplacés avec ces masses d'eaux plus riches. Ils peuvent aussi s'agréger au milieu de zones de converges dans les fronts.

- **Point de rencontre**

En supposant que les objets flottants sont plus rares que les poissons et se détectent à une plus grande distance, ils pourraient servir à augmenter les chances de rencontre entre des individus isolés ou des petits bancs. Il est admis que les bancs doivent avoir une taille minimale pour profiter de ses avantages (Fréon et Misund, 1999), et les points de rencontre serviraient à augmenter la taille des bancs qui sont en dessous de leur taille optimale. Les chances de rencontre sont encore accrues si les thons retournent à leur point de rencontre à des moments précis, ce que suggèrent des observations sur des thons marqués (Klimley et Holloway, 1999). Ils ont observé que des thons quittaient l'objet flottant dans des directions différentes, mais y revenaient à peu près au même moment (avec un décalage de moins de deux heures, parfois de moins d'une minute). D'un point de vue énergétique, il est plus économique d'attendre un congénère à un point de rencontre que de nager pour le trouver, surtout si le point de rencontre est un objet flottant dérivant avec la masse d'eau.

Dans le cas de l'albacore et du thon obès, on trouve une plus grande proportion de petits individus pêchés dans les bancs associés que dans les bancs libres¹³. Dans le cas du listao, aucune différence n'a été observée (Fréon et Dagorn, 2000). Un point de rencontre pourrait être particulièrement avantageux pour les petits individus. D'abord parce qu'ils sont plus vulnérables à la prédation (on croit que les gros bancs réduisent les risques). Ensuite parce qu'ils doivent être plus nombreux pour atteindre la taille optimale d'un banc. Les bancs composés d'individus nombreux étant plus difficiles à maintenir, les petits individus pourraient avoir besoin de grossir leurs rangs plus souvent. Cette hypothèse pourrait expliquer pourquoi la proportion de petits individus dans les bancs associés est plus grande que dans les bancs libres (Fréon et Dagorn, 2000).

1.4.5 Comportement gréginaire

Le comportement gréginaire est l'inverse du comportement territorial : c'est la tendance des individus de certaines espèces à se regrouper par attraction réciproque. Cette tendance peut amener la formation d'un *banc*, un ensemble d'individus se déplaçant activement dans la même direction. Par opposition, la *foule* désigne plutôt un groupe d'individus rassemblés sous la pression de contraintes externes (Collignon, 1991).

Un banc n'est pas une structure fixe ; il peut s'étaler et se déformer rapidement. Un banc peut durer des jours ou des mois, mais n'est jamais définitif ; il peut se disloquer, et deux bancs peuvent se fusionner. Les poissons se disposent de façon à réduire la résistance hydrodynamique, en gardant entre eux une distance qui dépend de leurs tailles. Les bancs les plus stables sont donc composés d'individus de tailles similaires (mais pas nécessairement de la même espèce). Outre l'avantage hydrodynamique, un modèle analytique

¹³ Toutefois, les individus sont plus nombreux dans les bancs associés, de sorte que le tonnage total est plus grand. Cette interprétation est fondée sur l'hypothèse que le tonnage pêché reflète le tonnage du banc plutôt qu'une différence de vulnérabilité due aux conditions différentes.

suggère que les bancs sont rentables pour la prospection de nourriture, jusqu'à un certain tonnage maximal (Petit, 1991).

Lorsque les thons sont en bancs, ils bénéficient de la prospection de l'ensemble : repérage plus rapide de la nourriture, plus de temps pour se nourrir, échantillonnage plus efficace et transfert d'informations (Petit, 1991; Fréon et Misund, 1999). Les bancs ont toutefois tendances à éclater dans les milieux pauvres (pas assez de nourriture pour tous) ou dans des milieux riches (pas de contraintes). C'est dans les milieux avec une densité de nourriture moyenne que l'on retrouve les plus gros bancs.

1.4.6 Reproduction

La fécondité de la plupart des poissons est élevée, de l'ordre de 100 000 à 1 million d'œufs par kilogramme de femelle. Leurs pontes libèrent une forte biomasse planctonique dans une zone restreinte (les frayères) pour un cours laps de temps. Ce concentré de biomasse inerte est immédiatement exploité par les prédateurs. L'éclosion des survivants survient de quelques jours à quelques semaines après la ponte, et libère une larve transparente (l'œuf aussi était transparent). Ces larves sont faiblement actives, souvent concentrées en « essaims » planctoniques, et elles aussi intensément exploitées par les prédateurs.

Les larves vivent sur les réserves de leur sac vitellin. Elles évoluent en post-larves et leur bouche s'ouvre à la phase dite du « sevrage ». Cette dernière phase s'accompagne toujours d'une forte mortalité pour les larves qui ne trouvent pas rapidement de la nourriture. Les post-larves sont encore victimes de prédation (dont les thonidés eux-mêmes), mais deviennent elles-mêmes des consommateurs voraces de plancton. Leur mobilité augmente progressivement et les post-larves survivantes atteignent la phase de juvéniles, où elles passent du plancton au necton (Collignon, 1991).

1.5 Synopsis des courants dans l'océan Indien

La circulation de l'océan Indien est dominée par la mousson. La mousson désigne un ensemble de phénomènes climatiques et océanographiques. Il comprend des vents saisonniers qui, au nord de 10°S, soufflent avec de légères variations pendant une moitié de l'année dans une direction et pendant l'autre moitié dans la direction opposée. La principale cause est la variation annuelle de la température sur de grandes surfaces continentales, plus importantes que les variations sur les surfaces océaniques voisines. Le contraste thermique cause un excès de pression sur le continent en hiver et un déficit en été. D'autres facteurs, tels que le relief, jouent aussi. Les vents qui en résultent, en plus de changer de direction, sont beaucoup plus forts durant la mousson d'été que durant la mousson d'hiver.

Le renversement des vents de mousson sur la partie nord de l'océan Indien force un renversement des courants dans les couches supérieures. Les figures 5 et 6 représentent les courants de surface durant les moussons d'été et d'hiver respectivement. Le courant océanique entre la mer d'Oman (*Arabian Sea* en anglais) et la baie du Bengale s'inverse selon un cycle saisonnier. Il s'écoule vers l'est durant la mousson d'été (mai à septembre), et vers l'ouest durant la mousson d'hiver (novembre à février). Ce courant est identifié par l'abréviation SMC dans la figure 5 et NMC dans la figure 6. Le courant de Somalie s'écoule vers le nord durant la mousson d'été, et vers l'équateur durant la mousson d'hiver. Un contre-courant équatorial (SECC) s'écoule vers l'est durant les deux moussons (en avril-mai et en octobre-novembre) et s'inverse entre les moussons. Contrairement au courant de mousson qui ne s'inverse que deux fois par année, le contre-courant équatorial le fait donc quatre fois. Le cycle de la mousson peut être résumé comme suit :

- **Mousson d'été**

La mousson d'été commence à se former dans le sud de la baie du Bengale en mai. Le vent souffle alors en provenance du sud-ouest sur la majorité de la partie nord de l'océan Indien. Il s'intensifie pour atteindre un sommet en juin, avec un forçage d'Ekman vers l'est dans la baie du Bengale et dans la mer d'Oman (avec une légère courbe anticyclonique dans ce dernier cas). Durant sa phase la plus mature (en juin), le courant dans la mer d'Oman est une continuation du courant de Somalie et du courant côtier d'Oman. Les courants atteignent alors 100 cm/s (2 noeuds) près de la côte de Somalie. Octobre est le mois de transition entre la mousson d'été et d'hiver, avec des vents faibles sur toute la partie nord de l'océan Indien.

- **Mousson d'hiver**

La mousson d'hiver commence à se former au sud du Sri Lanka en novembre. Le forçage d'Ekman est dirigé vers l'ouest dans la mer d'Oman et la baie du Bengale. La mousson s'intensifie en décembre pour atteindre un sommet en janvier, avec des vents en provenance du nord-est sur la majorité de la partie nord de l'océan Indien. Durant sa phase mature (décembre à mars), le courant s'écoule vers l'ouest au sud de la baie du Bengale. Les courants atteignent 30 cm/s dans le sud ouest de la mer d'Oman. La mousson commence à faiblir en février.

La majorité des variations du courant de mousson se produisent dans les 100 premiers mètres, ce qui en fait un courant peu profond comparativement au courant de Somalie (Shankar et al., 2002). La régularité du cycle saisonnier n'est qu'une facette de mousson. Une autre facette est la grande variabilité interannuelle du courant de mousson. De plus, les climatologies et les moyennes mensuelles suggèrent des courants relativement continus, alors que les images dérivées des données altimétriques satellitaires nous montrent des paysages assez différents. Les tourbillons y sont nombreux et sont parfois le trait dominant de la circulation (Shankar et al., 2002).

Au sud de 10°S, les variations saisonnières sont moins prononcées. Vers 17°S, un courant sud-équatorial (SEC) coule vers l'ouest et se sépare en deux branches près de Madagascar (figure 5 et 6). La branche sud atteint une zone de rétroflexion près de l'extrémité sud de Madagascar, où une partie retourne dans le gyre subtropical et l'autre partie continue vers le courant des Aiguilles. La branche nord passe le cap au nord de Madagascar et se fusionne avec le courant côtier africain est (EACC), mais une fraction tourne vers le sud vers le canal du Mozambique (Schott et al., 2002).

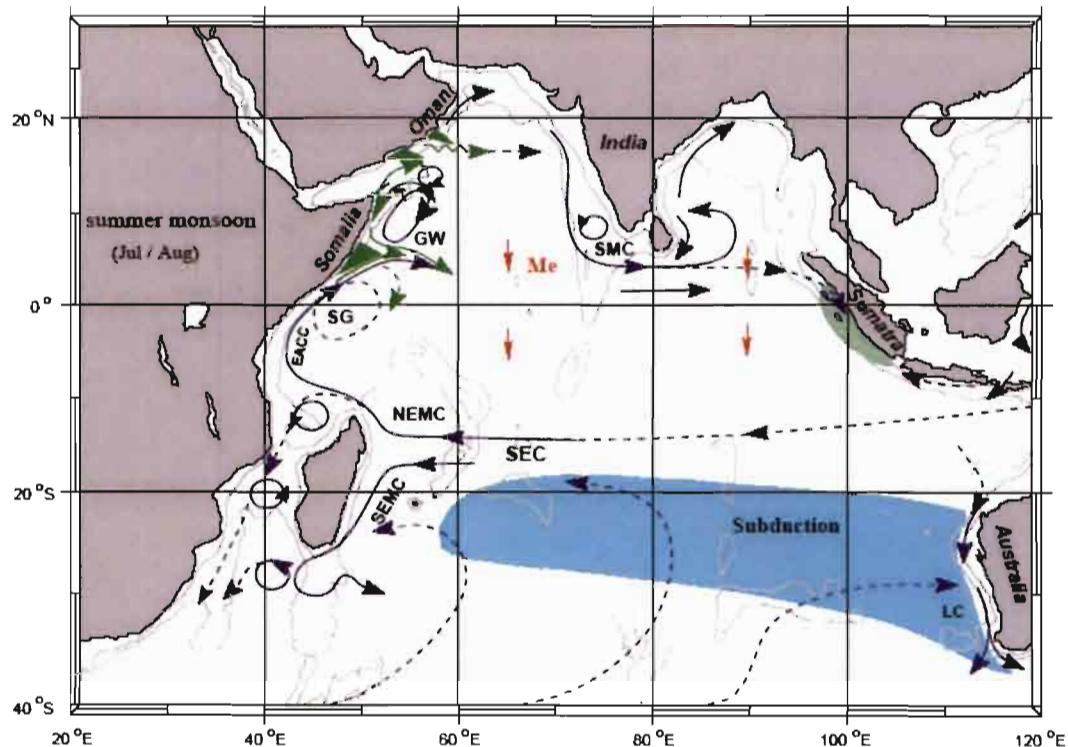


Figure 5: Circulation dans l'océan Indien lors de la mousson d'été

Les zones de subductions sont représentées en bleu, les remontées en vert, et le transport d'Ekman de part et d'autre de l'équateur par des flèches rouges.

Source: Schott et al., 2002

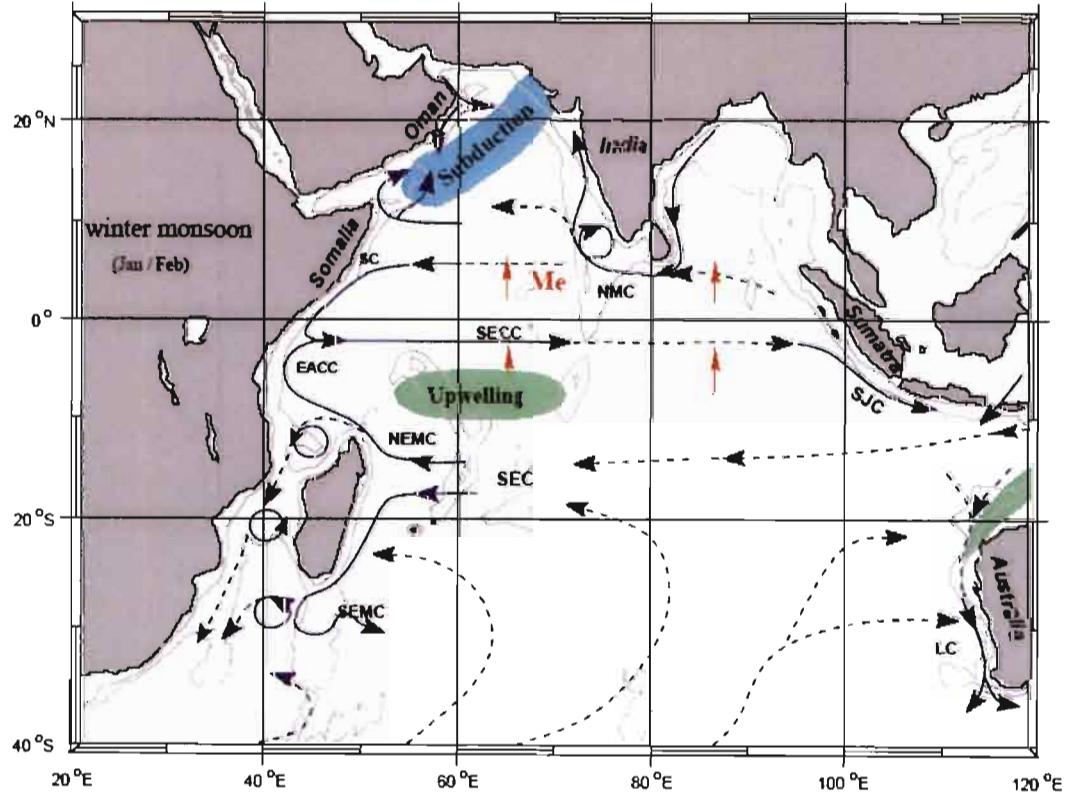


Figure 6: Circulation dans l'océan Indien lors de la mousson d'hiver

Même légende que pour la figure 5.

Source: Schott et al., 2002

1.5.1 Composantes du courant net

Le courant de mousson est causé à la fois par le forçage direct du vent (transport d'Ekman) et par la propagation de forçages qui ont eu lieu ailleurs. Cette propagation peut se faire par des ondes barocliniques, notamment l'onde de Kelvin équatoriale, l'onde de Rossby équatoriale et l'onde de Rossby côtière (Shankar et al., 2002). Ces dernières se propagent rapidement dans les tropiques. Or le bassin de l'océan Indien est confiné dans les tropiques à cause de sa bordure nord (environ 25°N), de sorte que ces ondes peuvent traverser le bassin en quelques mois.

- **Contribution du forçage d'Ekman**

La spirale d'Ekman, forcée par les vents de mousson, apporte une contribution importante (parfois dominante) aux courants de surface (Shenoi et al., 1999). Cette contribution atteint 100 cm/s près de la côte de Somalie en été. Le courant d'Ekman est assez régulier. Vers 5°N, il forme un flot continu vers l'est en juillet et vers l'ouest en janvier. L'intensité de ce courant croît et décroît à peu près en même temps à travers tout le bassin (Shankar et al., 2002).

- **Contribution du courant géostrophique**

Contrairement aux courants d'Ekman, les courants géostrophiques sont dominés par les tourbillons. Ils ne se forment ni ne décroissent en même temps partout dans le bassin. Ils ne forment un flot continu près de 5°N qu'en juillet (régime d'été) ainsi qu'en janvier-février (régime d'hiver). Pendant les autres mois, la circulation contient une « relique » du régime précédent. Ces reliques se trouvent de plus en plus confinées vers le bord ouest du bassin à mesure que l'on s'approche du régime suivant. Elles sont dues entre autres à la propagation vers l'ouest d'anomalies de la hauteur de l'eau associées au régime précédent. Ces propagations ont été associées à des ondes de Rossby (Shankar et al., 2002).

La mousson d'hiver est un courant principalement géostrophique, modulé par les courants d'Ekman. Durant la mousson d'été, de forts vents font dominer les courants d'Ekman à la surface, ce qui produit une structure verticale plus complexe durant la mousson d'été que durant la mousson d'hiver. Pendant la mousson d'été, la dominance du courant d'Ekman élimine le signal des ondes de Rossby en surface (environ 5 mètres). Les effets de ces ondes se font sentir à des profondeurs plus grandes (35 mètres), quoique plus faiblement que pendant les autres mois de l'année (Shankar et al., 2002).

1.5.2 Les remontées d'eaux

L'océan Indien se distingue des autres océans par une absence de remontée d'eaux à l'équateur, sauf lors d'événements climatiques inhabituels. Cette différence est causée par les alizés qui soufflent de l'ouest plutôt que du sud-est, entraînant une convergence plutôt qu'une divergence. La plupart des remontées d'eaux se font plutôt dans des zones côtières dans l'hémisphère nord durant la mousson d'été (Schott et al., 2002). Une zone de subduction subtropicale dans l'hémisphère sud (figure 5) est connectée aux zones de remontées côtières de Somalie et d'Oman par le courant de Somalie.

Les deux principales zones côtières de remontées d'eaux sont la côte de Somalie et la côte de la péninsule arabique. La remontée sur la côte de Somalie commence en avril-mai avec des vents soufflant vers le nord-est, parallèlement à la côte. Le courant qui traverse l'équateur tourne vers l'est lorsqu'il atteint 3 ou 4°N (figure 5). Au plus fort de la mousson d'été, la remontée d'eau de Somalie n'est pas due principalement à la divergence d'Ekman le long de la côte. Elle est davantage due à la remontée de la thermocline à la gauche du courant qui s'éloigne de la côte (Schott et al., 2002).

Au début du mois de juin, un fort vent anticyclonique s'établit et entraîne la formation d'un tourbillon dans la partie nord de la côte de Somalie (le « Great Whirl », abrégé GW dans la figure 5). Une zone de remontée s'établit aussi à gauche de la partie du tourbillon qui longe et s'éloigne de la côte.

La limite nord des alizés du sud-est se situe vers 5 à 10°S. Cette région est associée à une divergence d'Ekman, plus intense durant la mousson d'hiver (figure 6). On croit que cette région est le siège d'une remontée, ce qu'appuient les images satellitaires de température et de la couleur de l'eau. Ces dernières montrent parfois un bloom phytoplanctonique dans cette région (Schott et al., 2002).

Chapitre 2

Présentation des sources de donnée utilisées

2.1 Les données halieutiques

Pour cette thèse, toutes les observations sur les thons nous sont fournies par les pêcheurs. Comme expliqué dans la section 1.3.2 (page 22), les données de pêches sont considérablement plus abondantes que celles que l'on peut obtenir par des campagnes de marquage. Mais contrairement aux données de marquage, les informations des pêcheurs ne fournissent que des indices sur l'abondance des thons à des positions ponctuelles, sans aucune indication sur leur provenance. De plus, les positions prospectées par les pêcheurs ne sont pas aléatoires, mais tendent à suivre, entre autres, les distributions saisonnières des populations de thons visées. Cette corrélation devra être prise en compte lors de l'interprétation des résultats.

Cette thèse utilise les données de pêches à la senne, qui est une pêche active de surface (section 1.3.2). Les données de pêches à la palangre, un engin passif, et les analyses dont elles ont fait l'objet ont été présentées dans le rapport Palangre (Petit et Huyn, *sous presse*) et ne seront pas discutées dans cette thèse. Les données utilisées ici proviennent des flottilles de senneurs françaises et espagnoles dans l'océan Indien, auxquelles se mêlent quelques bateaux français battant pavillon étranger. Dans le cas des flottilles françaises, les pêcheurs ont fourni (pour certains) leurs informations en échange des images de température de surface captées par la station SEAS-Réunion (section 2.2.1), préfigurant le projet SEAGERH. L'IFREMER collecte les cahiers de pêches sur lesquels les patrons inscrivent leurs observations sur les zones prospectées et le tonnage des bancs qu'ils ont capturés. Mais la qualité de ces données est variable selon les individus et difficile à déterminer :

- Les positions sont approximatives, même si la plupart des pêcheurs disposent aujourd'hui d'un système de positionnement GPS. Souvent, la position dans le cahier de bord n'est pas celle de la pêche, mais plutôt la position relevée la dernière fois que le patron a fait le point (généralement le midi).
- La composition des espèces indiquée par le patron est généralement approximative. Par exemple, lorsqu'un patron déclare avoir capturé du listao, il s'agit en général d'un mélange d'espèces de tailles similaires au listao. Le thon obès et l'albacore sont aussi fréquemment confondus.
- Les pêcheurs ont généralement tendance à sous-estimer le tonnage capturé. Certains ne déclarent pas toutes leurs captures.

Nous ne pouvons pas rendre les données de pêches totalement objectives, mais nous pouvons tenter de réduire les écarts dus aux différences d'interprétation. A partir des informations fournies par les pêcheurs, deux corrections sont appliquées par l'IRD :

- En utilisant les données de campagnes d'échantillonnages et en tenant compte de la saison, la proportion probable des espèces est estimée. Cette correction ne change pas le tonnage total déclaré.
- En calculant le rapport du tonnage débarqué sur le tonnage déclaré, l'IRD obtient un facteur de correction (le tonnage débarqué étant la mesure la plus précise). Cette correction est ensuite appliquée sur les tonnages déclarés à chacun des coups de sennes.

Les tonnages sont détaillés par espèces pour le thon obès, l'albacore, le listao et le germon (chapitre 1.1.1). Une catégorie « autres » regroupe plusieurs espèces dont la précision des estimations est très incertaine. Détailer ces espèces donnerait une fausse impression de précision. Enfin, les tonnages par espèces sont subdivisés en tonnages par catégories :

- La catégorie 1 désigne les bancs composés d'individus de moins de 10 kg.
- La catégorie 2 désigne les bancs composés d'individus de 10 à 30 kg. Cette catégorie a été omise dans les données qui nous ont été fournies, car elle est souvent confondue avec la catégorie 3.
- La catégorie 3 désigne en principe les bancs composés d'individus de plus de 30 kg. En pratique, à cause de la confusion mentionnée au point précédent, nous interprétons la catégorie 3 comme des bancs d'individus de plus de 10 kg.

Conformément à la littérature (page 27), l'IRD observe que lorsqu'un pêcheur capture un banc composé d'une seule espèce de petite taille, il s'agit souvent d'un banc associé à un objet flottant. Les coups de sennes sur les objets flottants sont souvent composés de thon obès, de listao et d'albacore. Les proportions dans le

cas de nos données sont quantifiées à la figure 31 (page 95), sauf le mélange des espèces qui n'est pas représenté.

Les captures de bancs libres étaient plus nombreuses autrefois, car les pêcheurs connaissent aujourd'hui l'efficacité des dispositifs concentrateurs de poissons (section 1.4.4). Cette thèse analyse principalement les bancs libres, ce qui justifie l'intérêt d'utiliser des données historiques. Des données sont disponibles depuis 1980, mais sous forme de journaux de bord et sans distinction d'espèces. Les données que nous utilisons couvrent donc les années 1991 à 2000.

Les tonnages capturés ne sont pas nécessairement représentatifs de l'abondance de la ressource. Des mesures plus précises sont les captures par unités d'effort, que l'IRD estime à partir de la durée des prospections. Cette dernière s'estime en calculant (dans l'ordre) :

- **Nombre d'heures passées en mer**

Le pêcheur peut être en mer sans pêcher ni faire de prospection. C'est le cas par exemple lorsqu'il doit réparer une avarie.

- **Nombre d'heures passées en pêche (incluant les prospections)**

Ce nombre est toujours inférieur ou égal au nombre d'heures passées en mer.

- **Nombre d'heures passées en prospection**

Ce nombre est calculé à partir du nombre d'heures passées en mer auquel on soustrait le nombre d'heures utilisé pour les coups de sennes. La durée des coups de sennes n'est pas mesurée par le pêcheur ; elle est estimée par le laboratoire des études halieutiques de l'IRD en fonction de la saison et du tonnage. Ces estimations ne peuvent évidemment pas tenir compte d'éventuels problèmes rencontrés par le pêcheur, comme un filet qui se déchire. En outre, l'estimation du temps de pêche est compliquée par les progrès technologiques. On constate en effet que, à tonnage égal, la durée des coups de senne tend à diminuer depuis 10 ans.

Les données nous ont été fournies dans une table contenant une position pour tous les coups de sennes, nuls ou pas. Dans le cadre de cette thèse, elles ont été insérées dans une base de données dont la structure sera présentée dans la section 3.5.3 (page 78).

2.2 Les données satellitaires

La plupart des données satellitaires utilisées dans ce travail ont été élaborées par des organisations bien connues comme la NOAA. Plus spécifiquement, nous avons utilisé les produits suivants (ces produits sont décrits plus en détails dans Desruisseaux et al., sous presse) :

- **Concentration en chlorophylle-a**

Nous avons utilisé les données GAC (*Global Area Coverage*) dérivées du capteur SeaWiFS¹⁴. La résolution spatiale des données au nadir¹⁵ est de 1 km et la fauchée du capteur est de 2800 km. Ses huit canaux du vert au proche infra-rouge analysent avec une grande précision les composants de l'eau et leurs concentrations. Le canal 1 est utilisé pour distinguer les substances jaunes ; les canaux 2,3,4 pour calculer la concentration en pigments chlorophylliens ; les canaux 6,7,8 pour évaluer la turbidité et corriger les effets atmosphériques¹⁶. Dans le cas des eaux du large, la couleur de l'eau dépend essentiellement de la concentration chlorophyllienne. En revanche, pour les eaux côtières, l'interprétation des mesures radiométriques est plus complexe en raison de la présence simultanée de la matière organique dissoute, des particules dissoutes et des pigments chlorophylliens. La quantification de ces trois constituants demande de nombreuses données de terrain, et généralement on obtient des modèles locaux, assez peu génériques.

- **Température de surface de la mer**

Nous avons utilisé les données GAC (*Global Area Coverage*) dérivées du radiomètre à infrarouge thermique AVHRR sur les satellites NOAA. La résolution spatiale au nadir est de 1,1 km. La série de capteurs NOAA fournit des données de manière constante depuis plus de vingt ans. Ces données sont très utilisées en raison de leur faible coût et de la possibilité de les capter à partir d'une antenne HRPT comme celle de la station SEAS (section 2.2.1) à La Réunion. La mesure radiométrique permet de dériver la température des premiers micromètres de la colonne d'eau. Si les couches inférieures sont homogènes (en présence de vent ou en absence de réchauffement solaire superficiel), la mesure sera significative de la couche mélangée. En revanche si les couches inférieures sont très stratifiées, la mesure surfacique pourra être très différente de la mesure volumique. Les mesures de nuit ou en présence de vent moyen seront donc plus significatives. Si le vent est très fort, on assiste au phénomène inverse, moins marqué, de sous-estimation de la température de la mer à cause des embruns. La présence de nuages entraîne aussi une sous-estimation de la température.

La résolution des données GAC (à la fois pour la température de surface et pour la chlorophylle-a) est d'environ 9 km, même si la résolution du capteur était plus proche de 1 km. La résolution temporelle est de 8 jours. Dans les deux cas, seules les séries d'images du répertoire « *best pixels* » ont été utilisées. Dans le cas particulier de la température, une autre série d'images produite par la station SEAS (section 2.2.1) a été utilisée et comparée aux données GAC

- **Pompage d'Ekman**

Nous avons estimé le pompage d'Ekman à partir des mesures de la vitesse du vent à la surface de la mer. Ces dernières sont dérivées des données du diffusiomètre radar QuikSCAT. La précision de la vitesse du vent est d'environ 2 cm/s. La précision du pompage d'Ekman que nous calculons reste à déterminer. L'incertitude risque d'être élevée, puisque le calcul fait intervenir des différences de la vitesse du vent à des positions voisines, et que ces différences sont souvent inférieures à la précision déclarée de 2 cm/s (Desruisseaux et al., sous presse).

- **Anomalie de la hauteur de l'eau**

Nous avons utilisé les données dérivées du capteur Topex/Poseidon et traitées par le centre AVISO à Toulouse. L'orbite couvre la région entre 66° de latitude nord et sud et se répète tous les 10 jours. L'orbite choisi couvre la majorité des océans qui ne sont pas couverts par les glaces. Par rapport aux mesures de températures et de concentration en chlorophylle-a, les mesures de la hauteur de l'eau ont l'avantage de ne pas être perturbées par la présence de nuages.

¹⁴ SeaWiFS : Sea-viewing Wide Field-of-view Sensor.

¹⁵ « Au nadir » signifie à la verticale du passage du satellite.

¹⁶ Dans les longueurs d'onde du visible, la luminance mesurée par le capteur est perturbée en grande partie lors de son passage dans l'atmosphère. Les particules (eau, aérosols) absorbent fortement le rayonnement, et le diffusent aussi.

2.2.1 La station SEAS (Surveillance de l'Environnement Assistée par Satellite)

Mise en place par l'ORSTOM¹⁷ en 1991, la station SEAS est une des stations HRPT¹⁸ qui contribue à l'acquisition, l'archivage et le traitement des données pour le programme EARTHNET de l'agence spatiale européenne (ESA). Cette station participe à l'effort de coopération internationale pour intégrer dans un réseau coordonné les recherches qui se font à partir des données AVHRR. Elle procède à l'acquisition et l'archivage des données des satellites NOAA et SEAWiFS selon la norme de l'ESA. Son aire d'acquisition est conséquente, allant des grands lacs africains jusqu'au milieu de l'océan Indien. De 7 à 10 passages sont acquis par jour.

En plus de transmettre les données collectées à l'ESA, la station SEAS élabore aussi des produits finis à l'intention d'acteurs plus régionaux. Les algorithmes utilisés intègrent l'angle de visée, la correction atmosphérique et les spécificités régionales. Les produits obtenus incluent la température de surface LAC (*Local Area Coverage*), d'une résolution d'une minute d'angle (environ 1,6 kilomètre). Il ne s'agit pas du seul produit de cette station, mais les analyses statistiques des autres produits ont été reportées à une étude ultérieure.

Dans la région s'étendant du canal du Mozambique aux îles Chagos, soit dans la partie nord de l'ellipse d'acquisition de la station SEAS, s'est installé au milieu des années quatre-vingt l'essentiel de la pêche thonière industrielle de surface européenne. Cette installation était due au déplacement rapide de la flottille de senneurs français et espagnols de l'Atlantique vers l'océan Indien (section 1.1.2). Par ailleurs, dans la partie sud de cette aire d'acquisition, s'est développée depuis 1992 une pêcherie palangrière semi-industrielle au thon et à l'espadon très dynamique avec des unités de 16 à 25 mètres.

Trois types d'intérêts sont en jeu. Celui du pêcheur est de diminuer son temps de recherche et sa consommation d'énergie (fuel) dépensée par rapport à celle embarquée (poisson). Celui de l'administrateur gouvernemental des pêches est de pouvoir suivre la dynamique des stocks et de l'effort de pêche à la fois dans le temps et l'espace. Enfin, l'intérêt du scientifique est de mieux connaître les relations entre le poisson et son environnement. Ces trois groupes sont confrontés à la grande variabilité de l'environnement océanique. Ce contexte a favorisé un partenariat entre ces acteurs, dans lequel les différentes informations de la station sont mises à disposition sur un serveur. Les unités de pêches peuvent y accéder par le réseau de télécommunication INMARSAT B, D, M ou C. Le capitaine dispose ainsi à bord d'éléments décisionnels. En retour, les pêcheurs transmettent leurs carnets de pêches avec les prises détaillées et les conditions de captures, ainsi que la route des navires et de paramètres relevés *in-situ* tel que la température de surface. C'est ce partenariat qui nous a permis d'obtenir les données de pêches mentionnées dans le chapitre 2.1.

2.2.2 La topographie de la surface de l'eau

La masse liquide des océans ne présente pas une surface homogène ; au contraire, « vallées et collines » se succèdent et, de plus, sont mobiles. Ces déformations sont dues à plusieurs phénomènes physiques. L'attraction terrestre tout d'abord détermine la topographie statique dont l'océan s'écarte rapidement suite au mouvement de rotation de la Terre et des vents qui sans cesse entraînent avec eux les eaux de surface. C'est ce que l'on nomme la topographie dynamique de surface. Elle est de l'ordre de deux mètres d'amplitude. D'une connaissance de la topographie dynamique précise, on peut donc en déduire des informations sur la vitesse et la direction des courants. Les altimètres de ERS et surtout de TOPEX/POSEIDON ont la précision requise en mesurant la distance altimétrique avec une erreur de l'ordre du centimètre.

2.2.2.1 Principe des mesures de la hauteur de l'eau

L'altimètre fonctionne en envoyant une courte salve d'impulsions radar vers la Terre et en mesurant le temps requis pour que les impulsions reviennent au capteur après avoir été réfléchies par la surface de la mer. La distance obtenue est la *distance altimétrique*, illustrée à la figure 7. Le niveau de la mer est la différence entre l'altitude du satellite (figure 7) et la distance altimétrique.

¹⁷ ORSTOM est l'ancien nom de l'IRD.

¹⁸ HRPT: *High Resolution Picture Transmission*.

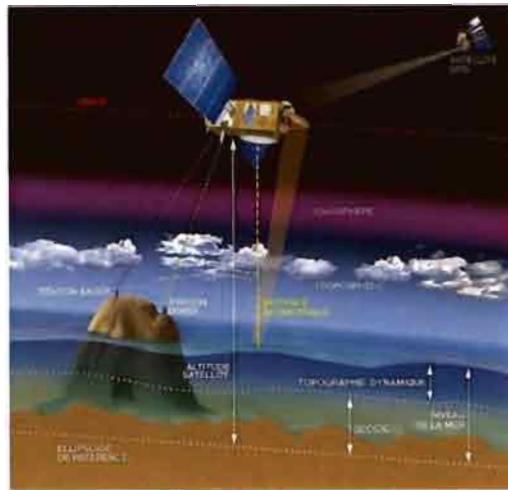


Figure 7: Principe des mesures altimétriques (source: AVISO)

Le niveau de la mer est mesuré par rapport à un ellipsoïde¹⁹ de référence (figure 7). Les mesures du niveau de la mer sont particulières du fait que leur valeur (en plus de leur position sur une carte) dépendent de l'ellipsoïde choisi. Les mesures altimétriques dont nous disposons utilisent un ellipsoïde spécifique au projet TOPEX/POSEIDON. Le tableau 3 donne les paramètres de cet ellipsoïde, ainsi que celui couramment utilisé WGS 1984.

Tableau 3: Paramètres de l'ellipsoïde WGS 1984 et de celle de TOPEX/POSEIDON

	WGS 1984	TOPEX/POSEIDON
Longueur du demi-axe équatorial:	6378,137 km	6378,1363 km
Longueur du demi-axe polaire:	6356,752 km	6356,7516 km
Aplatissement ²⁰ :	1/298,257223563	1/298,257

Le niveau de la mer est mesuré par rapport à l'ellipsoïde de référence, tandis que la topographie dynamique (la mesure qui nous intéresse) est plutôt mesuré par rapport au géoïde. Le géoïde est la hauteur qu'aurait la mer en l'absence de toute perturbation (marées, vents, courants, etc.). Il correspond à une surface équipotentielle du champ gravitationnel terrestre.

La topographie dynamique nous renseigne sur les courants, qui sont une des causes des écarts entre la hauteur réelle de la mer et son géoïde. Ces courants comprennent une circulation stationnaire (due entre autres à la rotation de la Terre et aux vents permanents) et une partie fortement variable qui dépend de la marée et des conditions météorologiques.

2.2.2.2 Le pourquoi des anomalies de la hauteur de l'eau

L'anomalie de hauteur de la mer est la hauteur de la mer par rapport à une moyenne, d'où le terme « anomalie » ou « déviation ». Il existe une variété de moyennes, qui peuvent être annuelles ou saisonnières. Les anomalies calculées dépendent donc des critères choisis pour déterminer la moyenne.

Nous utilisons des anomalies de hauteurs plutôt que la topographie dynamique (figure 7) afin de pallier notre connaissance insuffisante du champ gravitationnel terrestre, qui affecte les prédictions de l'orbite des satellites ainsi que du géoïde terrestre. Puisque la gravité ne varie pas dans le temps, elle affecte de la même

¹⁹ Un ellipsoïde est une forme géométrique relativement simple (une sphère aplatie aux pôles) qui donne une approximation de la forme complexe de la Terre. L'aplatissement polaire est pris en compte mais pas le relief de la croûte terrestre. Il est décrit par les longueurs de ses axes équatorial et polaire, dont les valeurs « idéales » dépendent de la région géographique étudiée. Les ellipsoïdes sont un des paramètres des projections cartographiques. Plusieurs images satellites (par exemple les images de température acquises par l'antenne de l'IRD-Réunion) sont projetées sur une carte qui utilise l'ellipsoïde WGS 1984.

²⁰ L'aplatissement est calculée par $(a-b)/a$ où a est la longueur du demi-axe majeur (équatorial) et b est la longueur du demi-axe mineur (polaire). Cette valeur est 0 pour une sphère et croît vers 1 pour les ellipsoïdes plus aplatis.

manière chaque passage du satellite au-dessus d'un même point. En retranchant la hauteur moyenne de l'eau, on supprime cette erreur mais on fait disparaître du même coup la circulation océanique moyenne. Néanmoins, les cartes d'anomalies sont utiles pour observer comment les courants varient sur de courtes périodes de temps et de courtes distances en fonction des tourbillons, des conditions de vents ainsi que des changements de la température dans la couche de surface de l'eau.

2.2.2.3 Capteurs disponibles

Le satellite GEOSAT²¹ de la marine américaine a été lancée le 12 mars 1985 et s'est arrêté en janvier 1990. Un an et demi après son lancement, il fut placé sur une orbite répétitive de 17 jours pour suivre les variations du niveau des océans pendant plus de trois ans. Ses données sont disponibles à la communauté scientifique via la NOAA²². Il a été remplacé par le satellite GFO²³, lancé sur la même orbite le 10 février 1998.

Le satellite TOPEX/POSEIDON (figure 8) est le fruit d'une coopération entre la France (CNES) et les États-Unis (NASA). Il a été lancé le 10 août 1992 avec pour mission « d'observer et comprendre la circulation océanique ». Son orbite se répète tous les 10 jours environ. Son successeur, JASON, a été lancé en 2001.

Le centre AVISO (situé à Toulouse) traite, valide, archive et distribue les données du satellite TOPEX/POSEIDON ainsi que de d'autres missions altimétriques (notamment les satellites ERS). Les opérations quotidiennes d'AVISO sont assurées par le laboratoire CLS²⁴.

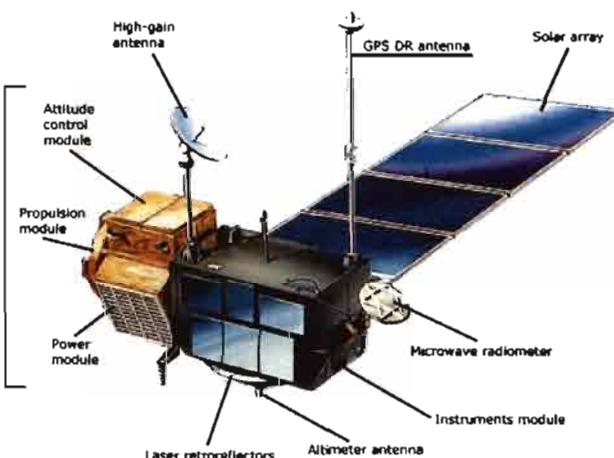


Figure 8: Schéma du satellite TOPEX/POSEIDON (source: AVISO)

Le satellite TOPEX/POSEIDON est composé de deux altimètres, TOPEX et POSEIDON, qui se partagent la même antenne. On ne peut donc faire fonctionner qu'un seul altimètre à la fois. L'altimètre POSEIDON étant expérimental, c'est l'altimètre TOPEX qui fonctionne dans la majorité des cas (90% du temps). Mais on observa une dégradation de la performance du capteur A de l'altimètre TOPEX à partir du cycle 130 (fin mars 1996). Afin d'éviter une dérive des mesures, l'instrument B fut utilisé à la place du A à partir du cycle 236 (début décembre 1999). Les mesures de ce dernier se sont révélées cohérentes avec celles des altimètres POSEIDON et ERS-2.

L'orbite de TOPEX/POSEIDON est inclinée à 66° par rapport à l'axe nord-sud de la Terre. Elle couvre donc la région entre 66° de latitude nord et sud (figure 9) et se répète tous les 10 jours. L'orbite choisit résulte d'un compromis entre plusieurs besoins. Elle couvre la majorité des océans qui ne sont pas couverts par les glaces. Elle est assez haute pour minimiser la friction dans l'atmosphère, ce qui aide à déterminer plus précisément l'orbite.

²¹ GEodesic SATellite.

²² National Oceanographic and Atmospheric Administrations.

²³ Geosat Follow-On.

²⁴ Collecte Localisation Satellite.

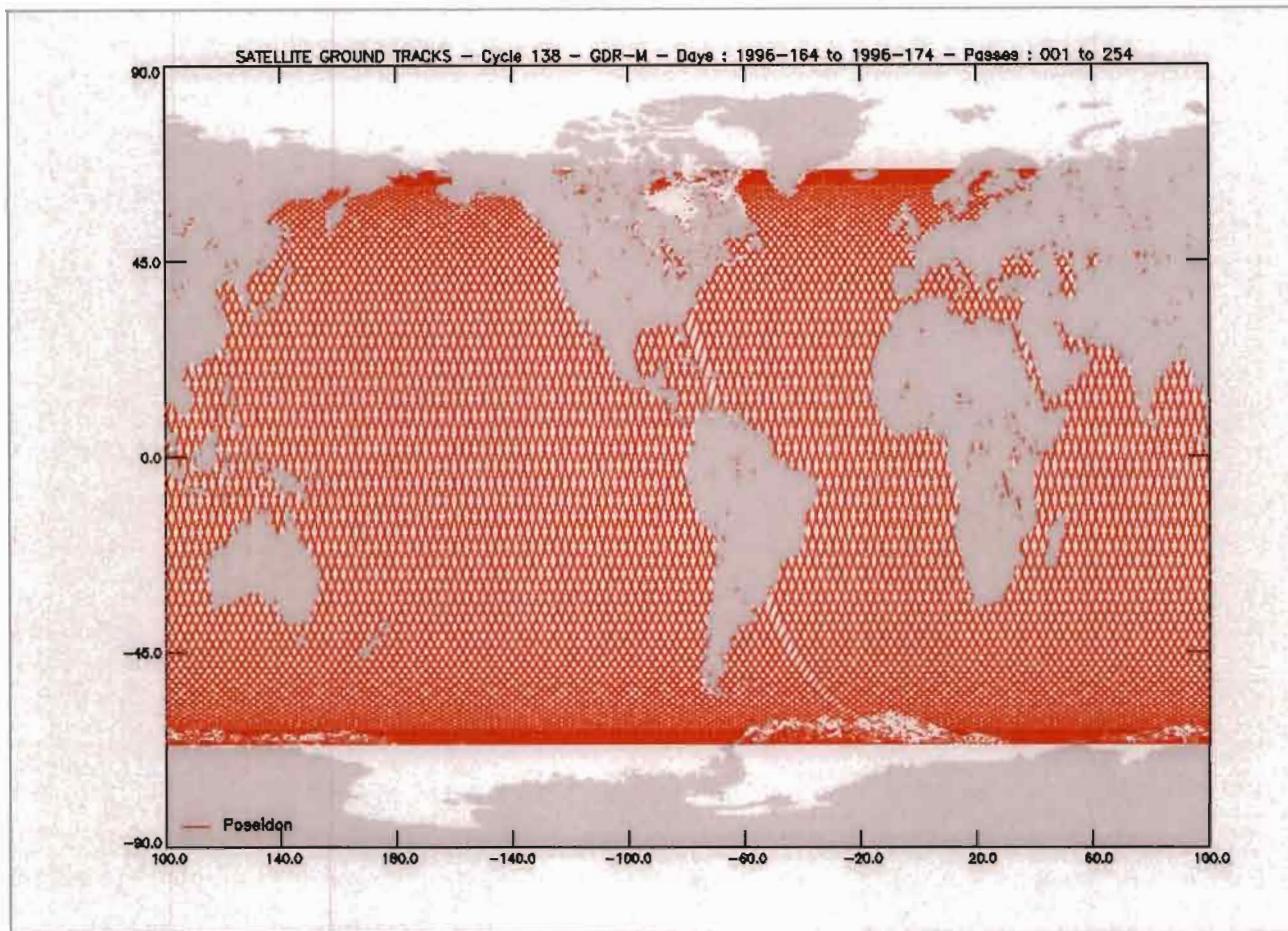


Figure 9: Trace au sol du satellite TOPEX/POSEIDON (source: AVISO)

La figure 9 illustre bien le fait que les traces ascendantes et descendantes dessinent des losanges. La distance moyenne entre les traces est égale à la demi-diagonale latitudinale. A l'équateur, cette distance est de 300 km pour TOPEX/POSÉIDON et de 85 km pour ERS. Dans les deux cas l'altimètre effectue une mesure par seconde, i.e. tous les sept kilomètres le long de la trace. L'échantillonnage spatial du premier est donc plus lâche que celui du second. En revanche, la durée du cycle de TOPEX/POSÉIDON est de 10 jours, soit un échantillonnage temporel plus de trois fois plus fin que celui d'ERS dont la période est de 35 jours. En combinant les données en provenance des deux satellites, on en optimise donc les caractéristiques.

2.2.2.4 Sources d'erreur

Trois types de sources d'erreurs peuvent être distinguées : des erreurs liées au trajet de l'onde, des erreurs liées au bruit d'autres phénomènes, enfin des erreurs liées à des manipulations techniques (généralement bien maîtrisées).

2.2.2.5 Effets lors du trajet de l'onde

Pour obtenir une estimation précise du niveau de l'eau, il faut que la vitesse de propagation de l'onde soit bien connue. Or, la propagation de l'onde est affectée par de nombreux facteurs, qui doivent être pris en compte pour corriger les observations de l'altimètre. La troposphère sèche amène l'altimètre à sous-estimer la hauteur d'environ 230 cm, tandis que la troposphère humide l'amène à sur-estimer la hauteur de 6 à 30 cm. L'ionosphère introduit aussi une erreur de 0,2 à 20 cm. L'importance de corriger ces erreurs apparaît clairement lorsque l'on songe que la précision désirée est de l'ordre de 2 cm.

La vitesse de propagation de l'onde radar est diminuée par les gaz et la quantité de vapeur d'eau dans l'atmosphère terrestre. L'influence de la masse totale de l'atmosphère (*troposphère sèche*) est assez prévisible. Mais l'influence de la vapeur d'eau (*troposphère humide*) est plus variable et plus difficile à déterminer.

Le nombre d'électrons libres dans l'*ionosphère* (figure 7) influence la vitesse de propagation de l'onde. Cette quantité d'électrons varie entre le jour et la nuit (il y a très peu d'électrons libres la nuit), entre l'été et l'hiver (il y en a moins durant l'hiver de l'hémisphère sud) et en fonction du cycle solaire.

2.2.2.6 Les perturbations « indésirables »

Si l'on s'intéresse à la topographie dynamique de la mer (figure 7), il faut aussi tenir compte des perturbations « indésirables », telle que la marée. Certaines erreurs peuvent provenir des vagues. Le capteur reçoit davantage d'ondes réfléchies dans les creux des vagues qu'à leurs crêtes. La hauteur moyenne à laquelle l'onde est réfléchie paraît donc légèrement plus basse que la hauteur moyenne de la mer. Cette erreur peut être corrigée si on connaît la *hauteur significative* des vagues ($\xi^{1/3}$). Or cette hauteur peut être déterminée à partir de la forme du pulse retourné vers l'altimètre. Une mer calme retourne un pulse étroit, tandis qu'une mer agitée avec des vagues plus hautes retourne un pulse plus étiré. En effet, les ondes réfléchies sur les crêtes des vagues parviendront à l'altimètre avant celles qui ont été réfléchies dans les creux des vagues. Les impulsions émises par les altimètres ont toujours une certaine étendue spatiale et une certaine épaisseur. Sur la figure 10, on observe que la courbure de l'onde sphérique du pulse, combinée avec la courbure de la Terre, fait que l'onde n'atteint pas la surface de la mer partout en même temps. La région dans laquelle l'onde ne tarde pas de plus de δz (l'épaisseur du pulse) est appelée « première zone de Fresnel » et les pulses des altimètres sont généralement un peu plus larges que la première zone de Fresnel. Notons que la mesure de la hauteur significative des vagues n'est possible qu'à condition que la hauteur $\xi^{1/3}$ des vagues soit supérieure à l'épaisseur δz du pulse (figure 10).

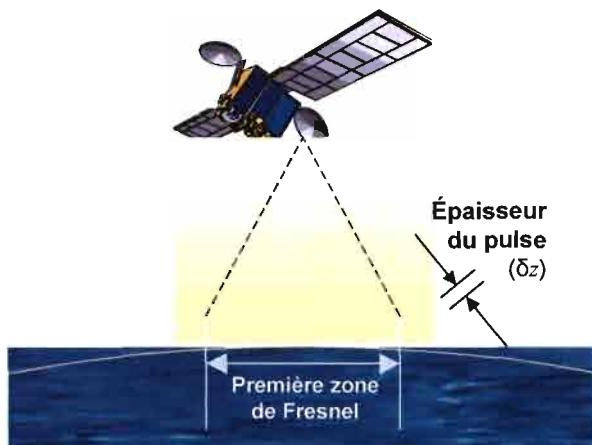


Figure 10: Géométrie d'un pulse d'altimètre sur la surface de la mer

La hauteur de la mer réagit aussi à la pression atmosphérique. Une augmentation de la pression atmosphérique de 1 mbar abaisse le niveau de la mer de 1 cm. Notons toutefois que la réponse de la surface de la mer aux variations de la pression atmosphérique est encore l'objet d'études. Les valeurs des effets barométriques sont séparées des valeurs corrigées dans les fichiers de données fournis ; ils pourront donc être recalculés si de besoin lorsqu'on connaîtra mieux cet effet.

Après avoir obtenu une mesure précise de la distance séparant l'altimètre de la surface de la mer, il faut encore prendre en compte les marées. Non seulement les forces gravitationnelles externes génèrent les marées océaniques, mais elles déforment aussi la Terre elle-même (ce qu'on appelle la marée solide). La réponse de la Terre à ces forces externes est suffisamment rapide pour que l'on puisse supposer un état d'équilibre. La surface de la marée solide est alors parallèle à la surface équipotentielle. Il existe donc une constante de proportionnalité entre la hauteur de la marée solide et le potentiel gravitationnel. Cette constante est à peu près indépendante de la fréquence de la marée, sauf à une fréquence proche de celle de la composante K1 à cause d'une résonance dans le manteau liquide.

L'axe de rotation de la Terre oscille avec des périodes apparentes de 12 et 14 mois. Du point de vue d'un observateur terrestre, il en résulte une force centrifuge supplémentaire dont l'effet sur la surface des océans ne se distingue pas des marées. On appelle ce phénomène la « marée polaire » et on considère sa période suffisamment longue pour permettre un équilibre avec à la fois l'océan et la Terre solide.

2.2.2.7 Les manipulations techniques

A l'altitude du satellite TOPEX/POSEIDON (1330 km), la friction de l'atmosphère est encore suffisante pour faire lentement baisser l'orbite du satellite. D'autres perturbations sont causées par des variations dans le champ gravitationnel de la Terre ainsi que dans la pression de la radiation solaire. De temps à autres, des manœuvres sont nécessaires pour maintenir le satellite sur son orbite. La fréquence de ces manœuvres dépend principalement du flux solaire, et varie entre 40 et 200 jours. Ces manœuvres durent de 20 à 60 minutes et sont effectuées de préférence lorsque le satellite survole les continents. Aucune mesure scientifique n'est prise pendant ces manœuvres.

En conclusion, les données apparaissant dans les fichiers fournis par CLS ont subi toutes ces corrections, excepté la correction de la pression atmosphérique qui apparaît dans un champ séparé. Pour le satellite TOPEX/POSEIDON, la précision moyenne instantanée de l'estimation locale du niveau de l'océan est inférieure à 5 cm, et la précision moyenne sur un mois inférieure à 2 cm.

Chapitre 3

Élaboration des outils d'investigation

3.1 Les services logiciels

3.1.1 Problématique

La problématique halieutique de cette thèse, à savoir mettre en évidence les relations entre les données de pêches et les données environnementales, nécessite le recouplement de trois domaines informatiques : il nous faut des méthodes d'interrogation des bases de données, des outils de système d'informations géographiques (SIG) et des outils statistiques. Chacun de ces trois groupes d'outils est l'objet de standards entérinés par des organismes internationaux, ou à tout le moins de standards de fait :

- L'industrie informatique a standardisé les accès aux bases de données relationnelles depuis plusieurs années, grâce au langage SQL²⁵. Ce langage est utilisable à partir du Java par un lien JDBC²⁶. Ce standard est fermement établi et ne sera pas présenté ici.
- Les méthodes de manipulation d'informations géographiques sont en voie de standardisation, grâce au consortium OpenGIS (section 3.1.3) et à l'ISO²⁷. Mais les efforts en ce sens sont plus récents et leur intégration dans les logiciels moins répandue. Au moment d'écrire ces lignes, aucun produit de notre connaissance n'implémente la totalité des spécifications. On trouve plutôt de nombreuses implémentations partielles, dont Geotools 2 (section 3.1.5).
- Les analyses statistiques ne sont pas, à notre connaissance, l'objet de standards informatiques. Toutefois, le langage S développé par le laboratoire Bell constitue un environnement largement accepté. Le logiciel commercial S+ et le projet libre R en sont deux implémentations. Notons qu'il existe des ponts entre R et Grass (un logiciel libre de SIG), ce qui ouvre d'intéressantes perspectives de géostatistiques. Des ponts JNI²⁸ vers le Java existent à la fois pour R et Grass.

Ces trois groupes d'outils sont représentés de gauche à droite dans la figure 11. Leurs services servent de fondations à « Seagis » (section 3.1.6), un logiciel qui tente de répondre à la problématique halieutique de cette thèse. Les boîtes bleues de la figure 11 représentent les standards, en vert les implémentations libres (*Open Source*) et en rouge les produits commerciaux. La partie jaune désigne les contributions faites dans le cadre de cette thèse.

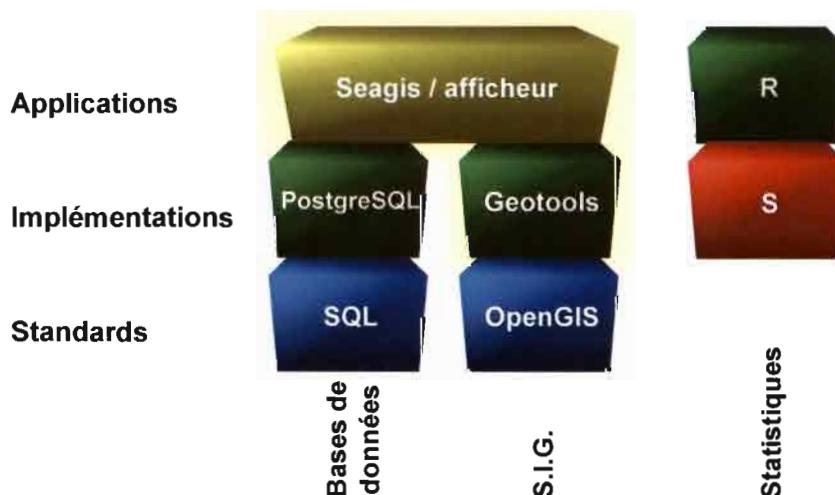


Figure 11: Services logiciels requis (de gauche à droite), des standards jusqu'aux applications

²⁵ **SQL:** *Structured Query Language* – Langage standardisé d'interrogation des bases de données.

²⁶ **JDBC:** *Java DataBase Connectivity* – Permet notamment d'exécuter des instructions SQL à partir du Java.

²⁷ **ISO:** *International Organization for Standardization* – Organisation internationale de normalisation.

²⁸ **JNI:** *Java Native Interface* – Permet les appels de fonctions en C/C++ à partir du Java, et vice-versa.

3.1.2 Interopérabilité

Les diverses informations nécessaires pour notre étude des corrélations entre les données de pêches et les données environnementales sont structurées de manière hétérogène. Elles comprennent à la fois des données matricielles (14 000 images réparties dans 30 séries) et des données vectorielles (150 000 positions de prospection des senneurs, ou 4 000 géométries des lignes palangrières). Les propriétés de ces données (enveloppe spatio-temporelle, système de coordonnées, codage des valeurs, *etc.*) varient en fonction du paramètre mesuré et des contraintes des appareils de mesure. De plus, l'exploitation de données de sources extérieures (Nasa, Aviso, *etc.*) accentue encore la difficulté. Une telle variété agit comme un frein aux études qui requièrent des combinaisons de données hétérogènes.

Nous ne pouvons pas imposer un format uniforme à l'ensemble des données, car la diversité des formats tient à des facteurs tels que la méthode de mesure et la distribution statistique des valeurs. Une solution plus flexible consiste à assurer l'interopérabilité des données à travers une interface de programmation (API²⁹) commune. Pour que cette démarche puisse être pérennisée dans des travaux postérieurs à cette thèse (prise en compte de nouveaux capteurs, extension à d'autres régions, à d'autres types de pêches, *etc.*), l'API doit être largement accepté par des développeurs indépendants. Autrement dit, l'API doit être un standard industriel.

De manière générale, le respect des standards exige un effort plus grand que ce qu'aurait requis un développement isolé, mais se rentabilise par un double avantage : en plus d'accroître l'interopérabilité de nos données avec celles des projets externes, il nous indique aussi une voie robuste à suivre pour nos propres développements internes. En effet, les groupes d'experts qui conçoivent les standards anticipent des difficultés qui échappent parfois à l'ingénieur en début de projet, mais qui risquent de le rattraper avant la fin.

3.1.3 Apport du consortium OpenGIS®

Une communauté d'informations géospatiales est un ensemble de systèmes ou d'individus capables d'échanger leurs données géospatiales grâce à des définitions et des standards communs ainsi qu'une reconnaissance réciproque. Le réseau SEASNET qui fournit les données satellitaires de cette thèse en est un exemple. Comme il existe une multitude de façons de représenter des informations géospatiales, chaque communauté est amenée à les structurer en fonction de ses centres d'intérêts. Cette diversité complique la tâche des utilisateurs de SIG en les plaçant devant une variété apparemment chaotique de formats et de structures de données. Le consortium OpenGIS est né de cette constatation et tente d'apporter de l'ordre dans ce chaos. Il regroupe des entreprises et institutions telles que MapInfo, Sun Microsystems, Oracle, la Nasa, l'O.N.U. et des universités, et offre deux technologies clés :

- permettre à une communauté d'annoncer leurs informations, de manière à ce que des individus ou des systèmes en dehors de cette communauté puissent les découvrir ;
- transférer des informations d'une communauté vers une autre en préservant leurs sémantiques, même si les deux communautés utilisent des représentations internes très différentes.

Ces standards sont fournis gratuitement à la communauté internationale sous la forme de **spécifications** ; le consortium OpenGIS (abrégé OGC) ne fabrique pas de logiciel. C'est le respect volontaire de ces spécifications qui permet à des groupes de recherche à priori indépendants d'échanger plus facilement des informations géographiques. Pour obtenir une **implémentation** de ces spécifications, nous devons choisir un des produits conformes disponibles sur le marché ou développer nous-même. Les objectifs affichés de l'OGC sont les suivants :

- la promotion de l'utilisation d'applications ouvertes pour le traitement de l'information géographique sur le marché des technologies de l'information ;
- la synchronisation des technologies de l'information géographique avec les standards des technologies de l'information (systèmes ouverts, applications distribuées, technologie de l'orienté objet, architectures en composants) ;
- la coopération entre les fournisseurs de systèmes et les utilisateurs pour le développement d'interfaces ouvertes ;

²⁹ API: Application Programming Interface.

- l'implication de l'ensemble de la communauté (industries, gouvernements, milieux académiques, organismes de normalisation) dans le processus d'interopérabilité ;
- la création d'une plate-forme d'échange pour promouvoir les partenariats, initiatives et projets communs de développement dans le domaine des technologies de l'information géographique.

3.1.4 Quelques spécifications

Les standards du consortium OpenGIS sont spécifiés dans une quinzaine de documents. Chaque document élabore un service, par exemple les transformations de coordonnées. Le fonctionnement de chaque service est décrit par un ensemble de classes d'objets et leurs interactions. La figure 12 illustre les deux services les plus fondamentaux utilisés dans le cadre de cette thèse, ainsi que leurs paquets de classes. Ces paquets sont illustrés dans l'ordre inverse de leurs dépendances. Par exemple, le paquet *Grid Coverage* a besoin de *Coverage*, qui a lui-même besoin de *Coordinate System* et ainsi de suite jusqu'à *Positioning*, le seul paquet qui n'a besoin d'aucun autre.

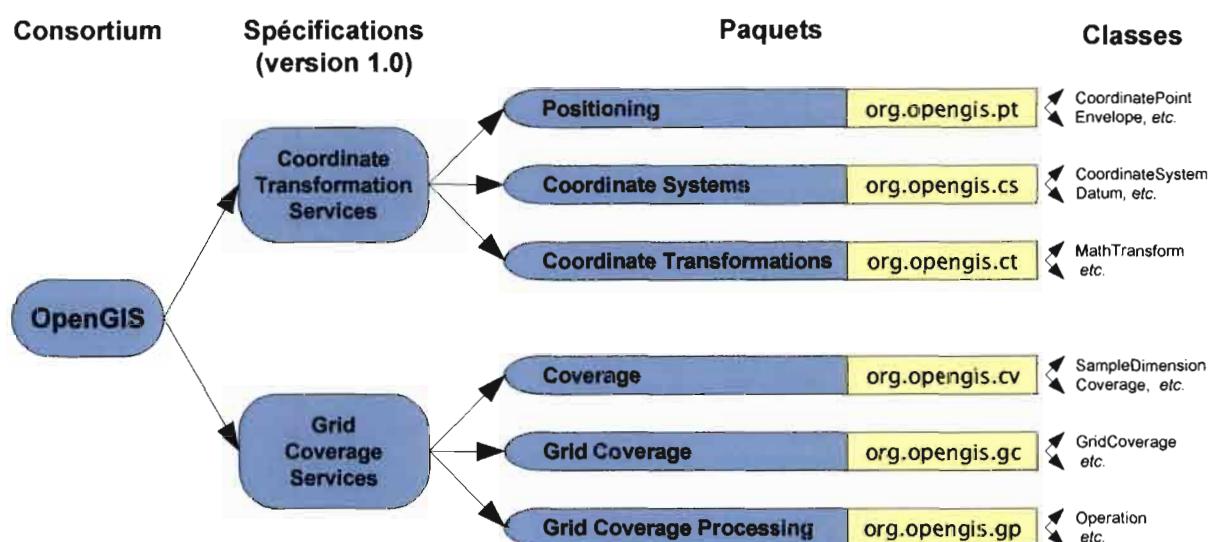


Figure 12: Spécifications d'implémentation d'OpenGIS utilisées dans cette étude (version 1.0)

Les classes développées pour cette thèse sont basées sur la version 1.0 des deux spécifications de la figure 12. Toutefois, au moment de rédiger cette thèse, de nouvelles spécifications d'OpenGIS ont été publiées. Certains éléments existants ont changé (par exemple la classe *CoordinateSystem* a été remplacée par *CoordinateReferenceSystem*)³⁰ et des éléments nouveaux ont été introduits, mais n'ont pas encore été pris en compte dans ce travail, notamment en matière de bases de données d'images. Lorsque les détails d'un changement n'influent pas sur la discussion, les pages suivantes emploieront la terminologie de la spécification la plus récente plutôt que celle qui a été implémentée. Ceci afin d'aider le lecteur à anticiper ce que le standard sera, plutôt que ce qu'il était.

³⁰ Ces changements sont souvent incompatibles avec les versions antérieures. Le consortium OpenGIS est conscient du problème que cela pose. Par exemple le projet INSPIRE (une initiative légale de l'Union Européenne) pourrait proposer l'adoption d'une partie des spécifications d'OpenGIS, sous réserve qu'elles se stabilisent suffisamment pour éviter les bris de compatibilité à l'avenir.

3.1.5 Apport des implémentations libres

Par rapport aux produits commerciaux, les projets libres nous offrent une plus grande souplesse d'adaptation à nos besoins, puisque nous disposons d'un accès direct à l'ensemble du code source. En retour, soumettre nos modifications au projet original permet de les perpétuer dans les versions futures, et donc d'accroître les chances que la solution choisie reste pertinente à notre problématique à long terme. Le projet Geotools 2 est particulièrement actif : il compte une dizaine de développeurs, suscite l'attention du consortium OpenGIS et a une base d'utilisateurs croissante.

Bien que les spécifications d'OpenGIS soient suffisamment complètes pour répondre aux besoins des études océanographiques autant que terrestres, dans la pratique les implémentations tendent à négliger les aspects rarement utilisés sur terre. En contribuant à Geotools 2, nous avons consacré une attention particulière aux problèmes océanographiques qui risquaient d'être de plus basse priorité pour les autres développeurs :

- **Dynamique temporelle :**

Les images terrestres présentent des contrastes forts sur de courtes distances, mais leurs grandes structures restent relativement stables d'une journée à l'autre. La tendance du marché est aux images de résolution très fine (submétriques) parfois espacées de plusieurs années. Le paysage océanographique en revanche est plus nuancé mais se transforme rapidement. Le temps doit être traité comme un continuum, au même titre que l'espace.

- **Image comme source d'informations quantitatives et qualitatives :**

Chaque pixel peut coder une information quantitative (une mesure d'un paramètre continu) ou qualitative (le résultat d'une classification³¹). Le caractère quantitatif ou qualitatif des valeurs détermine les algorithmes applicables. Le logiciel ERDAS Imagine formalise cette distinction en définissant deux types d'images : *continues* (valeurs quantitatives) et *thématiques* (valeurs qualitatives), que l'on associe par exemple aux modèles numériques de terrain et aux cartes d'occupation du sol respectivement. Ce logiciel applique le type à l'image entière ; la plupart des autres logiciels laissent la distinction à la charge de l'utilisateur. En océanographie, il est plus fréquent de combiner les deux types sur une même image, par exemple avec des mesures de température dont certaines valeurs sont masquées pour diverses raisons (nuage, terre, etc.). La prise en compte du caractère quantitatif ou qualitatif des valeurs sur une même image doit être aussi transparente que possible.

- **Encodage non-linéaire des pixels :**

Lorsque le paramètre mesuré est continu, l'encodage des valeurs géophysiques (nombres réels) en valeurs de pixels (entiers dans une plage restreinte) nécessite une transformation des valeurs. Cette transformation est spécifiée par une équation reliant les deux types de valeurs. Cette relation est propre à chaque image et presque toujours linéaire. Des exceptions existent toutefois, notamment les images de concentrations en chlorophylle-*a* qui utilisent une relation logarithmique. Le mécanisme d'encodage et de décodage des valeurs devra être suffisamment générique pour accepter ce genre de cas.

Dans les cas où les spécifications d'OpenGIS ne répondent pas pleinement à nos besoins, il suffit souvent de généraliser un concept déjà présent. Cette approche n'est possible qu'avec les projets libres ; on peut difficilement généraliser une bibliothèque dont le code est fermé (on peut lui ajouter de nouvelles fonctionnalités, mais ce n'est pas toujours satisfaisant)³². Cette approche a été appliquée notamment à l'encodage non-linéaire des pixels (section 3.4.4.2). En posant le problème au sein du

³¹ La classification affecte des valeurs numériques à des classes ou catégories. Par exemple 1=forêt, 2=rivière, 3=zone urbaine, etc. Contrairement aux valeurs continues, comparer les valeurs numériques des classes a peu de sens, hormis pour l'ordre d'apparition dans une liste.

³² Par exemple, une macro complémentaire de *Microsoft Excel 2000* (la transformation de Fourier rapide) stocke des nombres complexes dans des cellules. Les nombres réels ne sont qu'un cas particulier des nombres complexes ; pourtant *Excel* sait gérer les premiers mais pas les seconds. La macro complémentaire contourne le problème en codant les nombres complexes comme des chaînes de caractères de la forme "x + yi". Mais on

projet Geotools 2, les chances qu'il soit entendu par le consortium OpenGIS deviennent plus élevées que si cette réflexion demeurait une initiative isolée.

3.1.6 Les projets Seagis, Geotools et GeoAPI

Les solutions développées dans le cadre de cette thèse ne présentent pas toutes le même intérêt général. Certaines sont très spécifiques à notre problématique, par exemple les méthodes d'accès à nos bases de données (section 3.5.4). D'autres s'appliquent à la plupart des problèmes de géomatique, par exemple les systèmes de coordonnées (chapitre 3.3). Aussi avons nous effectué un développement par paliers, répartis dans trois projets, avec occasionnellement des transferts de code entre ces projets :

- **Seagis** (<http://seagis.sourceforge.net>)

Ce projet est propre à cette thèse et ne comportait à l'origine qu'un seul développeur (l'auteur de ces lignes). Il a été lancé le 20 juillet 2001. Au début, tous nos développements étaient effectués au sein de ce projet. Avec le temps, il a pris l'orientation d'un projet expérimental hébergeant les développements trop spécifiques pour les autres projets (Geotools et GeoAPI).

- **Geotools** (<http://www.geotools.org>)

Ce projet a été lancé par le *Centre for Computational Geography* de l'université Leeds (Angleterre) vers l'année 1997. Il a pris un nouveau départ au début de 2002 avec Geotools 2, une refonte pour mieux embrasser les nouvelles normes du consortium OpenGIS. Cette nouvelle version a suscité l'adhésion d'un nombre croissant de développeurs et sert de fondation à d'autres projets tel que GeoServer. Au moment d'écrire ces lignes, il compte une dizaine de développeurs actifs responsables du projet (« *Project Management Committee* », abrégé PMC) et un nombre difficile à évaluer d'utilisateurs. L'auteur de ces lignes s'est joint au groupe PMC vers juin 2002, en transférant le code du projet Seagis relatif aux systèmes de coordonnées (chapitre 3.3), à la gestion des images (chapitre 3.4) et à l'affichage de cartes (chapitre 3.2). Aujourd'hui, nous effectuons directement au sein de Geotools 2 ou de GeoAPI (voir point suivant) les développements d'intérêt général. Le reste est effectué d'abord au sein de Seagis, et migre vers Geotools 2 plus tard si jugé suffisamment mature.

- **GeoAPI** (<http://geoapi.sourceforge.net>)

Ce projet a été lancé le 23 décembre 2002 par les développeurs de Geotools 2, mais est devenu depuis un projet indépendant rassemblant un auditoire plus large. Conscient que Geotools 2 ne peut pas répondre à tous les besoins (par exemple le marché des périphériques mobiles³³) mais qu'un certain degré d'interopérabilité au niveau des interfaces de programmation serait tout de même bénéfique, nous souhaitions encourager les développeurs des autres projets à implémenter les mêmes interfaces que Geotools 2. Or, ces interfaces auraient beaucoup plus de chances d'être acceptées par des développeurs externes (y compris des commerciaux) si elles apparaissent dans un projet indépendant et collent au maximum aux spécifications du consortium OpenGIS. Le projet GeoAPI est né de ce calcul politique. La contribution de l'auteur de ces lignes à ce projet concerne les modules *Meta Data*, *Spatial Referencing by Coordinates*, *Feature Geometry* et *Grid Coverage*. Bien que l'expérience acquise au sein de Geotools 2 ait influencé la conception des interfaces de GeoAPI, elle n'en est pas la source unique. La source principale reste les spécifications du consortium OpenGIS, et ce dernier aura vraisemblablement la charge d'approuver ou de réclamer des modifications aux interfaces de GeoAPI.

ne peut les utiliser dans des formules conventionnelles, ni appliquer les formats de nombres, ni les tracer dans un graphique. Un analyseur doit convertir les chaînes de caractères en leurs paires de nombres réels et imaginaires avant chaque usage, ce qui ralentit les performances. Quelle que soit l'ingéniosité des macros complémentaires, elles ne peuvent apporter un support efficace des nombres complexes à un logiciel dont les fondations n'ont pas été conçues pour gérer plus que la partie réelle. Une telle généralisation aurait des répercussions dans à peu près tous les recoins du logiciel (calculs, formats, graphiques, etc.).

³³ Les périphériques mobiles sont équipés d'une version particulière du Java baptisée J2ME (*Java 2 Micro Edition*). En première approximation, on peut considérer J2ME comme une version allégée de la plateforme qui vise les ordinateurs de bureau, J2SE (*Java 2 Standard Edition*).

3.1.6.1 Définition des besoins

Le projet Seagis (incluant ses contributions aux projets Geotools et GeoAPI) vise à répondre aux besoins suivants :

- Accéder aux données satellitaires en faisant abstraction des pixels, c'est-à-dire ne travailler qu'en termes de coordonnées géographiques et de valeurs géophysiques (température en °C, etc.) plutôt qu'en coordonnées et valeurs de pixels. Les données apparaissent comme une fonction continue, et les valeurs sont interpolées au besoin. Aux fins de calcul, la notion même de résolution des images est estompée.
- Estimer l'erreur sur les données fournies ci-haut mentionnées³⁴. Cette erreur dépend de la précision des instruments de mesure, ainsi que des interpolations effectuées pour estomper le caractère discret de l'échantillonnage (par exemple une interpolation bilinéaire entre quatre pixels).
- Faciliter l'exploration des données selon une troisième dimension (en l'occurrence le temps, mais le problème aurait été le même avec la profondeur). Il s'agit par exemple d'afficher simultanément plusieurs images d'un même paramètre à différentes dates, et de synchroniser les zooms et défilements dans toutes ces images même si elles n'ont pas la même résolution.
- Rendre les services disponibles via une interface de programmation aussi indépendante que possible du format des données et de l'usage que l'on en fait (calcul ou affichage). Cela signifie par exemple de réduire au minimum le besoin de faire des traitements particuliers pour les valeurs manquantes (section 3.4.3), car ces dernières varient d'une source de données à l'autre. L'objectif est de faciliter l'automatisation de tâches et anticiper les besoins de travaux futurs.

³⁴ Ce besoin n'a pas été satisfait dans le cadre de cette thèse, mais sera pris en compte pour des perspectives futures (notamment les réseaux bayésiens, section 5.2.2). Toutefois, l'alignement de notre travail sur des spécifications internationales nous permet d'anticiper ce besoin. Par exemple la spécification ISO 19115 sur les méta-données contient un paquet « *Data Quality* » que nous pourrons exploiter.

3.2 Architecture de l'afficheur d'informations géographiques

Pour faciliter l'exploration des données décrites dans le chapitre 3.5, une application a été développée dans le cadre de cette thèse. Dans une telle application, le seul module affichant les informations géographiques (indépendamment de tous menus ou boîtes de dialogue) est un vaste sujet en soi. Contrairement à d'autres domaines tels que les systèmes de référence des coordonnées (chapitre 3.3) et les images (chapitre 3.4), les afficheurs n'étaient régis par aucun standard de notre connaissance au moment de débuter la rédaction de cette thèse. Une initiative en ce sens, *Geospatial objects* (GO-1), a déposé une proposition de standardisation devant le consortium OpenGIS en juin 2004. L'implémentation de ce standard a été remise à l'après-thèse.

La mise au point de l'afficheur discuté dans ce chapitre a démarré en 1997 au sein de Pêches et Océan Canada avant d'être transféré au projet Geotools 2 en 2002. Nous le qualifions de « *statefull renderer* » du fait qu'il retient sous une forme pré-traitée certaines informations relatives aux éléments à afficher, un peu comme un compilateur convertit un code source en un code binaire plus efficace. Cette approche est à l'opposée de celle d'un second afficheur développé dans Geotools 2, qualifié de « *stateless renderer* » du fait qu'il affiche immédiatement les informations qu'il reçoit sans stocker des objets intermédiaires, un peu comme un interpréteur.

Le principal avantage des afficheurs de type « *stateless* » est leur capacité à afficher de grandes quantités d'informations avec une consommation minime de mémoire. Les informations proviennent typiquement d'une base de données et transitent vers le périphérique de sortie (une imprimante, une fenêtre à l'écran ou une image créée par un serveur) en un flux continu, sans que les informations ne restent en mémoire après le rendu de chaque élément. Ce type d'afficheur a aussi l'avantage d'être plus simple à concevoir, ce qui le rend plus facile à maintenir. Toutefois, si la même région géographique doit être rendu plusieurs fois avec des modifications minimales, un afficheur de type « *stateless* » recommence le processus pour tous les éléments plutôt que seulement pour ceux dont l'état a changé.

Les afficheurs de type « *statefull* » ont les avantages et inconvénients inverses. L'utilisation d'une cache permet de répondre rapidement à certains changements, comme les zooms par un facteur modéré ou des modifications dans quelques éléments. Mais la quantité d'informations conservées en mémoire peut devenir considérable, et la complexité de ces afficheurs les rend rébarbatifs aux yeux d'éventuels nouveaux contributeurs.

Dans un rapport remis en mai 2004, la compagnie *Refractions Research* a comparé trois afficheurs : les versions « *stateless* » et « *statefull* » de Geotools 2, ainsi que JUMP (un projet indépendant). De ces trois afficheurs, la version « *statefull* » de Geotools 2 (mentionné dans le second paragraphe) apparaît comme le plus rapide après démarrage, à la condition que la totalité des données tiennent en mémoire. En outre, au moment de rédiger le rapport, il était le seul à être capable de projeter à la volée les données vers un système de référence de coordonnées arbitraire et à offrir un affichage progressif (par tuile) des images volumineuses. Il offre aussi la capacité de déterminer quelle géométrie se trouve à une position arbitraire (typiquement le curseur de la souris), ce qui est plus difficile à faire avec un afficheur de type « *stateless* ». Cela n'en fait pas nécessairement l'afficheur le plus recommandé pour tous les projets, les serveurs préférant souvent les afficheurs de type « *stateless* » pour les raisons de consommation de mémoire invoquées plus haut. On lui reproche aussi un temps de démarrage beaucoup trop long et une trop grande complexité. Le reste de ce chapitre décrit l'architecture actuelle de l'afficheur « *statefull* » de Geotools 2, et propose des moyens de corriger quelques unes de ses faiblesses. Cette description part de la structure la plus simple – un tableau de coordonnées – et assemble progressivement des structures plus complexes.

3.2.1 Abstraction d'un tableau de coordonnées (PointArray)

La norme ISO 19107 spécifie une interface de programmation décrivant les géométries à deux et à trois dimensions. De ces géométries, seules les bidimensionnelles sont gérées par notre afficheur. Ces géométries à deux dimensions se décomposent elles-mêmes en une séquence de segments de droites (*LineString*) ou une séquence d'arcs de cercles (*ArcString*). Seules les séquences de segments de droites sont gérées par notre afficheur³⁵. Ces dernières sont représentées par une liste de points de contrôles (*PointArray*), entre lesquels une infinité de points peuvent être interpolés linéairement.

³⁵ Le Java standard (J2SE 1.4) ne gère pas directement les arcs de cercles, car cette fonctionnalité n'est pas présente sur toutes les cartes graphiques. Il gère toutefois les courbes quadratiques (QuadCurve2D), plus répandues et qui font d'excellentes approximations des arcs de cercles.

On peut se représenter **PointArray** comme un simple tableau de **Positions** stockées directement en mémoire. Bien qu'inefficace en Java³⁶, ce mode de stockage est utilisé par certaines bibliothèques à cause de sa simplicité. Un mode de stockage plus efficace est d'utiliser un tableau de type `float[]` dans lequel on alterne les coordonnées x et y . Certaines applications exigeront plutôt des nombres réels à double précision, c'est-à-dire un tableau de type `double[]`. On pourrait aussi manipuler un tableau de type `byte[]` représentant des coordonnées compressées pour une consommation réduite de la mémoire. Enfin, la manipulation de bases de données plus volumineuses exigent que les coordonnées puissent résider sur le disque et leur chargement en mémoire limité à la région d'intérêt.

L'interface **PointArray** fonctionne selon le principe des containers; elle permet de faire abstraction du mode de stockage des coordonnées. L'afficheur utilise une implémentation de **PointArray** similaire à l'interface de la norme ISO 19107, mais adaptée à ses besoins spécifiques. Un itérateur permet de parcourir les coordonnées d'un objet **PointArray** vers l'avant seulement (« *forward-only iterator* »); les accès aléatoires y sont interdits. Tous les algorithmes de l'afficheur doivent obligatoirement fonctionner malgré cette restriction. On peut ainsi écrire des implémentations de **PointArray** lisant directement les données compressées d'un fichier ZIP par exemple.

3.2.1.1 Exemple de compression des coordonnées en mémoire

Lorsque le système de coordonnées est cartésien, que les points sont espacés d'une distance à peu près constante et que l'on n'a pas besoin d'une précision supérieure à environ 1% de la distance entre deux points, il est possible de représenter les coordonnées d'une manière plus compacte qu'un tableau de positions directes. Dans la figure 13 à gauche, les points de contrôle (en vert) d'un polygone sont localisés par les positions directes (x_i, y_i) . Dans la figure de droite, seul le premier point est localisé par une position directe (x_0, y_0) ; tous les autres points sont localisés par un déplacement ($\Delta x_i, \Delta y_i$) relatif au point précédent.

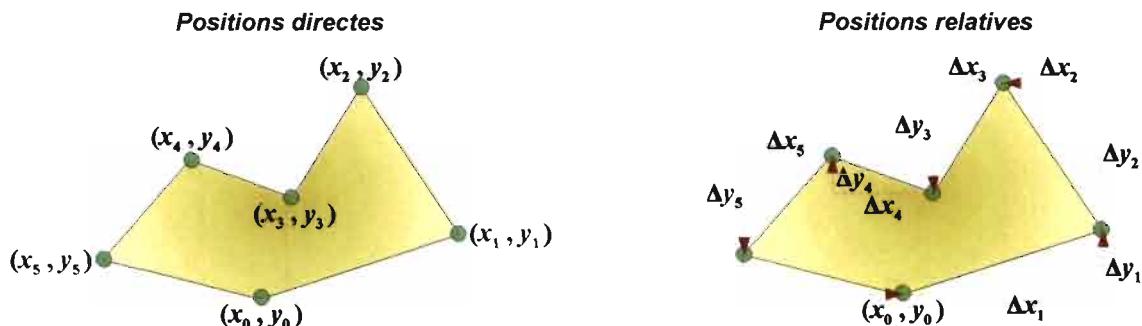


Figure 13: Positions directes et relatives des points de contrôles d'un polygone

Les valeurs des coordonnées x_i et y_i varient dans l'intervalle $[x_{\min} \dots x_{\max}]$ et $[y_{\min} \dots y_{\max}]$ respectivement, tandis que les valeurs Δx_i et Δy_i ne dépassent pas l'intervalle $[-\Delta_{\max} \dots \Delta_{\max}]$ où Δ_{\max} est la valeur maximale de $|x_{i+1} - x_i|$ ou $|y_{i+1} - y_i|$. Si l'erreur sur une valeur x_i ou y_i ne doit pas dépasser ε , alors le type de données utilisé pour le stockage doit être capable de représenter n valeurs différentes :

$$n = \frac{R}{2\varepsilon} + 1 \quad \text{où } R \text{ est la largeur de la plage de valeurs, soit } (x_{\max} - x_{\min}) \text{ ou } (y_{\max} - y_{\min}) \text{ pour la figure de gauche, et } 2\Delta_{\max} \text{ pour la figure de droite.}$$

Dans le cas des polygones ayant un contour très détaillé (c'est le cas de la plupart des structures naturelles, par exemple le contour des côtes), Δ_{\max} dépend de la résolution de la carte et est généralement beaucoup plus petit que $(x_{\max} - x_{\min})$ ou $(y_{\max} - y_{\min})$. Le type de données utilisé pour le stockage doit pouvoir distinguer un nombre de valeurs n beaucoup plus petit, et nécessite donc moins de bits. Si les valeurs Δx_i et Δy_i sont stockées sur 8 bits chacune, l'erreur ε sur les positions de x_i ou y_i ne dépassera pas $\Delta_{\max}/255$.

Une des implémentations de **PointArray** stocke les coordonnées dans un tableau de type `byte[]`. Une relation linéaire est utilisée pour dériver les valeurs de Δx_i et Δy_i à partir des entiers signés sur 8 bits. Pour les tableaux volumineux (c'est-à-dire lorsque devient négligeable l'espace occupé par des informations

³⁶ Les géométries peuvent contenir des millions de points. Stocker des millions d'objets **Position** en Java impose une très forte pression sur le ramasse-miette, tout en ajoutant 2 ou 3 octets par point avec les implémentations les plus répandues des machines virtuelles.

supplémentaires telles que les coefficients de l'équation linéaire), l'espace occupé par ce mode de stockage se compare à d'autres modes employés dans Geotools 2 comme suit :

- environ 25% de l'espace occupé par des positions directes bidimensionnelles stockées dans un tableau de type `float[]` ;
- moins de 3% de l'espace occupé par les tableaux de type `Coordinate[]`, où chaque objet `Coordinate` est une position directe tridimensionnelle (x, y, z) à double précision.

Ces pourcentages sont théoriques ; ils supposent qu'aucune valeur de Δx_i ou de Δy_i n'est supérieure à 255ϵ . Dans le cas contraire, les valeurs trop grandes doivent être diminuées en ajoutant des points intermédiaires entre les points trop espacés. L'ajout de points augmente les pourcentages ci-dessus, mais le gain reste néanmoins appréciable.

Pour un n constant (déterminé par le mode de stockage), la précision ϵ des coordonnées relatives est indépendante de la taille du polygone, mais dépend de Δ_{\max} . Pour un polygone donné, cela revient à dire que la précision augmente avec le nombre de points. Cette propriété contraste avec celle des coordonnées directes, pour lesquelles la précision ϵ est indépendante du nombre de points mais dépend de la largeur ($x_{\max} - x_{\min}$) et hauteur ($y_{\max} - y_{\min}$) du polygone. Pour un polygone très détaillé contenant plusieurs millions de points, la précision des coordonnées relatives dans un tableau de type `byte[]` peut finir par surpasser la précision des coordonnées directes dans un tableau de type `float[]`.

Les coordonnées relatives ont pour principal inconvénient de ne pas permettre les accès aléatoires. Cette contrainte est une des raisons pour lesquelles l'afficheur est conçu pour n'accéder aux coordonnées qu'à l'aide d'un itérateur.

3.2.2 Encapsulation d'un tableau de coordonnées (LineString)

Le standard ISO 19107 définit l'interface `LineString` comme une séquence de segments de droites. Cette séquence peut être représentée par un objet `PointArray` contenant les points de contrôles (c'est-à-dire les points aux extrémités des segments de droites). Pour notre afficheur, nous avons étendu le concept en associant deux objets `PointArray` pour chaque `LineString` plutôt qu'un seul :

- La première séquence peut être vue comme un échantillonnage de points le long d'un contour complexe, par exemple la côte de Terre-Neuve en bleu sur la figure de droite. Ces points sont habituellement rapprochés et nombreux, ce qui fait du `PointArray` un bon candidat au mode de stockage décrit dans la section 3.2.1.1.
- Les points de la seconde séquence sont peu nombreux et espacés. Les segments de droites sont souvent le résultat d'une définition plutôt que d'observations, par exemple une frontière politique, ou simplement les limites de la carte comme la ligne rouge dans la figure de droite.



Outre le fait que ces deux séquences de points de contrôle sont souvent stockées de manières différentes (comme positions relatives pour la première, et comme positions directes pour la seconde), elles sont aussi traitées différemment en cas de reprojection. Une ligne paraissant droite dans une projection cartographique apparaît souvent courbe dans une autre projection. Les éventuelles courbures ne sont prises en compte que pour la seconde séquence de points. Pour la première, les distances entre les points sont supposées trop courtes pour que les courbures soient visibles, et le nombre élevé de points dégraderait trop les performances.

3.2.3 Liste chaînée de séquences de segments de droites (Polyline / Polygon)

Un polygone (ou une polyligne si la forme géométrique n'est pas fermée) est structuré comme une liste d'objets `LineString`. Chaque objet `LineString` pris individuellement ne contient qu'une seule séquence de points « du trait de côte » et une seule séquence de points « du bord de la carte » (figure de la section 3.2.2). La constitution d'une liste chaînée permet de faire alterner ces deux types de séquence autant de fois que l'on veut dans un seul polygone, comme dans la figure ci-contre : l'intérieur de



la baie de Fundy, dans le bas de la carte, est inséré entre deux traits qui délimitent les bords de la carte.

Au moins un nouvel objet `LineString` est nécessaire chaque fois que l'on passe d'une « bordure de carte » à un « trait de côté », mais rien n'empêche de découper un polygone avec davantage de ces objets. Ce découpage des polygones en un nombre arbitraire de séquences de points est totalement invisible aux yeux de l'utilisateur, mais offre plusieurs avantages :

- Permet de diviser un polygone en tuiles de tailles et de formes arbitraires. On peut par exemple appliquer un quadrillage sur un polygone, et stocker les segments de droites apparaissant dans chaque cellule du quadrillage à l'intérieur d'un objet `LineString` différent. Si ces derniers utilisent une implémentation de `PointArray` qui n'extrait les données d'une mémoire auxiliaire que la première fois où elles sont requises, on réduit l'encombrement de la mémoire de l'unité centrale lorsque seule une partie du polygone est affichée³⁷.
- Accélère les tests d'inclusion de points à l'intérieur du polygone (opération `contains(Point2D)`), ainsi que d'intersection avec d'autres polygones. Les segments de droites du polygone ne sont parcourus que si le point à tester se trouve à l'intérieur du plus petit rectangle englobant complètement le polygone. Le découpage d'un polygone en une chaîne de plusieurs `LineString` permet de reproduire ce test à des échelles plus fines, en associant un rectangle par `LineString`.
- Permet de transformer les longues lignes droites en courbes lors des reprojections, sans alourdir les reprojections des courts segments de droites dans les parties détaillées du polygone.
- Augmente la précision de la généralisation des polygones. La généralisation est utilisée pour le passage à une échelle cartographique inférieure. Elle consiste à simplifier des formes en supprimant un certain nombre de points. Cet aspect sera discuté dans la section 3.2.3.2.

Chaque polygone connaît le système de référence des coordonnées (CRS) selon lequel sont exprimés tous les segments de droites qu'il contient. Le polygone est l'objet de plus bas niveau dans lequel ce concept apparaît ; les objets `LineString` qu'il contient n'ont aucune connaissance du CRS³⁸. Nous qualifions ce CRS de « interne », car c'est celui des coordonnées stockées.

En outre, chaque polygone peut offrir une vue de sa géométrie selon un CRS différent. Aux yeux de l'utilisateur, toutes les coordonnées du polygone semblent être exprimées selon le CRS de la vue. Mais en interne, les coordonnées restent toujours selon leur CRS d'origine et les transformations ne sont effectuées qu'à la volée³⁹. En évitant d'écraser les données originales avec les données transformées, nous évitons les pertes de précision (voir les pertes complètes de données, surtout près des pôles) que peuvent entraîner des successions de reprojections. On évite aussi de dupliquer les données si plusieurs vues dans différents CRS sont désirées pour la même géométrie, par exemple en fonction de l'échelle. Enfin nous supposons que le CRS original avait été choisi par le fournisseur de données parce que bien adapté à la géométrie, et qu'il est plus précis pour les calculs.

3.2.3.1 Découpage des polygones avant traçage

Chaque polygone peut être découpé dans les limites d'une région rectangulaire. Ce découpage est effectué à la volée avant un traçage et englobe une région plus grande que la seule région à tracer à l'écran. La totalité de la région pré-découpée (rectangle jaune dans la figure ci-contre) est transmise à la bibliothèque graphique Java2D, qui fera le découpage final en fonction de la fenêtre d'affichage à l'écran (rectangle bleu). La région pré-découpée doit être suffisamment petite pour minimiser la quantité de données que devra



³⁷ Au moment de rédiger ce rapport, une telle implémentation de `PointArray` est la seule pièce manquante pour que le chargement en différé des tuiles d'un polygone soit effectif dans Geotools 2.

³⁸ Dans le standard ISO 19107, seules les positions directes et les géométries sont explicitement associées à un système de référence des coordonnées. Toujours selon ce standard, ni `LineString` et à plus forte raison ni `PointArray` ne sont des géométries. Ce ne sont que des structures internes aux géométries.

³⁹ Sur le plan de l'implémentation, il suffit d'une seule association vers un objet `CoordinateOperation` (section 3.3.1) pour gérer à la fois le CRS interne (`sourceCRS`) et celui de la vue (`targetCRS`). Puisqu'en général tous les polygones d'une carte sont associés au même objet `CoordinateOperation`, la consommation de mémoire se limite à 4 octets par polygone.

traiter Java2D lors de chaque affichage, et suffisamment grande pour maximiser les possibilités de déplacements de la fenêtre sans invalider le pré-découpage. En effet, le pré-découpage est une opération relativement coûteuse qui provoque des pauses perceptibles lors des déplacements de la fenêtre. Un nouveau pré-découpage ne sera effectué que si l'une des conditions suivantes est rencontrée :

- Au moins un bord de la fenêtre sort des limites du précédent pré-découpage, à la suite d'un déplacement horizontal ou d'un zoom arrière.
- Un zoom avant a réduit la superficie de la fenêtre à moins de 6,25% de celle de la région pré-découpée. Ce seuil correspond à une largeur et hauteur égales au quart de celles de la région pré-découpée. Les valeurs de ces seuils ont été déterminées empiriquement.

Les pré-découpages sont conservés dans une cache (section 3.2.3.2) et réutilisés pour chaque traçage. Les tableaux de points (section 3.2.1) ne sont jamais copiés ; seuls les objets qui les enveloppent sont copiés et modifiés pour référencer les parties des tableaux incluses dans le découpage. On peut donc stocker un grand nombre de découpages avec une faible consommation de mémoire. L'afficheur en profite pour conserver un historique des pré-découpages effectués suite aux déplacements de la fenêtre. Lorsque le découpage actuel n'est plus valide, il vérifie d'abord si un découpage plus ancien pourrait être réutilisé.

Ce système de pré-découpages a décuplé la vitesse d'affichage d'une carte de la Méditerranée en présence d'un fort zoom. Un affichage prenant une demi-seconde devient instantané, ce qui rend fluides des défilements qui, sans le pré-découpage, étaient très saccadés.

3.2.3.2 Extraction de la séquence de points à tracer

Au moment du traçage, les objets de type `PointArray` (section 3.2.1) sont convertis en tableaux de type `float[]` généralisés (c'est-à-dire en appliquant une décimation des points) et transformés vers le système de référence des coordonnées (CRS) de l'afficheur. Le tableau 4 ci-dessous énumère les systèmes de références impliqués aux différentes étapes de la transformation, du CRS interne des géométries jusqu'au CRS du périphérique de sortie. Les termes en bleu sont dérivés des définitions de l'initiative *Geospatial objects* d'OpenGIS, tandis que les termes en jaune désignent des étapes propres à l'implémentation de Geotools 2.

Tableau 4: Systèmes de référence des coordonnées à différentes étapes de l'affichage

 Transformations	<p>CRS interne « <i>internal CRS</i> »</p> <p>CRS selon lequel sont exprimées les données originales (les objets <code>PointArray</code>). Il s'agit le plus souvent d'un CRS projeté, mais pourrait aussi être géographique.</p> <p>CRS objectif « <i>objective CRS</i> »</p> <p>CRS associé aux données modélisées par l'afficheur. Il s'agit le plus souvent du même que le CRS interne, mais peut être différent si l'utilisateur effectue une projection à la volée, par exemple en fonction de l'échelle.</p> <p>CRS de l'afficheur « <i>display CRS</i> »</p> <p>CRS associé à la géométrie de l'afficheur. Il s'agit généralement du même que le CRS du périphérique (définition suivante), excepté que les coordonnées sont exprimées en points plutôt qu'en pixels. La taille d'un point est égale à 1/72^e de pouce, tandis que celle d'un pixel dépend du périphérique de sortie.</p> <p>CRS du périphérique « <i>device CRS</i> »</p> <p>CRS associé à la géométrie du périphérique de sortie. Ces coordonnées sont généralement exprimées en pixels. Pour une sortie à l'écran, il est associé à un système de coordonnées cartésien. Pour une sortie sur le dôme d'un planétarium, il est associé à un système de coordonnées sphérique.</p>
---	---

Seuls les `PointArrays` appartenant aux polygones découpés dans la section 3.2.3.1 seront traités. Les tableaux `float[]` produits sont conservés dans une cache et réutilisés lors des traçages subséquents. Les étapes de leur création sont :

1. Convertir la résolution désirée de l'afficheur (par défaut ½ point) en résolution selon le CRS objectif (typiquement en mètres ou en degrés d'angle). La valeur obtenue dépend de l'échelle de la carte. Elle n'est calculée qu'une fois pour l'ensemble de la région à tracer, parce que la conversion du CRS de l'afficheur vers le CRS objectif est généralement affine.

2. Pour chaque **PointArray**, transformer la résolution obtenue à l'étape 1 en résolution selon le CRS interne. Lorsqu'une projection cartographique est faite à la volée, la résolution obtenue varie en fonction de la coordonnée à laquelle elle a été calculée. Aux fins de l'étape 3, nous calculons la résolution au centre du plus petit rectangle englobant complètement le **PointArray**, et supposons cette résolution constante. Cette approximation est d'autant meilleure que la région couverte par **PointArray** est petite. Cette condition peut être assurée même pour les polygones couvrant de vastes espaces, en les construisant comme une liste chaînée d'objets **LineString** tel que décrit à la page 52 (division en tuiles).
3. A l'aide d'un itérateur, copier les coordonnées de l'objet **PointArray** vers un tableau **float[]** en ne retenant que celles qui sont espacées d'une distance supérieure à la résolution calculée à l'étape 2. Cette généralisation accélère les affichages lors de forts zooms arrière, et complète ainsi le découpage décrit dans la section 3.2.3.1 qui accélère l'affichage lors de forts zooms avants.
4. Projeter (si nécessaire) les coordonnées obtenues du CRS interne vers le CRS objectif.
5. Convertir les coordonnées obtenues du CRS objectif vers le CRS de l'afficheur. La dernière étape (conversion vers le CRS du périphérique) est assurée par Java2D.

En cas de zoom avant, arrière, de translation ou de rotation, l'afficheur vérifie si la résolution utilisée à l'étape 1 lors du dernier traçage reste dans certaines limites (par défaut entre $\frac{1}{4}$ et 2 points). Dans l'affirmative, le *changement* d'échelle ou de position est calculé sous forme de conversion affine et appliqué directement à l'étape 5. Dans le cas contraire, une nouvelle généralisation est nécessaire et la chaîne est ré-exécutée à partir de l'étape 1.

Comme nous l'avons vu dans la section 3.2.1, le mode de stockage des coordonnées initiales (selon le « CRS interne » défini dans le tableau 4) est à la discréétion de l'utilisateur ; il peut utiliser des nombres réels à double précision si nécessaire. Mais à partir du moment où ces coordonnées entrent dans la chaîne de traitement de l'afficheur, tous les résultats intermédiaires sont stockés comme nombres réels à simple précision (type **float[]**). Ce type suffit largement au traçage des coordonnées selon le CRS de l'afficheur, puisque les écrans ont rarement plus de quelques milliers de pixels en largeur ou hauteur. Il peut toutefois être insuffisant pour exprimer des coordonnées selon le CRS objectif. Cette limitation survient lorsque les variations des coordonnées dans la région d'intérêt sont faibles par rapport à la valeur absolue de ces coordonnées. Elle peut être contournée en déplaçant l'origine du CRS objectif au centre de la région à afficher. Si cette translation est effectuée avant que les coordonnées selon le CRS interne ne soient forcées au type **float**, alors cette simple précision peut suffire aux calculs⁴⁰. Cette approche peut sembler compliquée comparativement à une utilisation systématique des nombres réels à double précision, mais elle évite de consommer deux fois plus de mémoire que ce qui est strictement nécessaire.

Au moment de rédiger, les travaux restant à faire comprennent :

- Rendre l'API davantage similaire à la norme ISO 19107. Les sections 3.2.1 à 3.2.3 décrivaient une implémentation dédiée à la problématique de l'affichage, mais qui peut néanmoins implémenter au moins partiellement les interfaces dérivées de la norme ISO. Ce rapprochement rendrait l'API plus familier à des contributeurs extérieurs, ce qui pourrait atténuer une de ses principales critiques : sa complexité.
- Extraire les méthodes rarement utilisées (notamment tout ce qui concerne la compression des coordonnées décrite dans la section 3.2.1.1) et les déménager dans un paquet dédié aux algorithmes. L'objectif est encore de simplifier le code en réduisant la taille du noyau.
- Pour les tests d'inclusion d'un point à l'intérieur d'un polygone, utiliser les coordonnées présentes dans la cache plutôt que la totalité des coordonnées internes. Les tests s'en trouveront accélérés du fait de la

⁴⁰ Les océanographes effectuent une opération similaire avec les mesures de la densité de l'eau de mer. Cette densité varie typiquement entre 1000 et 1040 kg/m³. Lorsque l'on stocke ces valeurs sous forme de nombres réels à simple précision, on ne peut retenir plus de 4 chiffres après la virgule. Par commodité, les océanographes utilisent plutôt la densité *sigma-t*, définie comme la densité moins 1000 kg/m³. Les valeurs de densité *sigma-t* varient typiquement entre 0 et 40 kg/m³. Pour cette gamme de valeurs, les nombres réels à simple précision peuvent retenir environ 6 chiffres après la virgule. Le passage de 4 à 6 chiffres correspond au passage d'une résolution à la limite insuffisante vers une résolution supérieure à la précision de la plupart des capteurs. Pour les calculs (notamment ceux qui exigent des mesures de densités classiques plutôt que *sigma-t*), les conversions vers des nombres réels à double précision peuvent se faire à la volée.

généralisation décrite à la page 54. La précision sera réduite, mais les résultats correspondront autant (sinon mieux) à ce que l'utilisateur voit.

- Pour les tests d'inclusion d'un point à l'intérieur d'un polygone, profiter des découpages en tuiles afin de généraliser de manière plus agressive les tuiles dans lesquelles ne se trouvent pas le point. Il s'agit en fait d'implémenter le deuxième item énuméré à la page 52.
- Modifier le pré-traitement effectué avant la compression décrite dans la section 3.2.1.1. L'implémentation actuelle interpole les points à distance constante, ce qui peut créer des géométries dans lesquelles tous les points sont le résultat d'une interpolation. Elle devrait être remplacée par une implémentation créant de nouveaux points (en interpolant sur un arc de cercle) seulement entre les points trop distancés, et laissant les autres inchangés.
- Fournir une implémentation de `PointArray` qui diffère le chargement des coordonnées jusqu'au moment où elles seront requises (section 3.2.1).

3.3 Systèmes de référence des coordonnées

Le chapitre 3.2 présentait un afficheur capable de projeter à la volée les données vers un système de référence de coordonnées arbitraire. La conversion de coordonnées d'un système de référence vers un autre est un des problèmes centraux pour assurer l'interopérabilité des données. Ce chapitre présente les concepts et les choix d'implantation effectués dans le cadre de cette thèse.

Un système de coordonnées (classe `CoordinateSystem`) est composé d'un ensemble d'axes associés à des unités de mesures (classe `Unit`). Ces unités sont habituellement linéaires (mètres) ou angulaires (degrés de longitude ou de latitude), parfois temporelles, mais elles peuvent aussi être plus indirectes ; on peut par exemple utiliser des unités de pression pour mesurer l'altitude. Chaque système de coordonnées a son propre jeu de relations mathématiques permettant de calculer des éléments géométriques tels que la distance entre deux coordonnées ou leur angle. Par exemple dans un espace bidimensionnel, on définit entre autres le système cartésien (classe `CartesianCS`) et le système polaire (classe `PolarCS`).

Un référentiel (classe `Datum`) spécifie la relation entre un système de coordonnées et la terre. Il permet d'appliquer un concept mathématique abstrait (un système de coordonnées) au problème pratique qu'est la description d'une position près de la surface terrestre. Ces référentiels sont généralement établis par des autorités nationales ou scientifiques et varient en fonction des pays ou de l'historique des données. La combinaison d'un référentiel et d'un système de coordonnées produit un système de référence des coordonnées (classe `CoordinateReferenceSystem`, abrégée CRS). Soulignons que les objets CRS ne contiennent jamais de coordonnées ; ces dernières sont mémorisées séparément, typiquement dans un tableau de nombres réels. Un objet CRS ne fait que décrire la façon d'interpréter les valeurs numériques contenues dans de tels tableaux.

Tous les systèmes de référence des coordonnées sont représentés par une sous-classe de CRS. Chacune de ces sous-classes représente un type de système de référence. Par exemple la sous-classe `TemporalCRS` représente un système avec une seule dimension le long de l'axe du temps, tandis que la sous-classe `GeographicCRS` représente un système ellipsoïdal à deux dimensions (degrés de longitudes et de latitudes). Des systèmes de référence composés (`CompoundCRS`) peuvent être construits par l'assemblage d'au moins deux autres systèmes de référence. Par exemple on peut construire un système à trois dimensions en assemblant les systèmes `GeographicCRS` et `TemporalCRS` mentionnés précédemment. La figure 14 illustre de manière simplifiée la hiérarchie des principales classes (flèches pleines) avec les associations (losanges) telles qu'elles ont été utilisées dans cette thèse (les spécifications standards définissent des associations plus générales, qui ne sont pas illustrées ici).

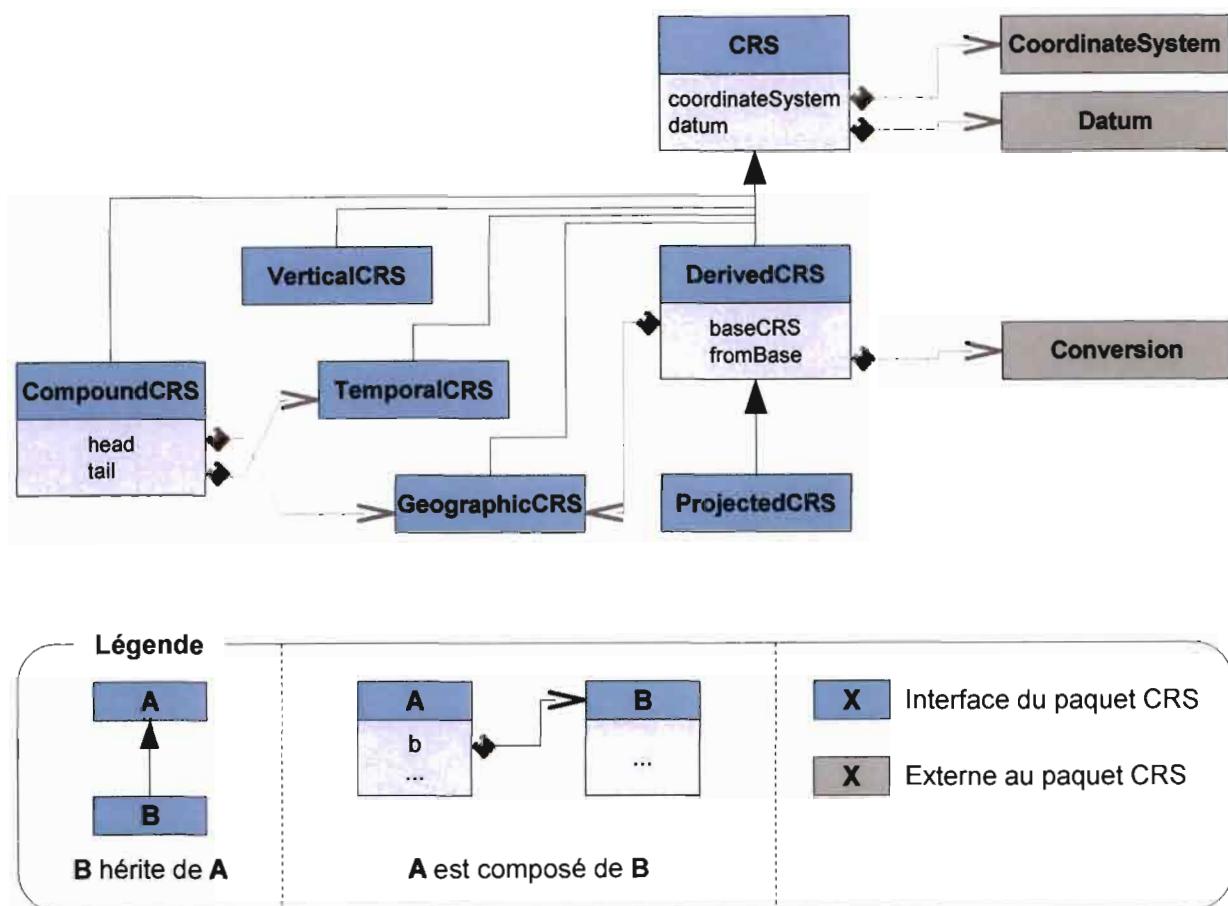


Figure 14: Hiérarchie simplifiée des principales classes de systèmes de référence des coordonnées

Une multitude de combinaisons est possible, du fait qu'il existe plusieurs variantes pour chaque sous-classe de CRS. Par exemple, un système vertical qui mesure les hauteurs avec un baromètre n'est pas identique à un autre système vertical qui mesure les hauteurs avec un GPS⁴¹. Un système temporel utilisant l'heure UTC n'est pas identique à un autre système utilisant l'heure GMT⁴². Enfin, deux systèmes géographiques peuvent différer par l'ellipsoïde de référence qu'ils utilisent. Ces informations sont précisées notamment par le référentiel (Datum) qui compose chaque système CRS (figure 14).

Dans le cadre de la présente étude, nous positionnons les données satellitaires dans un système de référence de coordonnées à trois dimensions, soit dans l'ordre la longitude, la latitude et le temps (x, y, t). Il s'agit donc d'un assemblage d'un objet GeographicCRS (qui référence les coordonnées x, y) avec un objet TemporalCRS (qui référence la coordonnée t), conformément aux associations illustrées dans la figure 14. Les angles sont mesurés en degrés sur l'ellipsoïde WGS 1984⁴³, tandis que le temps est mesuré en jours écoulés depuis le 1^{er} janvier 1950 à 00:00 UTC (convention du centre Aviso pour les données altimétriques). La figure 15 illustre cet assemblage.

⁴¹ **GPS:** *Global Positioning System* – Système de positionnement par satellites.

⁴² L'heure UTC (*Universal Time Clock*) est basée sur une horloge atomique, tandis que l'heure GMT (*Greenwich Mean Time*) est basée sur des observations astronomiques. L'heure GMT est fonction de l'orbite terrestre, tandis que l'heure UTC est régulière. De temps à autre, l'heure UTC est synchronisée sur l'heure GMT par l'introduction d'une seconde supplémentaire à la fin de certaines années.

⁴³ Sauf les données dérivées de l'altimétrie, qui utilisent leur propre ellipsoïde. Toutefois, l'ellipsoïde utilisé pour les données de Topex/Poseidon étant très proche de l'ellipsoïde WGS 1984, la version actuelle de nos programmes ignore la différence.

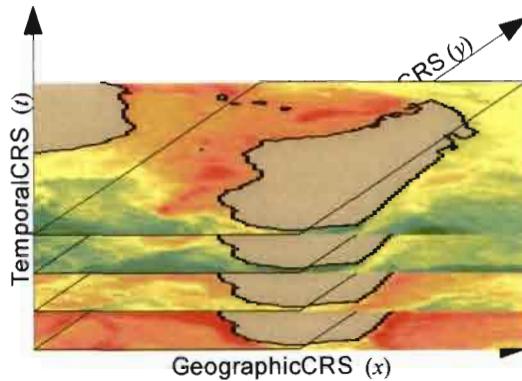


Figure 15: Système de référence des coordonnées 3D (x,y,t) utilisé

Quelques-unes de nos données font exception à ce schéma. Par exemple nous disposons de quelques séries d'images pour lesquelles la composante horizontale GeographicCRS est remplacée par ProjectedCRS avec une projection cylindrique de Mercator. Si des transformations de coordonnées sont nécessaires, elles seront effectuées à la volée. Des transformations s'appliquent aussi à l'axe du temps, puisque les bases de données et le Java ont déjà leur propre type Date dont la définition diffère de celle du centre Aviso (qui fournit les données altimétriques).

3.3.1 Conversions et transformations de coordonnées

Combiner des données provenant de sources hétérogènes exige que l'on soit capable de traduire ces données dans un système de référence commun, choisi par l'utilisateur. Pour cette étude, nous effectuons toutes les combinaisons dans le système tridimensionnel illustré à la figure 15. La plupart de nos données sont déjà référencées selon ce système. Mais pour les quelques séries d'images référencées autrement, des *conversions* ou *transformations* de coordonnées sont nécessaires. Ces deux types d'opérations se distinguent par la nature de leurs paramètres :

- **Conversion** : une opération mathématique qui n'implique pas de changement de référentiel. Les paramètres de la conversion sont définis plutôt que déterminés empiriquement. Les projections cartographiques entrent dans cette catégorie.
- **Transformation** : une opération mathématique qui implique habituellement un changement de référentiel. Les paramètres sont de nature stochastique, déterminés empiriquement à partir d'un ensemble de positions exprimées selon les deux systèmes de référence des coordonnées. À chaque transformation est donc associée une estimation statistique de l'erreur. Les changements d'ellipsoïde de « Clarke 1866 » vers « WGS 1984 » font partie de cette catégorie.

Les conversions ou transformations définissent une relation mathématique d'un objet CRS d'origine (source – variable selon nos données) vers le système désiré (target – figure 15 dans le cas particulier de cette thèse). Dans le cas général, il est pratiquement impossible de stocker les opérations entre toutes les paires possibles de systèmes de référence des coordonnées. Le logiciel doit être capable de dériver automatiquement de nouvelles opérations à partir de concaténations de plusieurs opérations connues, ou de leurs inverses. Par exemple une base de données peut spécifier une transformation T_{AB} du système A vers le système B ainsi qu'une transformation T_{AC} du système A vers le système C , mais pas de transformation de B vers C . En employant une notation matricielle⁴⁴, une telle transformation doit pouvoir se dériver par $T_{BC} = (T_{AC}) \times (T_{AB})^{-1}$. Les changements d'unités, d'ordre et de direction des axes peuvent s'exprimer sous forme de conversions affines et être concaténés de la même manière.

La figure 16 donne un exemple de chaîne des opérations que notre implémentation peut construire. Chaque rectangle représente un système de référence des coordonnées, avec en bleu les systèmes intermédiaires qui utilisent des axes standardisés (unités en degrés ou en mètres, axes dans l'ordre x, y, z, t et directions conventionnelles). Chaque flèche représente une opération. Bien que la figure 16 explicite bon

⁴⁴ Cette notation s'applique particulièrement aux transformations affines. La multiplication matricielle de deux transformations affines T_1 et T_2 équivaut à transformer les coordonnées par T_2 d'abord, puis par T_1 . Autrement dit, les concaténations telles que $T_1 \times T_2$ doivent se lire de droite à gauche.

nombre d'entre elles, dans la pratique plusieurs opérations seront fusionnées en une seule. Par exemple deux transformations affines consécutives seront concaténées à l'aide d'une multiplication matricielle. De plus, certaines flèches se traduisent souvent par l'opération identité. Par exemple toutes les étapes ombragées seront omises si le référentiel source et destination est le même.

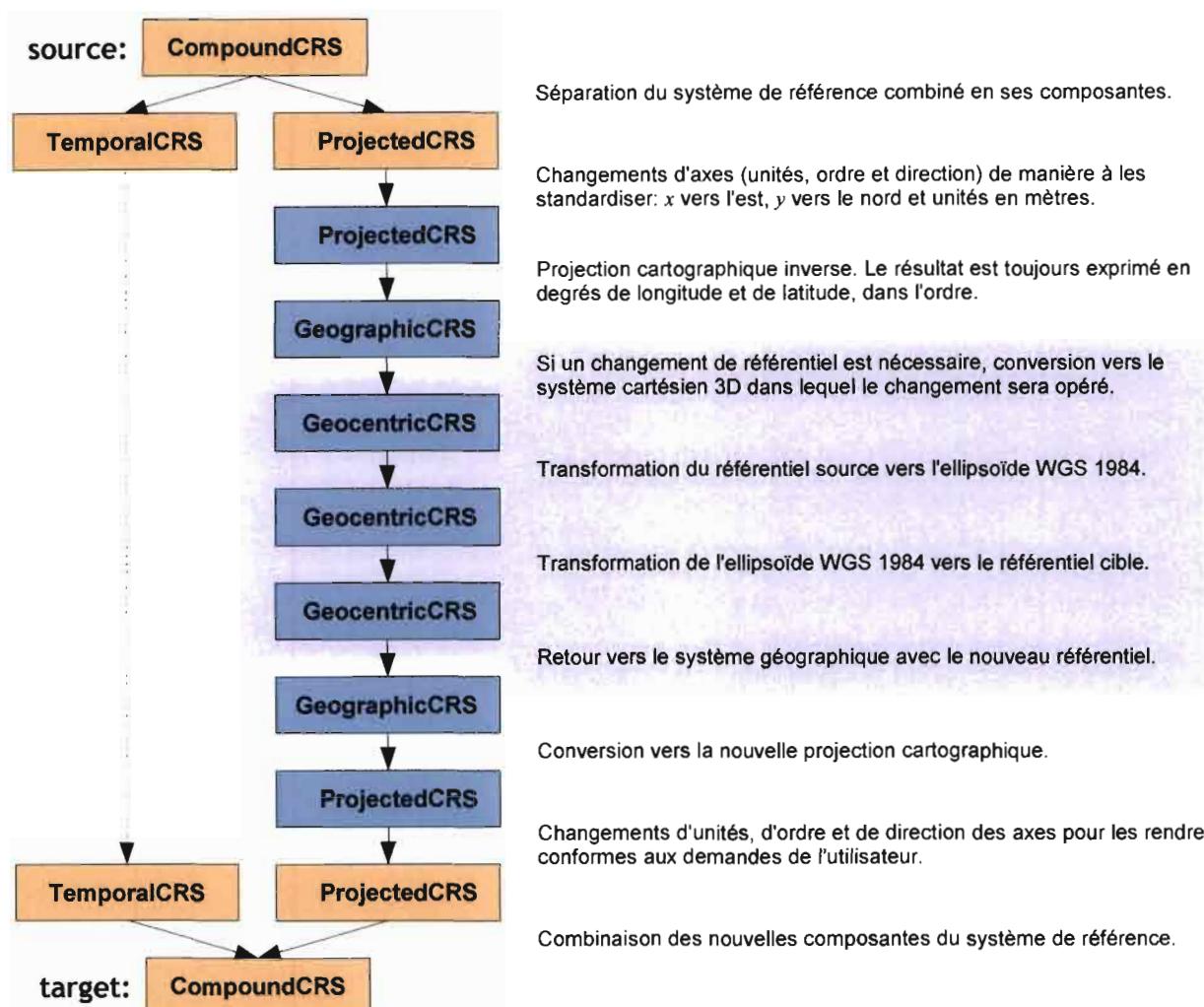


Figure 16: Exemple de chaîne des opérations

Dans la figure 16, les deux composantes du CompoundCRS donnent lieu à des chaînes d'opérations indépendantes. Cela ne sera pas toujours le cas. Par exemple si une composante verticale (VerticalCRS) est présente, elle sera prise en compte lors des conversions entre les coordonnées géographiques et les coordonnées géocentriques tridimensionnelles (partie ombragée de la figure 16). Un nombre restreint de cas peuvent aussi impliquer une chaîne d'opérations qui dépendent du temps (par exemple un système de référence associé à un navire qui se déplace), mais notre implémentation n'a pas encore atteint ce degré de sophistication.

Crédits :

L'implémentation des projections cartographiques est basée sur la bibliothèque PROJ.4, remodelée et portée en Java au sein de Pêches et Océans Canada, puis portée vers l'API d'OpenGIS lors de cette thèse. L'algorithme qui enchaîne les opérations a été développé lors de cette thèse, mais PROJ.4 dispose aussi de son propre algorithme. Une batterie de tests fournie par OpenGIS permet de vérifier l'exactitude des opérations. Je tiens à remercier les participants du projet Geotools 2 pour l'aide qu'ils m'ont apportée dans la réalisation des bibliothèques de CRS à travers leurs tests et remarques.

3.3.2 Crédits :

Les objets CRS et leurs dépendances sont créés par des « fabriques » de bas niveau : <Class>Factory, où <Class> dépend de la classe des objets fabriqués et sera omis pour le reste de la discussion. Différents Factory

peuvent fournir différentes implémentations des objets CRS à partir des mêmes informations géographiques, par exemple pour des utilisations en réseaux suivant différents protocoles (section 3.3.3). L'utilisation de ces Factory de bas niveau est néanmoins laborieuse en raison de la quantité d'informations qu'il faut fournir : parfois une demi-page de code pour un seul système de référence des coordonnées. Il est plus commode de désigner le système désiré par un simple nom ou un code reconnu par un organisme faisant figure d'autorité (habituellement une organisation nationale ou scientifique). La création d'objets à partir de ces noms est alors dévolue à une fabrique de plus haut niveau, «*Class*·AuthorityFactory».

Comme pour les Factory de bas niveau, plusieurs implémentations des AuthorityFactory sont possibles, mais cette fois en fonction des autorités nationales, scientifiques ou autres. Par exemple les paramètres qui définissent les référentiels, leurs relations avec l'ellipsoïde standard WGS 1984⁴⁵, l'ordre, la direction et les unités des axes, les méthodes de projection cartographiques et la façon dont tout ceci s'assemble pour former des systèmes de référence des coordonnées sont répertoriés dans une base de données maintenue par l'EPSG⁴⁶. Cette base de données recense plus de 3000 systèmes en usage sur la planète ainsi que près d'un millier de transformations entre ces systèmes. Une de nos implémentations de AuthorityFactory exploite cette base de données : elle en extrait les informations correspondant à un code de l'EPSG, puis transmet ces informations de plus bas niveau à un Factory qui procédera à la construction du système de référence. Ce flux de données est illustré à la figure 17, d'un code de l'EPSG jusqu'à un objet CRS.

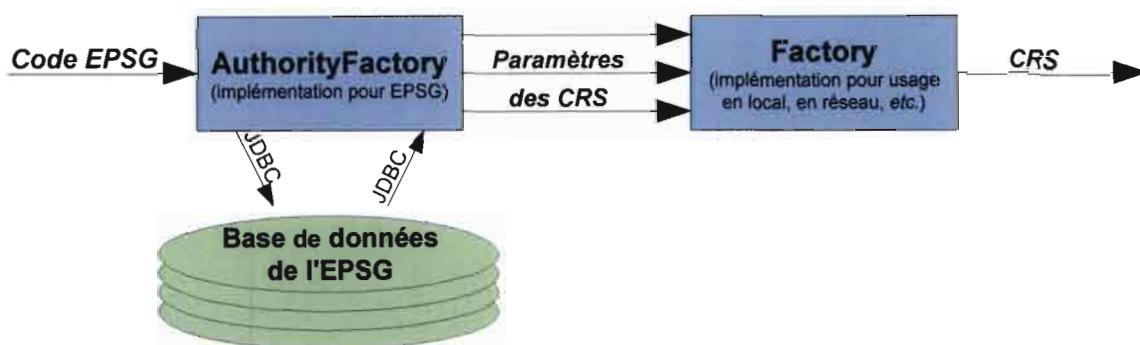


Figure 17: Flux de données à travers les fabriques de systèmes de référence des coordonnées

Comme l'illustre la figure 17, la fabrique de bas niveau Factory intervient en bout de chaîne. Elle intervient aussi dans d'autres processus de fabrication d'objets CRS, par exemple à partir de données au format XML⁴⁷ ou WKT⁴⁸. En ne remplaçant que l'implémentation du Factory, il est possible de s'adapter aux différents types de réseaux. Une approche en ce sens a été brièvement tentée à partir de Rimouski, alors que les 40 Giga-octets de données environnementales se trouvaient à Montpellier. Elle n'a pas été menée à son terme faute de temps, mais semblait techniquement réalisable.

3.3.3 Systèmes et transformations de coordonnées distribuées

Les données sont souvent gérées plus efficacement là où elles sont produites. En conséquence, les études impliquant de nombreux paramètres doivent puiser à plusieurs sources. OpenGIS a conçu ses spécifications dans l'esprit d'un monde où les données et les calculs seraient distribués sur plusieurs machines distantes. En Java, les technologies utilisées à cet effet se traduisent en interfaces RMI⁴⁹. On associe souvent ces interfaces avec le protocole JRMP⁵⁰ propre au Java, ce qui donne l'impression que les RMI ne permettent pas de communiquer avec d'autres langages. En réalité nous pouvons choisir parmi une variété de protocoles

⁴⁵ **WGS 84:** *World Geodetic System 1984*.

⁴⁶ **EPSG:** *European Petroleum Survey Group* (<http://www.epsg.org/>).

⁴⁷ **XML:** *eXtensible Markup Language* – Langage à balises d'apparence similaire au HTML.

⁴⁸ **WKT:** *Well Known Text* – Format standard décrivant les systèmes de référence sous forme de texte. Ce format est utilisé dans un champ texte de la base de données d'images (section 3.5.2).

⁴⁹ **RMI:** *Remote Method Invocation* – Modèle de programmation pour application distribuée en Java.

⁵⁰ **JRMP:** *Java Remote Method Protocol* – Protocole de transfert pour RMI propre au Java.

de transport, incluant RMI-IIOP⁵¹ (basé sur CORBA) et SOAP⁵². Une fois la connexion établie, le protocole utilisé est transparent à l'utilisateur. La substitution d'un **Factory** par un autre dans le paragraphe précédent est un moyen de s'adapter aux différents protocoles.

Permettre les appels à distance impose certaines contraintes sur la signature et la sémantique des méthodes. Par exemple certains types d'arguments sont passés par référence lorsqu'une méthode est appelée en local, alors que les mêmes arguments passés à la même méthode transitent par valeur sur le réseau. Cette complication affecte particulièrement les méthodes agissant sur des tableaux, surtout les conversions ou transformations de coordonnées. En effet, si une méthode de transformation avait la garantie que les tableaux seront passés par référence, alors sa signature serait typiquement celle de la méthode 1 ci-dessous (signature inspirée de la classe **AffineTransform** du Java standard) :

transform — appel sur la machine locale seulement

```
public void transform(double[] srcPts, int srcOff,
                      double[] dstPts, int dstOff, int numPts)
```

Transforms a list of coordinate point ordinal values. This method is provided for efficiently transforming many points. The supplied array will contain packed ordinal values. For example, if the source dimension is 3, then the ordinals will be packed in this order: ($x_0, y_0, z_0, x_1, y_1, z_1, \dots$).

Parameters:
 srcPts - The array containing the source point coordinates.
 srcOff - The offset to the first point to be transformed in the source array.
 dstPts - The array into which the transformed point coordinates are returned.
 May be the same than srcPts.
 srcOff - The offset to the location of the first transformed point that is stored in the destination array.
 numPts - The number of point objects to be transformed.

Throws: TransformException if a point can't be transformed.

Méthode 1: Transformation de coordonnées sur une machine locale (API de Geotools 2)

La signature de la méthode 1 répond efficacement à deux besoins essentiels : la possibilité de transformer une partie seulement du tableau, et la possibilité d'écrire le résultat dans un tableau pré-existant (incluant le tableau des coordonnées sources, qui seront alors écrasées). Ces fonctionnalités sont utilisées intensivement par les afficheurs afin de ne transformer que les parties visibles des géométries, et de placer le résultat dans une cache qui sera mise à jour à chaque changement d'échelle.

Efficace pour un usage local, la signature de la méthode 1 est toutefois inadaptée aux appels sur une machine distante (au moins avec le transport JRMP²⁵ par défaut). D'abord, elle expédie la totalité du tableau source sur le réseau alors que la transformation ne concerne peut-être qu'une petite portion. Pire, elle expédie inutilement le tableau de destination, dont les données sont destinées à être écrasées de toute façon. Enfin, puisque les tableaux qui transitent sur le réseau ne sont que des copies, les écritures faites par le serveur dans le tableau de destination ne seront pas reflétées chez le client. Il en résulte un dysfonctionnement de la méthode 1 dans une architecture client-serveur.

La signature proposée par OpenGIS (méthode 2 ci-dessous) est un compromis qui fonctionnera mieux sur un réseau, tout en limitant les coûts en usage local. Les tableaux ont exactement la longueur nécessaire, et le résultat est retourné par le serveur plutôt qu'écrit dans un tableau fournit par le client. Le compromis réside dans la description de la méthode, qui autorise (mais n'oblige pas) à retourner le tableau source après avoir écrasé ses données. Ainsi, on peut éviter dans certains cas d'allouer de la mémoire pour stocker le résultat, un objectif important de la méthode 1.

⁵¹ **IOP:** *Internet InterORB Protocol*, où ORB est l'acronyme de *Object Request Broker*.

⁵² Les appels à des méthodes distantes à base de messages XML sont le sujet du JSR-224 : *Java API for XML-Based RPC*. Du point de vue de l'utilisateur, ces méthodes ressemblent à celles d'une interface RMI classique.

transform — appel sur une machine locale ou distante

```
public double[] transformList(double[] srcPts)
```

Transforms a list of coordinate point ordinal values. This method is provided for efficiently transforming many points. The supplied array will contain packed ordinal values. For example, if the source dimension is 3, then the ordinals will be packed in this order: ($x_0, y_0, z_0, x_1, y_1, z_1, \dots$). The size of the passed array must be an integer multiple of source dimension. The returned ordinal values are packed in a similar way. In some implementations, the ordinals may be transformed in-place, and the returned array may be the same as the passed array. So any client code should not attempt to reuse the passed ordinal values (although they can certainly reuse the passed array). If there is any problem then the server implementation will throw an exception. If this happens then the client should not make any assumptions about the state of the ordinal values.

Parameters:	srcPts - Packed ordinates of points to transform.
Returns:	The packed transformed points. May be srcPts.
Throws:	RemoteException if a remote method call failed.

Méthode 2: Transformation de coordonnées sur une machine locale ou distante (API d'OpenGIS)

Contrairement à la méthode 1, la méthode 2 force l'utilisation d'un tableau temporaire. D'abord parce que la transformation d'un sous-ensemble des coordonnées nécessite leur copie dans un tableau de longueur appropriée⁵³, ensuite parce qu'une copie défensive est nécessaire de toute façon puisque le tableau source peut être écrasé. De même, le contenu du tableau obtenu en sortie devra souvent être copié à son emplacement définitif. Ces nombreuses copies, ajoutées aux allocations de mémoire que la méthode 2 devra faire si elle n'écrase pas le tableau source (par exemple lorsque les points n'ont pas la même dimension en sortie qu'en entrée), rend difficile une utilisation efficace de cette méthode. Sans parler de l'encombrement des RemoteException qu'il faut gérer...

Dans nos tentatives de concilier un usage local efficace avec un usage en réseau, les conversions et transformations de coordonnées sont le principal écueil rencontré (méthodes 1 et 2)⁵⁴. La spécification *Coordinate Transformation Services 1.0* (d'où est issue la méthode 2) est peut-être plus contraignante que nécessaire du fait qu'elle déclare pratiquement toutes les interfaces comme utilisables à distance⁵⁵, ce qui nous semble exagéré. Par exemple dans le cas de l'interface Ellipsoid, que l'on peut résumer à une structure contenant deux nombres réels (les longueurs des semi-axes majeure et mineure), déléguer les services à un serveur distant ne paraît pas très utile. Cette remarque peut aussi s'appliquer à des opérations plus complexes, comme les transformations de coordonnées discutées dans ce chapitre : il est parfois préférable que le serveur envoie le code permettant au client d'exécuter l'opération en local, surtout s'il a des milliers de points à transformer. Transférer sur le réseau le code ou les données : le choix dépend de la complexité de l'opération et du volume de données. Ces paramètres sont trop variables pour être figés dans la hiérarchie des interfaces.

Notre contribution actuelle à Geotools 2 n'implémente pas directement les interfaces d'OpenGIS, afin de contourner les contraintes propres à une exécution distribuée. Nous avons plutôt implémenté indirectement ces interfaces via un ensemble d'adaptateurs, soit une approche qui s'écarte des idiomes habituels de la programmation orientée objet. Nous avons procédé ainsi parce que nous n'avons pas autorité pour modifier les interfaces d'OpenGIS, alors que nous pouvons modifier l'implémentation de Geotools 2. Toutefois, l'alignement récent des interfaces d'OpenGIS sur les standards ISO 19107 (géométries), ISO 19111 (références spatiales) et ISO 19115 (meta-données) nous offre la possibilité d'intervenir sur cet aspect des prochaines interfaces Java, via notre participation au projet GeoAPI.

⁵³ Dans une fonction en C/C++, on pourrait simplement modifier la valeur du pointeur vers le tableau. Cette approche est possible parce que contrairement au Java, les tableaux en C/C++ n'ont pas d'en-tête spécifiant la classe de l'objet et la longueur du tableau ; la longueur doit être spécifiée explicitement par le programmeur. Le C++ peut aisément contourner le problème en enveloppant le tableau et sa longueur dans une classe. Cette approche facilite la création de sous-tableaux et les ré-allocations de tableaux existants, au prix d'un niveau supplémentaire d'indirection (les accès à une instance de cette classe reviennent à utiliser un pointeur vers un pointeur). Les machines virtuelles Java auraient pu implémenter les tableaux de la même façon, mais ont évité de le faire jusqu'à maintenant pour des raisons de performance liées au niveau supplémentaire d'indirection.

⁵⁴ Avec le protocole JRMP, la signature de la méthode 1 pourrait être retenue même en réseau si l'élément de remplacement (« RMI stub ») est écrit à la main plutôt que confié au compilateur rmic. Cela éviterait tous les problèmes de performance de la méthode 2 en local, mais risque d'accroître les coûts de maintenance.

⁵⁵ En Java, une interface est déclarée exécutable sur une machine distante si elle hérite de l'interface Remote.

3.4 Matrices de données (images)

Toutes les données qui peuvent être représentées sur une grille régulière (par exemple une image) sont gérées par la classe `GridCoverage`. Les données qui ne sont pas nécessairement distribuées sur une grille régulière sont gérées par la classe parente (plus générale) `Coverage`, que l'on pourrait traduire par une « couverture de données ». Les couvertures, régulières (« Grid ») ou pas, possèdent toutes au moins les propriétés suivantes :

- Un système de référence des coordonnées (CRS) qui s'applique à l'enveloppe (voir le point suivant) ainsi qu'à toutes les coordonnées utilisées avec cette couverture. Dans le cadre de la présente étude, ce système de référence est celui de la figure 15 (page 58).
- Une enveloppe rectangulaire (`Envelope`) qui donne les coordonnées extrêmes de la couverture. Cette enveloppe doit avoir le même nombre de dimensions que le système de référence des coordonnées. Dans la présente étude, chaque couverture possède une enveloppe à trois dimensions qui donne dans l'ordre : la longitude minimale et maximale, la latitude minimale et maximale, ainsi que la date de début et de fin de l'image.
- Au moins une bande (décrise par `SampleDimension`), parfois plusieurs. Chaque bande contient une mesure géophysique (par exemple la température) à la position de chaque pixel. Les images de cette étude n'ont qu'une seule bande, sauf les images dérivées des données altimétriques qui en ont trois (l'anomalie de la hauteur de l'eau ainsi que les composantes *U* et *V* du courant géostrophique).

Comme pour les objets `CoordinateReferenceSystem` (chapitre 3.3), les objets `SampleDimension` ne contiennent jamais de mesures ; ils ne font que décrire la façon d'interpréter les valeurs numériques d'une bande. Le mode de stockage des données est à la discréction des implémentations. Dans le cas particulier de Geotools 2, il est délégué à une classe standard du Java : `RenderedImage`. La figure suivante illustre les relations entre les principales classes utilisées par l'implémentation de Geotools 2.

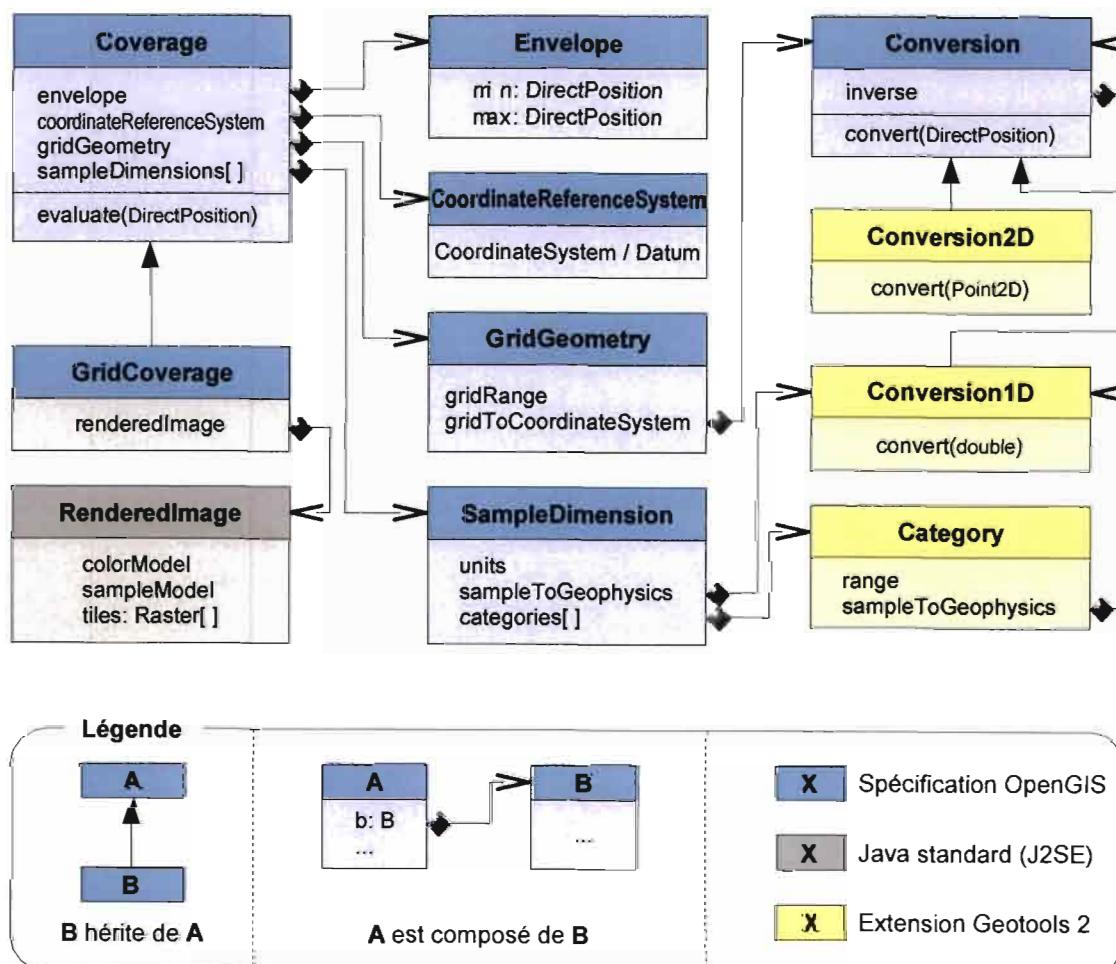


Figure 18: Diagramme simplifié de `GridCoverage` et de ses dépendances dans Geotools 2

La classe Coverage (dont hérite GridCoverage) dispose de méthodes evaluate(...) qui permettent d'obtenir la valeur des données géophysiques à une coordonnée quelconque. A chaque appel d'une méthode evaluate(...), les valeurs sont calculées pour toutes les bandes et retournées dans un tableau. Si l'image n'a qu'une seule bande, la longueur de ce tableau ne sera que de 1. La méthode 3 ci-dessous est extraite de la documentation *javadoc* :

evaluate

```
public float[] evaluate(DirectPosition coord, float[] dest)
```

Returns a sequence of float values for a given point in the coverage. A value for each sample dimension is included in the sequence. The default interpolation type used when accessing grid values for points which fall between grid cells is nearest neighbor. The coordinate system of the point is the same as the grid coverage coordinate system.

Parameters: coord - The coordinate point where to evaluate.
dest - An array in which to store values, or null to create a new array. If non-null, this array must be at least getSampleDimensions().size() long.
Returns: The dest array, or a newly created array if dest was null.
Throws: PointOutsideCoverageException if coord is outside coverage.

Méthode 3: Évaluation de valeurs à une coordonnée géographique (API de Geotools 2)

À première vue, une couverture de données ne fait qu'envelopper une image en y ajoutant un référencement géographique. Cette lecture incomplète de l'architecture du système (figure 18) a amené plusieurs utilisateurs de Geotools 2 à croire qu'un objet GridCoverage pouvait accepter n'importe quel objet RenderedImage, notamment ceux qui proviennent de fichiers au format JPEG. Ce raccourci ignore une différence majeure entre ces deux classes d'objets : alors que RenderedImage s'intéresse seulement à la couleur des pixels, GridCoverage se préoccupe de ce que ces pixels mesurent (température en °C, etc.). Les seules images acceptées par notre implémentation sont celles dont le modèle de couleurs fait correspondre chaque valeur de pixel à un index dans une table (IndexColorModel). Ces images sont généralement accompagnées de deux tables ou relations :

- Une table indiquant quel triplet de couleurs RGB (rouge-vert-bleu) correspond à chaque valeur d'index. Cette table est fréquemment modifiée afin de changer la palette de couleurs sans modifier les valeurs de pixels d'une image.
- Une relation donnant la valeur géophysique (comme nombre réel avec ses unités) correspondant à chaque valeur d'index. Cette relation est habituellement fixée une fois pour toute et déclarée dans les métadonnées qui accompagnent une image.

Tout modèle de couleurs dit « direct » (DirectColorModel) est inutilisable par notre implémentation, car ils sacrifient les deux relations ci-haut mentionnées. La première est mise en ligne à l'intérieur de l'image qui stocke directement les valeurs RGB (ce qui rend la palette difficilement modifiable), et la seconde est totalement perdue. Cette perte est sans conséquence pour RenderedImage, du moment que l'aspect visuel est préservé. Mais elle invalide GridCoverage, dont une des missions premières est de manipuler des images dont les pixels ont un sens géophysique.

3.4.1 Conversion des valeurs de pixels vers les valeurs géophysiques

Pour des raisons d'économie de mémoire, les valeurs sont le plus souvent stockées sous forme d'entiers de 8 ou 16 bits. Utiliser une méthode evaluate(...) qui retourne des valeurs réelles (comme la méthode 3) ne fait que forcer le type sans changer la valeur numérique. La conversion des valeurs encodées dans l'image vers les valeurs géophysiques (par exemple une température en degrés Celsius) est une opération séparée, presque toujours définie par une relation linéaire. Cette relation est spécifiée indépendamment pour chaque bande. Par exemple dans le cas des images de température produites à la station SEAS de l'IRD sur l'île de La Réunion, les valeurs encodées de l'unique bande s'interprètent comme suit :

Tableau 5: Signification des valeurs de pixels des images de température de la Réunion

Plage de valeurs	Signification
0	Trait de côte
1 à 9	Nuage
10 à 239	Température ($^{\circ}\text{C}$) calculée par $T = 10 + p/10$ où p est la valeur du pixel
240 à 254	Terre
255	Absence de données

Lorsque les méthodes `evaluate(...)` retournent ces valeurs encodées de pixels, leur interprétation en fonction du tableau 5 (ou de tout autre tableau pertinent pour l'image utilisée) est laissée à la charge de l'utilisateur. OpenGIS spécifie simplement quelques attributs à l'intérieur de `SampleDimension` qui permettent à l'utilisateur d'obtenir les informations du tableau 5 :

Tableau 6: Principaux attributs de `SampleDimension` donnant les informations du tableau 5

Type	Attribut	Valeur pour les images décrites par le tableau 5
double	<code>offset</code>	: 10 $^{\circ}\text{C}$.
double	<code>scale</code>	: 0,1 $^{\circ}\text{C}/\text{valeur d'indice de pixel}$.
double[]	<code>noDataValue</code>	: L'ensemble des entiers de 0 à 9 et de 240 à 255 inclusivement.
String[]	<code>categoryNames</code>	: Un tableau de 256 noms, dont les éléments 1 à 9 contiendraient le texte « Nuage », les éléments 240 à 254 le texte « Terre », etc.

Les interfaces d'OpenGIS n'obligent aucunement à utiliser des entiers pour encoder les valeurs des pixels. L'utilisation des entiers découle plutôt du mode de stockage des données, ainsi que des contraintes d'affichage liées aux capacités des cartes graphiques⁵⁶. Pour fins de calcul, il est plus pratique de voir les données comme une matrice de nombres réels sans encodage particulier, ce qui revient à convertir les images de manière à ce que `offset = 0`, `scale = 1` et `noDataValue = {NaN}` (*Not-a-Number*). Cette conversion pourrait être effectuée par l'utilisateur à l'aide des opérations décrites dans la section 3.4.2. Nous verrons toutefois dans la section 3.4.3 (page 69) qu'une solution plus pratique est requise, et que les attributs du tableau 6 ne suffisent pas. Une généralisation de ces attributs (les boîtes jaunes de la figure 18) sera proposée dans la section 3.4.4.2 (page 72).

3.4.2 Opérations sur les matrices de données

Le calcul effectué par les méthodes `evaluate(...)` peut être modifié. Par défaut, elles ne font que retourner la valeur du pixel sous la coordonnée spécifiée, sans transformation d'aucune sorte (sauf un éventuel forçage de type). Mais on peut par exemple leur faire appliquer à la volée la conversion linéaire du tableau 5, suivi d'un calcul de gradient utilisant les pixels voisins, et finalement interpoler le résultat lorsque la coordonnée demandée n'est pas au centre d'un pixel. Ces opérations sont gérées par la classe `GridCoverageProcessor`, qui permet de :

- appliquer des opérations algébriques élémentaires (multiplication par une constante, *etc.*) ;
- changer le nombre de bandes ou leur ordre ;
- ré-échantillonner une image (projection, changement de résolution, sous-région, *etc.*) ;
- changer la méthode d'interpolation pour les points qui ne sont pas au centre des pixels ;
- accentuer l'image, filtres, magnitude du gradient, *etc.*

L'application d'une opération nécessite les éléments suivants :

- le nom de l'opération (voir la liste ci-dessous) ;
- un et un seul objet `GridCoverage` source ;
- des arguments optionnels qui dépendent de l'opération.

⁵⁶ Il est techniquement possible d'afficher des images composées de nombres réels en Java, mais le rendu est très lent. Les seules valeurs compactées qui profitent d'une accélération matérielle sont les entiers non-signés (qui n'autorisent pas de valeurs négatives) sur 8 et 16 bits.

Le résultat est un nouvel objet `GridCoverage` qui contiendra le résultat de l'opération demandée. Les données originales (c'est-à-dire l'objet `GridCoverage` source) ne sont jamais modifiées. Retenir à la fois les données originales et modifiées peut ressembler à un gaspillage de mémoire, mais dans la pratique `GridCoverageProcessor` ne calcule pas immédiatement le résultat de l'opération. Il construit plutôt une chaîne de traitements qui calculera les données à la volée, seulement lorsqu'elles seront nécessaires. Chaque image est divisée en plusieurs tuiles, et seules les tuiles requises pour la région d'intérêt seront calculées⁵⁷. Les opérations reconnues par Geotools 2 comprennent :

Rescale

Applique une transformation linéaire sur les valeurs de pixels d'une image. Cette opération attend en argument les valeurs des coefficients `scale` et `offset`. Les valeurs des pixels de destination sont définies par le pseudo-code suivant (x et y sont des coordonnées pixels et b désigne la bande) :

$$\text{destination}[x, y, b] = \text{source}[x, y, b] \times \text{scale} + \text{offset};$$

En donnant aux coefficients `scale` et `offset` les valeurs du tableau 6, cette opération peut servir à convertir certains indices de pixels (valeurs entières de 10 à 239 dans le tableau 5) en valeurs géophysiques (11,0 à 33,9 °C pour cet exemple). Toutefois, cette opération ne gère pas les données manquantes. Par exemple la valeur 255 du tableau 5 sera convertie en 36,5 °C alors qu'elle devrait être convertie en NaN. Une solution plus générale sera proposée dans la section 3.4.4.3.

Resample

Ré-échantillonne une image en utilisant une géométrie différente. La « géométrie de la grille » comprend la taille de l'image (en pixels) et la relation entre les coordonnées de chaque pixel et les coordonnées géodésiques au centre des pixels (géoréférencement)⁵⁸. Le ré-échantillonnage d'une image peut nécessiter des interpolations dans l'image d'origine, auquel cas le type d'interpolation peut être spécifié comme un argument optionnel. L'opération « Resample » peut effectuer en une seule étape n'importe quelle combinaison des fonctionnalités suivantes :

- **Ré-échantillonnage sans reprojection**

L'image est ré-échantillonnée à une résolution différente.

- **Reprojection**

La nouvelle image peut utiliser un système de référence de coordonnées différent (par exemple une conversion d'un système géographique vers une projection de Mercator).

- **Découpage d'une sous-région**

Une sous-région peut être obtenue comme une image séparée en utilisant cette opération avec le même géoréférencement que l'image d'origine, mais une région plus petite.

Dans tous les cas, le ré-échantillonnage est représenté par une conversion ou une transformation (section 3.3.1) des coordonnées en pixels de l'ancienne géométrie vers celles de la nouvelle. Si l'opération combine plusieurs des points énumérés plus haut, les transformations correspondantes seront concaténées. Le résultat est très souvent une transformation affine⁵⁹, auquel cas elle sera donnée à

⁵⁷ Notre implémentation de `GridCoverageProcessor` fait calculer les données par une chaîne d'opérations de *Java Advanced Imaging*, une bibliothèque développée par Sun Microsystems avec des contributions de partenaires tels que la Nasa. Cette bibliothèque est conçue pour les traitements d'images en général, que ce soit dans le domaine satellitaire, médical ou astronomique. Son lot d'opérations est déjà assez riche. Dans Geotools 2, nous ne faisons qu'y ajouter les notions de positionnement géographique et de catégories typiques des systèmes d'informations géographiques.

⁵⁸ Selon la spécification d'OpenGIS, le géoréférencement fait correspondre le centre des pixels aux coordonnées géodésiques. Dans *Java Advanced Imaging*, le référencement se fait plutôt avec le coin supérieur gauche des pixels. Les deux conventions sont d'usage courant, et l'on passe de l'une à l'autre en ajoutant ou retranchant 0,5 aux coordonnées pixels. Cette correction de 0,5 ne doit pas être négligée ; l'oublier fait dériver les calculs vers des erreurs très vite inacceptables.

⁵⁹ Les transformations ou conversions affines préservent la colinéarité des points, c'est-à-dire que deux points sur une ligne avant la transformation resteront sur la même ligne après la transformation. Elles préservent aussi le parallélisme et les rapports entre les distances. Elles figurent parmi les opérations les plus utilisées, car elles englobent les échelles, translations et rotations. Aussi disposent-elles d'optimisations particulières.

l'opération « *AffineTransform* » de *Java Advanced Imaging*. Dans le cas plus général, la transformation sera utilisée par une opération « *Warp* » de plus bas niveau.

Interpolate

Pour la plupart des paramètres mesurés par télédétection (température, couleur de l'eau, *etc.*), la valeur d'un pixel est représentative de la moyenne sur toute la surface géographique du pixel. Par défaut, la méthode *evaluate(...)* retourne donc la valeur du pixel sous la coordonnée spécifiée, peu importe que cette coordonnée soit au centre du pixel ou près d'un bord. Mais on peut aussi considérer que la valeur est représentative du centre du pixel, et interpoler linéairement à mesure que l'on se rapproche d'un bord. La valeur à un point situé exactement à la frontière entre deux pixels équivaut alors à la moyenne arithmétique de ces deux pixels.

L'unique argument de cette opération indique le type d'interpolation désirée : « *NearestNeighbor* » (absence d'interpolation), « *Bilinear* » (nécessite 2×2 pixels) ou « *Bicubic* » (nécessite 4×4 pixels). Notre implémentation étend la définition d'*OpenGIS* en permettant de spécifier des interpolations de second recours. Par exemple si une interpolation bicubique a produit la valeur *Nan* parce qu'une des valeurs dans son voisinage de 4×4 pixels était elle-même *Nan*, notre implémentation peut se rabattre sur l'interpolation bilinéaire (à la condition que l'utilisateur l'ait autorisé). Celle-ci encourt moins de risque d'échouer puisqu'elle nécessite un voisinage plus petit (2×2 pixels).

NodataFilter

Le couvert nuageux masque plusieurs valeurs dans les images de température et de concentration en chlorophylle-*a*. Plusieurs valeurs masquées se situent sur un pixel isolé entouré de valeurs valides. Ces absences de données sont problématiques pour une étude halieutique pour les raisons suivantes :

- Les opérations qui nécessitent les valeurs des pixels voisins peuvent augmenter l'étendue des valeurs manquantes. Cela affecte notamment les calculs de gradients (discutés dans l'opération suivante), pour lesquels chaque valeur *Nan* se répercute dans un voisinage de la taille des matrices de convolution (tableau 7).
- Les *observations* (dans notre cas, les données de pêches) sont beaucoup plus rares et souvent plus coûteuses à acquérir que les pixels. Lorsqu'une étude nécessite les valeurs des pixels aux positions des observations (par exemple pour des analyses de corrélations), une valeur *Nan* peut entraîner l'exclusion pure et simple de toute observation à cette position. Ces pertes peuvent être réduites lorsque les valeurs des pixels adjacents fournissent une approximation satisfaisante.

L'opération « *NodataFilter* » remplace les valeurs *Nan* par la moyenne des valeurs valides dans une fenêtre de taille arbitraire (3×3 pixels par défaut), à la condition que le nombre de valeurs valides soit égal ou supérieur à un certain seuil (4 par défaut). La moyenne est pondérée par la distance des pixels moyennés au pixel à remplacer.

GradientMagnitude

Calcule l'amplitude du vecteur de gradient, en calculant d'abord deux composantes orthogonales de ce vecteur. Cette opération est définie par :

$$\text{destination}[x,y,b] = \sqrt{(\nabla_h \text{source}[x,y,b])^2 + (\nabla_v \text{source}[x,y,b])^2}$$

Équation 1: Calcul de gradient à partir d'opérateurs de convolution

où ∇_h et ∇_v sont des opérateurs calculant les composantes horizontale (est-ouest) et verticale (nord-sud) du gradient, respectivement. Ces opérateurs sont définis analytiquement par des dérivées partielles, ce qui implique que (comme la plupart des dérivées) ils tendent à accentuer le bruit. En traitement d'images, ces opérateurs sont des convolutions qui approximent les dérivées par une différence finie. Quelques opérateurs de convolution couramment utilisés à cet effet sont :

Tableau 7: Opérateurs de convolutions utilisés pour les calculs de magnitude du gradient

Opérateurs de Sobel			Opérateurs isotropiques		
∇_h			∇_v		
-1	0	1	-1	-2	-1
-2	0	2	0	0	0
-1	0	1	1	2	1

Avant d'appliquer les opérateurs de convolution sur les données, l'opération « **GradientMagnitude** » les teste d'abord sur des gradients horizontal U_h et vertical U_v de 1 unité/pixel. Pour ces gradients, les opérations $\nabla_h(U_h)$ et $\nabla_v(U_v)$ calculent 8 avec l'opérateur de Sobel, tandis que $\nabla_h(U_h)$ et $\nabla_h(U_v)$ calculent 0. La valeur 8 doit être encore multipliée par la taille des pixels, qui peut varier pour chaque image. Le résultat est le produit d'un facteur de normalisation et d'un facteur spatial (Simpson, 1990), qui doit diviser chaque élément des opérateurs de Sobel du tableau 7. La même procédure empirique est appliquée à l'opérateur isotrope, ou à tout autre opérateur peu importe sa taille. Elle permet d'obtenir des valeurs de gradient exprimées selon des unités plus proches des unités conventionnelles, par exemple en °C/km ou en °C/degré d'angle⁶⁰.

Sur les images de température, les pixels contaminés par les nuages paraissent plus froids qu'en réalité. Les calculs de gradients tendent à amplifier ce bruit, d'où l'importance d'un filtre de nuages efficace. Puisque aucun filtre de nuages n'est totalement efficace, les opérateurs de gradients ∇_h et ∇_v devront être aussi robustes que possible en présence de bruit. Dans le cas d'un bruit blanc gaussien⁶¹, les opérateurs isotropiques (tableau 7) apparaissent plus robustes que les opérateurs de Sobel (Simpson, 1990). Notre définition de l'opération « **GradientMagnitude** » continue toutefois à proposer les opérateurs de Sobel comme opérateurs par défaut, par souci de cohérence avec l'opération « **GradientMagnitude** » de *Java Advanced Imaging*.

Nous avons aussi expérimenté quelques extensions à l'opérateur de Sobel. Le tableau ci-dessous représente un opérateur horizontal de taille 9×9 . Le centre (encadré bleu) est l'opérateur de Sobel 3×3 classique (tableau 7). La bordure rouge délimite un opérateur horizontal 5×5 , que nous utiliserons dans la section 4.4 pour les températures LAC (*Local Area Coverage*).

Tableau 8: Opérateur similaire à Sobel étendu à une taille de 9×9 .

-0,25	-0,24	-0,20	-0,12	0,00	0,12	0,20	0,24	0,25
-0,32	-0,33	-0,31	-0,20	0,00	0,20	0,31	0,33	0,32
-0,40	-0,46	-0,50	-0,40	0,00	0,40	0,50	0,46	0,40
-0,47	-0,60	-0,80	-1,00	0,00	1,00	0,80	0,60	0,47
-0,50	-0,67	-1,00	-2,00	0,00	2,00	1,00	0,67	0,50
-0,47	-0,60	-0,80	-1,00	0,00	1,00	0,80	0,60	0,47
-0,40	-0,46	-0,50	-0,40	0,00	0,40	0,50	0,46	0,40
-0,32	-0,33	-0,31	-0,20	0,00	0,20	0,31	0,33	0,32
-0,25	-0,24	-0,20	-0,12	0,00	0,12	0,20	0,24	0,25

$$\frac{2 \cdot \Delta x}{(\Delta x)^2 + (\Delta y)^2}$$

Ces valeurs ont été calculées par $\frac{2 \cdot \Delta x}{(\Delta x)^2 + (\Delta y)^2}$ où $(\Delta x, \Delta y)$ est la position de la cellule relativement au centre. Cette expression est déduite de la magnitude du vecteur gradient, proportionnelle à $1/\sqrt{(\Delta x)^2 + (\Delta y)^2}$, de laquelle on ne retient que la partie horizontale en multipliant par le cosinus de l'angle. Un calcul similaire est effectué pour l'opérateur vertical. Le facteur 2 a été ajouté afin d'obtenir dans l'encadré bleu du tableau 8 les mêmes valeurs que l'opérateur de Sobel 3×3 . Ce facteur facilite les comparaisons, mais n'a aucun impact sur le résultat final puisque l'opérateur sera de toute façon normalisé avant d'être appliqué.

⁶⁰ Cette normalisation est le principal travail fait par Geotools 2. Une fois les opérateurs normalisés, le calcul effectif est laissé à l'opération « **GradientMagnitude** » de la bibliothèque *Java Advanced Imaging*.

⁶¹ Le bruit produit par la présence de nuages n'a généralement pas une distribution gaussienne. Sa distribution dépend plutôt du type, de l'abondance et de la hauteur des nuages, ainsi que du nombre de couches de nuages (Coakley et Bretherton, 1982).

Appliquer l'opérateur du tableau 8 revient à calculer la moyenne de toutes les composantes horizontales du gradient comprises dans la fenêtre de taille 9×9 pixels. Ces gradients sont toujours calculés entre un pixel quelconque et le centre de la fenêtre. Un calcul similaire étant effectué avec l'opérateur vertical, la valeur de gradient obtenue au final (équation 1) provient de la moyenne des composantes, non de la moyenne des amplitudes. Elle sera donc proche de zéro lorsque calculée au centre d'une masse d'eau délimitée par un front circulaire (au milieu d'un tourbillon par exemple), contrairement à ce que pourrait donner la moyenne des amplitudes.

SelectSampleDimension

Sélectionne N bandes d'une image source et copie leurs données vers une image de destination dans l'ordre spécifié. Bien que la sémantique de l'opération soit celle d'une copie, l'implémentation tente de les éviter en ne changeant que les valeurs de certaines références lorsque possible. Cette opération est utilisée avec les images muti-bandes⁶² pour spécifier celles sur lesquelles sera effectué un calcul. Notre implémentation étend la définition d'OpenGIS en permettant un argument optionnel spécifiant la bande à rendre visible au moment de l'affichage.

3.4.3 Limites des spécifications du consortium OpenGIS

Le chapitre 3.4 introduisait quelques éléments des spécifications d'OpenGIS utilisées. Certaines extensions mineures offertes par l'implémentation Geotools 2 ont été présentées au fil du texte. Par exemple, l'opération « `Interpolate` » peut utiliser des méthodes d'interpolation de second recours si la méthode principale trouve NaN. D'autres extensions sont des ajouts de fonctionnalités qui n'étaient pas spécifiées, par exemple l'opération « `NodataFilter` » qui réduit le nombre de pixels ayant la valeur NaN. Ces ajouts ne violent pas la spécification, ces dernières ne faisant que définir le minimum requis. Toutes ces extensions mineures n'entraînent pas la création de nouvelles classes publiques, ni la dépréciation d'interface de programmation d'OpenGIS.

Les problèmes océanographiques mentionnés dans la section 3.1.5 (page 46) jettent un éclairage sur une limitation plus contraignante, entraînant la dépréciation de certaines méthodes d'OpenGIS. Ces contraintes concernent l'encodage des valeurs à l'intérieur d'une image (section 3.4.1) et la manière de faire correspondre ces valeurs encodées aux valeurs géophysiques utilisées pour les analyses. OpenGIS permet de décrire cette conversion par les attributs `scale`, `offset` et `noDataValue` du tableau 6 (page 65), mais n'utilise pas ces attributs lui-même. Cette approche a quelques inconvénients :

- Les attributs `scale` et `offset` ne permettent de définir que des relations linéaires. Or, nos données de concentration en chlorophylle-*a* utilisent une relation logarithmique.
- La gestion des données manquantes (attribut `noDataValue`) est laissée à la charge de l'utilisateur. Une gestion explicite de ces valeurs est laborieuse et rend invalides des bibliothèques existantes. En outre, des tests suggèrent qu'elle peut dégrader les performances⁶³.
- Les conversions entre les images géophysiques et compactées peuvent devenir coûteuses en mémoire et temps de calcul, car laisser ces conversions à la charge de chaque utilisateur ne favorise pas la réutilisation des résultats précédemment calculés.

Pour fins de calcul, nous souhaitons que toutes les données retournées par la méthode `evaluate` (décrise à la page 64) soient toujours les valeurs géophysiques des descripteurs (par exemple une température en degrés Celsius), jamais les valeurs entières des pixels. Les conversions devront être effectuées à la volée. Si la donnée est manquante (par exemple si la coordonnée spécifiée tombe sur un nuage ou sur la terre ferme), alors la valeur renournée sera NaN (*Not-a-Number*). Les valeurs NaN du Java ont plusieurs propriétés utiles pour les traitements d'images, parmi lesquelles :

- Toute opération arithmétique faisant intervenir une valeur NaN produit un résultat NaN. Ainsi, il est possible d'appliquer des algorithmes déjà existants sur une image sans crainte d'obtenir des valeurs aberrantes, résultantes par exemple de l'addition d'une valeur de température avec un nuage. Si une

⁶² Notamment les images dérivées de l'altimétrie, qui contiennent trois bandes : l'anomalie de la hauteur de l'eau (cm) ainsi que les composantes est-ouest et nord-sud du courant géostrophique (cm/s).

⁶³ Voir le rapport GEOT-67 à <http://jira.codehaus.org/browse/GEOT-67>.

des données était NaN (le nuage), le résultat sera automatiquement NaN quelle que soit la valeur de température.

- Plusieurs valeurs NaN sont possibles⁶⁴. On peut avoir une valeur NaN pour les nuages et une autre pour la terre. Il est ainsi possible de continuer à distinguer les nuages de la terre ferme même si les images sont le résultat d'un calcul (une convolution par exemple).

Une raison supplémentaire pour ne travailler que sur les valeurs géophysiques vient de la distribution de certaines données, notamment de l'anomalie de la hauteur de l'eau (SLA). Ces données se distribuent selon une courbe en forme de cloche (pas tout-à-fait gaussienne), dans laquelle la majorité des valeurs se situent entre -30 et +30 cm. Toutefois, on observe quand même des événements extrêmes de l'ordre du mètre. Ces événements sont souvent associés à des tempêtes. Bien que relativement rares, ils ont quand même une importance océanographique. Les tempêtes contribuent à la dissolution d'oxygène gazeux dans une proportion difficile à quantifier. En se brisant, les vagues déferlantes entraînent une portion de l'atmosphère dans des nuages de bulles qui persistent plusieurs minutes et atteignent des profondeurs de 4 fois la hauteur significative des vagues. Lors de fortes pluies, les gouttes d'eau de plus de 2 millimètres de diamètre peuvent elles aussi entraîner de l'air sous forme de bulles Thorpe, 1995. Compresser les valeurs de SLA dans une plage d'index sur 8 bits nous aurait forcé à tronquer l'effet des tempêtes, ce que nous avons voulu éviter.

3.4.4 Extensions aux spécifications du consortium OpenGIS

Le tableau 5 à la page 65 donnait un exemple de règles établissant une correspondance entre les valeurs des pixels sur 8 bits et les valeurs géophysiques désirées. Bien qu'une multitude d'encodages soient possibles, deux formes présentent un intérêt particulier :

- Images compactées :**

Selon ce modèle, les valeurs des pixels sont des nombres entiers positifs sur 2, 4, 8 ou 16 bits. Ces types de données nécessitent moins d'espace de stockage que les nombres réels (32 ou 64 bits), se compressent plus facilement (par exemple dans des images au format PNG⁶⁵) et conviennent mieux aux accélérations matérielles par les cartes graphiques des ordinateurs. Ces types d'entiers positifs font des index efficaces pour la consultation de tables de correspondances, que ce soit pour la valeur géophysique (réelle) à utiliser dans les calculs, ou pour la couleur associée à chaque valeur d'index. Dans le tableau 9 donné en guise d'exemple, les index correspondent à la colonne ombragée.

- Images géophysiques :**

Selon ce modèle, les valeurs des pixels sont des nombres réels donnant directement les valeurs du descripteur du paysage océanique (par exemple la température en degrés Celsius). Les données manquantes sont représentées par des valeurs NaN. L'utilisateur n'a pas à se soucier de convertir les valeurs avant de les utiliser dans des calculs. Autrement dit, la colonne ombragée du tableau 9 n'existe pas.

Le tableau 9 donne un exemple de tables de correspondances pour les images de température de surface produites à l'île de la Réunion. Ce tableau est une version plus détaillée du tableau 5 de la page 65. Pour chaque index (colonne 1), les valeurs des tables de correspondances (colonnes 2 à 4) sont distribuées dans différents objets :

⁶⁴ Pour les nombres réels à simple précision codés selon le standard IEEE 754, les séquences de bits dont la valeur hexadécimale est comprise entre 7FC0 0000 et 7FFF FFFF sont toutes des valeurs NaN, ce qui nous offre plus de 4 millions de possibilités. Ce n'est pas la seule plage de valeurs ayant cette propriété, mais nous nous limitons à celle-ci. La plage choisie permet de distinguer les valeurs NaN en Java au moins sur la plate-forme Windows, mais la portabilité sur les autres systèmes n'est pas garantie. La principale difficulté découle de l'existence de deux types de valeurs NaN : *actif* et *passif* (en anglais : *signaling* et *quiet* respectivement). Sans détailler les différences entre ces deux types, on retient surtout que les valeurs de type actif peuvent (à la discrétion de l'implémentation) être converties en valeurs de type passif lorsqu'elles sont passées en argument à une fonction, ce qui nous empêche de retrouver la séquence de bits initiale. Le type des différentes valeurs NaN dépend du système, et le Java n'offre pas de méthode fiable permettant de les distinguer.

⁶⁵ **PNG:** Portable Network Graphics. Format d'image proposé en remplacement au format GIF.

1. les **valeurs encodées** proviennent des images compactées (`RenderedImage` du Java standard) ;
2. les **valeurs géophysiques** pour chaque index sont listées dans une table « `Lookup` » de JAI ;
3. les **couleurs** pour chaque index sont listées dans un objet `IndexColorModel` du Java standard ;
4. les **catégories** pour chaque index sont listées dans un objet `SampleDimension` d'OpenGIS.

Valeur encodée sur 8 bits	Valeur géophysique (table « Lookup »)	Couleur RGB (table IndexColorModel)	Nom de la catégorie (table SampleDimension)
0	NaN ₀	000000 (noir)	Trait de côte
1	NaN ₁	C3C3C3 (gris)	
...	
9	NaN ₁	C3C3C3 (gris)	Nuage
10	11,0 °C	822382 (violet)	
11	11,1 °C	...	
...	
238	33,8 °C	A0505C (rouge)	Température
239	33,9 °C	...	
240	NaN ₂	D2C8A0 (beige)	
...	
254	NaN ₂	D2C8A0 (beige)	Terre
255	NaN ₃	FFFFFF (blanc)	Absence de données

Tableau 9: Exemple de table de correspondances pour les images de température

La distinction entre les images compactées et géophysiques est purement conceptuelle ; elle n'est pas strictement liée au type des données sous-jacent. Une image géophysique n'est pas synonyme d'une image utilisant des valeurs à virgule flottante. Une image géophysique peut très bien stocker toutes ses valeurs sous forme d'entiers s'il n'y a aucune valeur manquante (NaN) et si la précision des données ne justifie pas l'inclusion des parties fractionnaires. Inversement, une image dans laquelle la valeur -99 désigne une donnée manquante n'est pas une image géophysique, peu importe le type des données sous-jacent, parce que toutes les occurrences de la valeur -99 nécessiteront un traitement particulier (typiquement un conversion en valeur NaN) avant toute utilisation dans un calcul numérique.⁶⁶

Ces deux formes particulières (images compactées et géophysiques) ne sont pas formalisées dans les spécifications du consortium OpenGIS. Cette omission se justifie aisément dans le cas de la forme compactée, puisque la description donnée à la page 70 (entiers positifs sur 16 bits ou moins) découle de contraintes liées à l'implémentation. En revanche, l'absence de statut particulier pour les images dont toutes les valeurs sont déjà géophysiques oblige l'utilisateur à effectuer cette vérification lui-même, et à convertir les données si nécessaire.

La solution proposée dans le cadre de Geotools 2 est illustrée partiellement par les boîtes jaunes de la figure 18 à la page 63. Ce mécanisme est l'extension la plus significative que nous ayons fait par rapport à la spécification d'OpenGIS. Il peut être décrit en trois parties :

3.4.4.1 Description plus détaillée des catégories

À chaque valeur entière de pixel peut être associée une catégorie. Par exemple une classification pourrait attribuer à la valeur 0 la catégorie « zone forestière », à la valeur 1 la catégorie « zone urbaine », etc. La spécification d'OpenGIS permet déjà d'associer une chaîne de caractères (le nom de la catégorie) à chaque valeur. Pour le milieu océanique, cela ne suffit pas. Nos images mélangeant à la fois des catégories dites *qualitatives* (« nuage », « terre », etc.) et des catégories dites *quantitatives* (« température en °C », etc.). Il nous fallait une structure plus complexe capable de distinguer les deux types de catégories et d'effectuer des

⁶⁶ Cette discussion implique que certaines images ne correspondent à aucune de ces deux formes. Par exemple les images de bathymétrie terrestre GTOPO30 représentent les données manquantes par la valeur -9999. Cette seule caractéristique suffit à exclure ces images des formes géophysique (puisque une conversion est nécessaire avant tout calcul numérique) et compactée (puisque les valeurs négatives sont impropre à une utilisation directe par `IndexColorModel`). Ces images restent malgré tout utilisables par `GridCoverage`, mais au prix d'une dégradation considérable des performances puisque notre implémentation doit alors renoncer aux accélérations matérielles.

conversions adaptées à chacune. Là où OpenGIS n'utilisait que la classe String, Geotools 2 étend le concept en définissant un véritable objet `Category` qui :

- au lieu de ne s'appliquer qu'à une valeur entière, s'applique à une plage de valeurs ;
- outre le nom de la catégorie, contient aussi la transformation vers les valeurs géophysiques (c'est-à-dire l'équivalent des attributs `scale` et `offset` du tableau 6).

Une bande (`SampleDimension`) est décrite par une liste d'objets `Category`. En spécifiant les valeurs de `scale` et `offset` dans les objets `Category` plutôt que `SampleDimension`, on obtient un contrôle plus fin qui permet de spécifier des relations linéaires ne s'appliquant qu'à certaines plages de valeurs plutôt qu'une seule relation linéaire s'appliquant à la totalité des valeurs d'une bande. Ce genre de découpage correspond à l'opération « *Piecewise* » de la bibliothèque *Java Advanced Imaging*. Il permet de décrire plus précisément les informations des tableaux 5 et 9 par exemple, ce qui correspond au second objectif énuméré dans la section 3.1.5 (page 46).

3.4.4.2 Généralisation des conversions pour les relations non-linéaires

Les attributs `scale` et `offset` de type « nombres réels » ont été remplacés par un seul attribut de type `MathTransform1D` (figure 18), une interface permettant de programmer une conversion arbitraire. Cette généralisation rend possible l'utilisation de conversions logarithmiques comme celle qui est utilisée pour les images de chlorophylle-*a*, ce qui correspond au troisième objectif énuméré dans la section 3.1.5 (page 46). Cette interface prend aussi en charge la conversion des valeurs énumérées par `noDataValue` en valeurs NaN.

Avec cette généralisation, les attributs `scale`, `offset` et `noDataValue` définies dans `SampleDimension` par OpenGIS deviennent obsolètes. Geotools 2 les conserve toutefois par commodité, les conversions linéaires représentant la majorité des cas. L'unique attribut qui les remplace se nomme `sampleToGeophysics`. Bien que cette substitution soit propre à Geotools 2, elle s'insère naturellement dans les spécifications existantes. En effet, `sampleToGeophysics` joue un rôle tout-à-fait comparable à `gridToCRS` définie dans `GridGeometry` (figure 18) :

- Toutes deux sont de type `MathTransform`, directement ou indirectement.
- Alors que `gridToCRS` convertit les coordonnées des pixels en coordonnées « du monde réel », `sampleToGeophysics` convertit les valeurs de pixels en valeurs « du monde réel ». En d'autres mots, la première applique une transformation dans les dimensions du système de référence des coordonnées (`CoordinateReferenceSystem`), tandis que la seconde agit dans les dimensions d'échantillonnage (`SampleDimension`).
- OpenGIS note que la transformation `gridToCRS` est généralement affine, ce qui implique la linéarité. De même, la transformation `sampleToGeophysics` est généralement (mais pas obligatoirement) linéaire. Les propriétés `scale` et `offset` de la spécification originale ne représentaient que le cas particulier.

3.4.4.3 Basculement entre les deux formes de chaque image

Chaque objet `GridCoverage` contient une référence vers un autre objet `GridCoverage` de la forme opposée : celui qui gère une vue compactée des données contient une référence vers celui qui gère une vue géophysique, et vice-versa. La forme opposée n'est construite que la première fois où elle est demandée. Puisque la création de ces vues et leur stockage sont gérés par `GridCoverage` lui-même, ce dernier peut éviter tout doublon. On peut basculer à la forme géophysique d'une image en appelant la méthode `geophysics(true)`, peu importe la forme d'origine et sans craindre de dupliquer des conversions. De même, on peut basculer inconditionnellement à la forme compactée (pour affichage, enregistrement ou transmission sur le réseau) avec `geophysics(false)`. L'exemple suivant effectue un ensemble d'extractions après avoir basculé une fois pour toute aux valeurs géophysiques. À aucun moment l'utilisateur n'a à se soucier des attributs `scale`, `offset` et `noDataValue` :

```
GridCoverage    coverage = ...;           // Image contenant les données.  
DirectPosition[] positions = ...;        // Coordonnées auxquelles extraire des valeurs.  
float[]         values   = null;          // Valeurs dans toutes les bandes à une coordonnée  
coverage = coverage.geophysics(true);     // Demande que les valeurs soient géophysiques.  
for (int i=0; i<positions.length; i++) {    // Extrait les valeurs pour chaque position  
    values = coverage.evaluate(positions[i], values);  
    // Effectuer les calculs ici sur les valeurs géophysiques à la coordonnée courante.  
}
```

3.5 Les bases de données

Les données de pêches et les informations sur les données satellitaires apparaissent dans deux bases de données disjointes. La base de données satellitaires est commune à plusieurs projets, tandis que la base de données de pêches peut être remplacée par une base de données d'observations des cétacés ou des cas de choléra dans d'autres sujets de recherche. Ces bases de données nous servent à la fois :

- de catalogues pour recenser les données disponibles et identifier les trous ;
- de moteurs de recherche pour retrouver rapidement des images ou des données de pêches à des coordonnées spatio-temporelles arbitraires ;
- d'organiseurs pour construire à la volée des tableaux croisés entre les données de pêches et les données environnementales ;
- d'enregistreurs pour recevoir les résultats de calculs coûteux ou laborieux.

La base de données satellitaires ne contient pas les données environnementales elles-mêmes, mais plutôt des **références** vers les fichiers qui contiennent ces données. On pourrait plus pertinemment la qualifier de base de *méta*-données satellitaires. Ces bases de données sont gérées par les logiciels Access ou PostgreSQL. Leurs données sont exploitées d'abord par une application écrite en Java via des liens JDBC⁶⁷, puis par les routines statistiques en langage S via un lien ODBC⁶⁸ (figure 19).

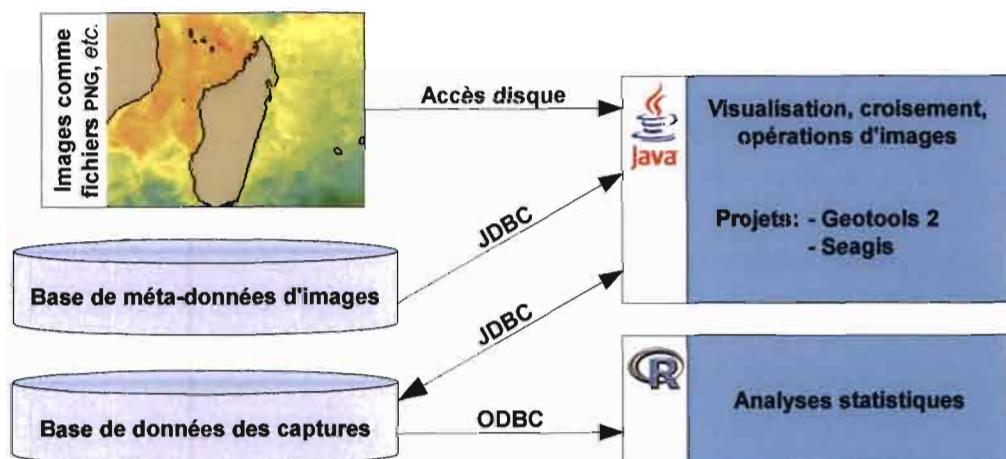


Figure 19: Aperçu des outils et de leurs inter-connections

L'application en Java utilise les données en provenance de toutes les sources. Quelques résultats de ces croisements sont conservés dans la base de données des pêches pour analyse statistique ultérieure, d'où la flèche à double-sens dans la figure 19. Son code a été développé dans des projets collaboratifs, notamment Geotools, et est disponible librement sous licence LGPL. Cette application complète les fonctionnalités des logiciels de bases de données en servant :

- de fournisseur de données environnementales, en assurant tous les services suivants pour des enregistrements quelconques de la base de métadonnées d'images :
 - décoder les fichiers qui contiennent les données (images au format PNG, RAW, etc.) ;
 - transformer les valeurs compactées des pixels en valeurs géophysiques ;
 - appliquer les projections cartographiques, si nécessaire ;
 - calculer à la volée un produit dérivé (par exemple l'amplitude du gradient), si demandé ;
 - interpoler dans l'espace et dans le temps à la coordonnée géographique demandée ;
- d'afficheur capable de superposer quelques couches (produit de base ou dérivé tel que présenté au point précédent, captures, bathymétrie, courants géostrophiques) ;

⁶⁷ **JDBC:** Java DataBase Connectivity – Interface similaire à ODBC, mais adapté au langage Java.

⁶⁸ **ODBC:** Open DataBase Connectivity – Interface standardisée d'accès aux bases de données.

- de navigateur, en présentant un sommaire des données disponibles et en rafraîchissant toutes les couches affichées dès que la sélection change ;
- de modèle, en simulant des déplacements de bancs de thons en fonction de leur environnement.

Ce chapitre décrit d'abord la structure des deux bases de données (satellitaires et de pêches), puis l'interface de programmation (API) développée pour y accéder. Notre API s'appuie sur des standards existants autant que possible (notamment les spécifications du consortium OpenGIS introduites dans la section 3.1.4), auxquels nous avons apporté quelques extensions. Pour aider le lecteur à distinguer les standards des extensions, nous utiliserons la convention suivante :

- les classes, méthodes ou propriétés définies par le consortium OpenGIS seront écrites en bleu (exemple : `CoordinateReferenceSystem`) ;
- les classes, méthodes ou propriétés ajoutées dans le cadre de cette thèse seront écrites en jaune (exemple : `Category`) ;
- les autres entités (noms de tables, classes standards du Java) n'auront pas de couleur particulière (exemple : `RenderedImage`).

3.5.1 Termes employés pour désigner les données

Les descriptions des bases de données (chapitre 3.5) et des tests statistiques (chapitre) désignent les données par différents noms. Les données de pêche sont parfois appelées *données de prospection* pour souligner que les positions notées n'incluent pas que les endroits où il y a eu un coup de senne.

De façon générale, les *paramètres environnementaux* sont des quantités géophysiques dérivées des capteurs (par exemple la température de surface de la mer), tandis que les *descripteurs du paysage océanique* sont des quantités dérivées des paramètres environnementaux et associées à des données de pêche. Le terme « paysage océanique » dans ce contexte s'interprète comme le paysage perçu par le thon, directement (*via* ses sens) ou indirectement (par les effets sur d'autres descripteurs). Par exemple le température dans un milieu 15 jours avant la présence ou l'absence de thon, ainsi que le gradient de température le jour même, font deux descripteurs du paysage océanique dérivés du même paramètre environnemental : la température.

Chaque donnée de pêche avec son ensemble de descripteurs du paysage océanique forme une unité d'échantillonnage. En océanographie classique, une unité d'échantillonnage correspondrait au prélèvement d'un échantillon d'eau, et les descripteurs correspondent aux différentes caractéristiques mesurées sur cette eau. Cet emploi du terme « descripteur » est conforme à l'usage que nous en ferons, selon la terminologie employée dans l'analyse des séries statistiques multiples (Legendre et Legendre, 1979).

3.5.2 Structure de la base de métadonnées satellitaires

Nous enregistrons les données satellitaires dans des formats textes (`ASC`), binaires simples (`RAW`) ou dans des formats d'images compressées tels que `PNG`⁶⁹. Ces formats n'enregistrent pas certaines informations importantes telles que les coordonnées spatio-temporelles des données. Bien qu'une partie de ces informations soient codées dans les noms des fichiers⁷⁰, nous ne nous y fions pas parce qu'elles sont trop incomplètes, trop coûteuses à utiliser comme critère de recherche et suivent des conventions qui varient trop selon les fournisseurs. Il nous fallait une base de métadonnées capable de contenir des informations sur les différents paramètres (unités et précision de la température, hauteur de l'eau, etc.) enregistrés dans différents formats, dans différents répertoires et couvrant différentes régions géographiques. La structure retenue est bien adaptée à nos données, mais ne saurait couvrir tous les besoins. Les personnes désirant une structure beaucoup plus puissante (quitte à en payer le prix sur le plan de la complexité) sont invitées à consulter les propositions du consortium OpenGIS⁷¹.

⁶⁹ **PNG:** *Portable Network Graphics*.

⁷⁰ Par exemple le fichier « SYN99036.png » contient une image de synthèse de température au format `PNG` acquise le 36^{ème} jour julien (5 février) de l'année 1999.

⁷¹ *in : OpenGIS simple features specification for SQL* dans <http://www.opengis.org/techno/specs.htm>.

La figure 20 illustre la structure de notre base de données. Il s'agit du minimum que s'attendra à trouver le programme Java qui l'interrogera. L'utilisateur ne devrait pas supprimer ou renommer des champs qui appartiennent à cette structure minimale. Il peut toutefois ajouter de nouveaux champs, par exemple le nom d'une personne responsable pour chaque série. La structure actuelle s'inspire de quelques interfaces d'OpenGIS énumérées dans le tableau suivant. Il est à noter que ces interfaces évitent les termes « images » et « bandes », sans-doute parce que les données ne proviennent pas nécessairement du domaine visible. Bien que nous utiliserons ces termes par simplicité dans ce document, ils devront être interprétés comme les équivalents de GridCoverage et SampleDimension respectivement.

Tableau 10: Interfaces d'OpenGIS reflétées dans la base de données

GridCoverage	Données réparties sur une grille régulière, comme dans une image ou matrice.
GridGeometry	Géométrie d'un GridCoverage, incluant sa taille en pixels et la transformation affine passant d'une coordonnée pixel vers une coordonnée géographique.
Envelope	Coordonnées géographiques englobant complètement un GridCoverage.
CoordinateSystem	Système de coordonnées (géographique, projetée...) ⁷² de l'enveloppe.
SampleDimension	Une des bandes de l'image représentée par GridCoverage.
Category	La signification d'une valeur de pixel dans une bande.

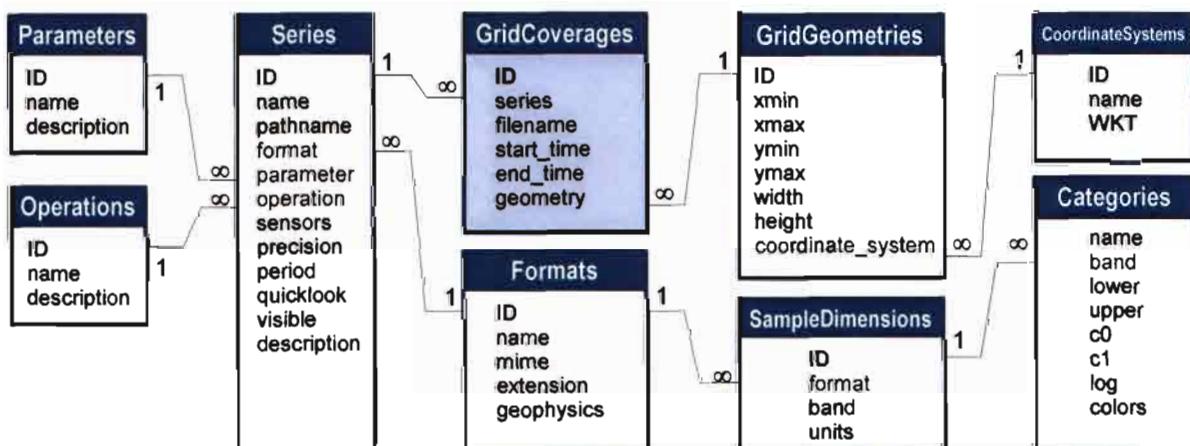


Figure 20: Structure de la base de données d'images

Les références vers toutes les données sont déclarées dans la table GridCoverages. Chaque entrée de cette table comprend le nom du fichier qui contient les données (généralement sous forme d'image), ainsi que les coordonnées spatio-temporelles de son enveloppe. La composante spatiale de l'enveloppe apparaît dans une table séparée (GridGeometries) car elle est identique pour de nombreuses images. La composante temporelle apparaît directement dans la table GridCoverages sous la forme d'une date de début et une date de fin d'acquisition. Pour les images de moyennes ou de synthèses, il s'agira des dates de la première et dernière image ayant servi à calculer la moyenne ou synthèse.

Chaque image (GridCoverage) appartient à une série (un enregistrement de la table Series), qui représente la combinaison d'un paramètre et d'une opération ainsi que d'autres critères laissés à la discréction de l'utilisateur. Par exemple le paramètre « *température* » sur lequel on applique l'opération « *synthèse* », en ajoutant qu'il s'agit de données acquises par l'antenne de l'île de la Réunion, donne la série « *Synthèses de température autour de l'île de la Réunion* ». Les principaux paramètres sont :

⁷² Le terme **CoordinateSystem** était employé dans la version 1.0 de la spécification [Coordinate transformation services implementation](#). Les dernières révisions utilisent plutôt le terme **CoordinateReferenceSystem**.

Tableau 11: Principaux paramètres dans la base de métadonnées satellitaires

Abréviation	Description	Unités
CHL	Concentration en chlorophylle- α	mg/m ³
SST	Température de surface de la mer	°C
SLA	Anomalie de la hauteur de l'eau	cm
U, V	Composantes est-ouest (U) et nord-sud (V) du courant géostrophique	cm/s
EKP	Pompage d'Ekman, assimilée à une vitesse de remontée de la thermocline	cm/jour
FLR	Altitude du plancher océanique par rapport au niveau moyen de la mer	m

La table Operations énumère les algorithmes qui ont servi à produire les valeurs des séries. Pour un même paramètre (par exemple la température), on peut avoir une série dans laquelle chaque image représente une synthèse sur cinq jours, ainsi qu'une série où chaque image représente une moyenne mensuelle. Ces traitements ont été effectués par les fournisseurs de données et ne sont pas reproduits par nos programmes⁷³; la table Operations ne nous sert donc qu'à des fins de classification. Elle contient principalement les entrées suivantes :

- **Analyse objective**

Image dont chaque pixel a été calculé par analyse objective à partir de plusieurs valeurs qui n'étaient pas disposées sur une grille régulière. Cette méthode de calcul ne concerne que les paramètres dérivés des altimètres (SLA, u et v).

- **Analyse objective en temps quasi-réel**

Analyse objective calculée en temps quasi-réel. La précision peut être inférieure à celle qui résulte des analyses objectives calculées plus tard.

- **Synthèse**

Image qui synthétise plusieurs données prises sur une plage de temps de quelques jours. Elle combine plusieurs passages d'un ou plusieurs satellites. En cas de chevauchement, elle retient les valeurs de pixels qui semblent les plus probables.

- **Moyenne**

Image dont chaque pixel est la moyenne des valeurs des pixels correspondants de plusieurs images de synthèse.

Toutes les données d'une même série sont codées dans le même format d'image. Ces formats sont énumérés dans la table Formats et comprennent non seulement le type de fichier (PNG, RAW, etc.)⁷⁴, mais aussi la manière d'interpréter les valeurs des pixels. Une transformation des *valeurs compactées* vers les *valeurs géophysiques*⁷⁵ est nécessaire du fait que les images codent généralement les pixels sous forme de valeurs entières de 0 à 255 inclusivement. Cette transformation ainsi que les unités des valeurs géophysiques résultantes peuvent être différentes pour chaque bande de l'image. La plupart de nos images n'ont qu'une seule bande, mais certaines données dérivées des altimètres font exception avec trois bandes : SLA, u et v.

La table SampleDimensions liste les bandes attendues pour chaque format d'image. Chaque bande contient une liste de catégories énumérées dans la table Categories. Une catégorie est une plage de valeurs compactées que l'on interprète comme une information qualitative (*terre*, *nuage*, etc.), ou comme une information quantitative à transformer en valeurs géophysiques. Dans ce dernier cas, la catégorie fournit les coefficients C_n de l'équation $y = C_0 + C_1x$ effectuant cette transformation. La figure 21 donne une vue des tables Formats, SampleDimensions et Categories. Dans cet exemple, le premier format contient trois bandes (SampleDimension), et la troisième bande contient deux catégories : une pour les données manquantes, et une pour la composante nord-sud du courant géostrophique. Cette dernière est reliée aux valeurs compactées (x) par $y = 0,5x - 64$ cm/s, tel que l'indiquent les coefficients C_0 et C_1 de la table.

⁷³ Ces opérations ne doivent pas être confondues avec celles des chapitres suivants. Le reste de ce document utilisera le terme « opérations » pour des traitements d'images plus simples et relativement standards, que nous effectuerons à l'aide d'un processeur d'images (GridCoverageProcessor).

⁷⁴ Les formats de fichiers sont en fait identifiés par leur code MIME (*Multipurpose Internet Mail Extensions*) autant que possible, par exemple « image/png ».

⁷⁵ Cette transformation a été nommée `sampleToGeophysics` au chapitre 3.4, par analogie avec la transformation `gridToCoordinateSystem` d'OpenGIS. Ces deux transformations jouent fondamentalement le même rôle, la première dans l'espace des valeurs des pixels et la seconde dans l'espace des coordonnées des pixels.

Nom		MIME	Extension	Géophysique
Bandes	Unités			
ASC Anomalie de hauteur de l'eau et courants géostrophiques	text/x-grid-CLS	asc		<input checked="" type="checkbox"/>
1 cm				
2 cm/s				
3 cm/s				
Nom de la catégorie		Minimum	Maximum	C0
Absence de données		0	0	<input type="checkbox"/> #000000
Composante V du courant géostrophique		1	255	-64,000
				<input type="checkbox"/> file:SLA-Aviso.pal
ASC Bathymétrie				
PNG Chlorophylle [0,01 .. 64,57 mg/m³]	text/x-grid-bathy	asc		<input checked="" type="checkbox"/>
PNG Chlorophylle [0,01 .. 66,83 mg/m³] (Méditerranée)	image/png	png		<input type="checkbox"/>
PNG Pompage d'Ekman [-43,2 .. 43,2 cm/jour]	image/png	png		<input type="checkbox"/>
PNG potentiel de pêche [-2 .. 2]	image/png	png		<input type="checkbox"/>
PNG sans paramètres	image/png	png		<input type="checkbox"/>
PNG Température [15 .. 31 °C]	image/png	png		<input type="checkbox"/>
Bandes		Unités		
1 °C				
Nom de la catégorie		Minimum	Maximum	C0
Tout le ciel		0	0	<input type="checkbox"/> #000000
Nuage		1	9	<input type="checkbox"/> #C3C3C3
Plage inutilisée		10	49	<input type="checkbox"/> #B22382
Température de surface		50	210	10,000
Plage inutilisée		211	239	0,100
Terre		240	254	<input type="checkbox"/> #A0505C
Absence de données		255	255	<input type="checkbox"/> #D2CBA0
PNG Température [-3 .. 35,25 °C]	image/png	png		<input type="checkbox"/> #FFFFFF
PNG Température [6,1 .. 31,4 °C]	image/png	png		<input type="checkbox"/>
RAW Anomalie de hauteur de l'eau	image/raw-msla	raw		<input checked="" type="checkbox"/>
RAW Bathymétrie	image/raw	raw		<input checked="" type="checkbox"/>
RAW Pompage d'Ekman [-43,2 .. 43,2 cm/jour]	image/raw	raw		<input checked="" type="checkbox"/>

Figure 21: Table des formats, avec bandes (SampleDimensions) et catégories

3.5.3 Structure de la base de données des captures

Cette base de données comprend deux tables principales : une table Captures qui contient la quantité de poissons pêchés (en tonnes) à certaines coordonnées spatio-temporelles, et une table Environnement qui contient la valeur des paramètres environnementaux (température, *etc.*) dans le voisinage des captures. La table des captures a été créée par l'IFREMER à partir des informations fournies par les pêcheurs. La table de l'environnement est remplie automatiquement en interrogeant la base de données d'images pour chaque coordonnée spatio-temporelle des captures. Le nombre de données de pêches dans la table Captures est de l'ordre de 150 000. Le nombre d'enregistrements dans la table Environnements dépend du nombre de paramètres que l'on souhaite évaluer, mais se chiffre habituellement en plusieurs millions. Pour les modèles linéaires du chapitre , la table Captures fournira les variables expliquées *y* (section 4.1.1) tandis que la table Environnements fournira les variables explicatives *x* (section 4.1.2). La figure 22 illustre partiellement la structure de cette base de données.

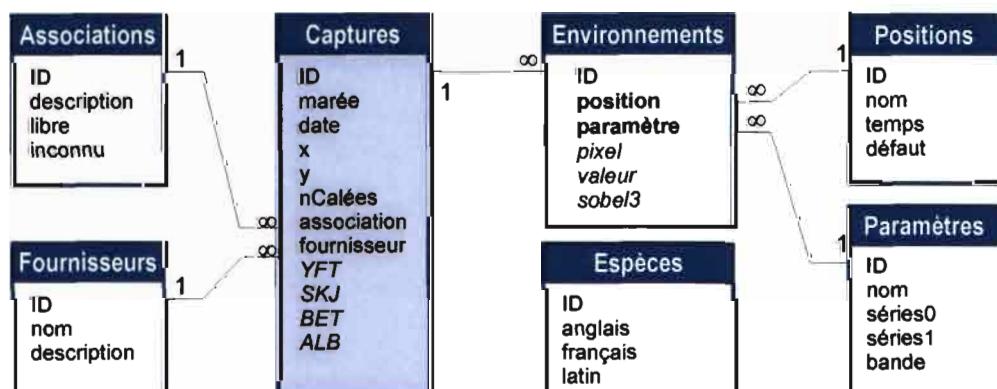


Figure 22: Structure partielle de la base de données des captures

Les bancs capturés par les pêcheurs peuvent être associés à une structure fixe, un objet flottant ou être libres. Le caractère « libres » ou « non-libres » des bancs est spécifié par la table Associations et sera un élément important lors des analyses ultérieures. La table Fournisseurs indique l'origine des données, au

nombre de deux : l'IFREMER (c'est-à-dire les véritables données de pêches), et un générateur de coordonnées spatio-temporelles aléatoires. Ces dernières nous serviront à comparer les conditions environnementales des pêches aux conditions environnementales « standards ».

La table Captures détaille la quantité de poissons pêchés (en tonnes) par espèce et catégorie. Ces quantités apparaissent dans des colonnes nommées d'après le code de la FAO de l'espèce (première colonne du tableau 12). La figure 22 n'illustre que les colonnes YFT, SKJ, BET et ALB, mais une quantité arbitraire de colonnes peuvent être ajoutées à la condition qu'elles soient déclarées dans la table Espèces⁷⁶.

Tableau 12: Quelques codes de la FAO (extrait de la table Espèces)

Code	Anglais	Français	Latin
ALB	Albacore tuna	Germon	<i>Thunnus alalunga</i>
BET	Bigeye tuna	Thon obèse	<i>Thunnus obesus</i>
BIP	Indo-Pacific Bonito	Bonito oriental	<i>Sarda orientalis</i>
BLM	Black Marlin	Malkaire noir	<i>Makaira indica</i>
BLT	Bullet tuna	Bonitou	<i>Auxis rochei</i>
BLZ	Indo-Pacific Blue Marlin	Makaïre bleu de l'Indo Pacifique	<i>Makaira mazara</i>
COM	Narrow barred Spanish Mackerel	Thazard rayé	<i>Scomberomorus commerson</i>
FRI	Frigate tuna	Auxide	<i>Auxis thazard</i>
FRZ	Frigate & Bullet tuna	Auxides et Bonitous	<i>Auxis spp.</i>
GUT	Indo-Parific king mackerel	Thazard ponctué	<i>Scomberomorus guttatus</i>
KAW	Kawakawa	Thonine orientale	<i>Euthynnus affinis</i>
LOT	Longtail tuna	Thon mignon	<i>Thunnus tonggol</i>
MLS	Striped Marlin	Marlin rayé	<i>Tetrapturus audax</i>
OTH	Others species	Autres espèces	<i>Scombridae & Xiphioidei</i>
SBF	Southern Bluefin tuna	Thon rouge du sud	<i>Thunnus maccoyii</i>
SFA	Indo-Pacific Sailfish	Voilier de l'Indo-Pacifique	<i>Istiophorus platypterus</i>
SHX	Sharks	Requins	<i>Lamnidae & Carcharhinidae</i>
SKJ	Skipjack	Listao; Bonite à ventre rayé	<i>Katsuwonus pelamis</i>
STS	Streaked seerfish	Thazard cirrus	<i>Scomberomorus lineolatus</i>
SWO	Swordfish	Espadon	<i>Xiphias gladius</i>
TUN	Tunas & Bonitos (others)	Thons et bonites (autres)	<i>Thunnini and Sardini</i>
YFT	Yellowfin tuna	Albacore	<i>Thunnus albacares</i>

Certaines espèces sont divisées en catégories en fonction du poids des individus. Les moins de 10 kilogrammes sont classés dans la catégorie 1 et les plus de 10 kilogrammes dans la catégorie 3⁷⁷. La catégorie est spécifiée par le suffixe « c1 » ou « c3 » ajouté après le code FAO de l'espèce. Dans le cas de nos données, nous avons donc les espèces/catégories YFT, YFTc1, YFTc3, SKJ, BET, BETc1, BETc3, ALB et OTH. Les colonnes sans indication de catégorie comprennent la somme des captures toutes catégories confondues (par exemple YFT = YFTc1 + YFTc3).

Les enregistrements de la table Captures ne représentent souvent que la synthèse d'une journée. La coordonnée géographique (x,y) est notée à une heure arbitraire (souvent à midi) et les tonnages peuvent englober les captures totales de la journée. Lorsque l'heure n'est pas spécifiée (ce qui est le cas la plupart du temps), nous avons supposé que la coordonnée a été notée à midi. Dans tous les cas, nous supposons que le fuseau horaire est celui de l'île de la Réunion⁷⁸. L'enregistrement peut se référer à une journée sans pêche, une prospection ou une tentative de capture (réussie ou non). Ces cas sont différenciés par les colonnes hPêches, nSennes et nCalées, qui s'interprètent comme suit :

⁷⁶ La déclaration dans la table des espèces n'est pas strictement obligatoire, mais son omission fera ignorer cette espèce par l'application en Java. Ces données ignorées resteront évidemment utilisables dans des requêtes SQL classiques.

⁷⁷ La catégorie 2 était une catégorie intermédiaire de fiabilité incertaine. Elle est assimilée par l'IFREMER à la catégorie 3.

⁷⁸ La zone de pêche s'étendant sur plusieurs fuseaux horaires, cette hypothèse peut introduire une erreur allant jusqu'à 6 heures. Cette erreur reste inférieure à la précision de la date de la majorité des données ainsi qu'à la résolution temporelle des images satellitaires.

- **Journée sans pêche** ($hPêche = 0$)

Le pêcheur n'était ni en train de pêcher, ni en prospection. Il pouvait par exemple réparer une avarie ou retourner vers le port les cales pleines. Ces enregistrements sont identifiés par un nombre d'heures de pêche nul ($hPêches = 0$) et seront ignorés, puisqu'ils ne reflètent pas la présence ou l'absence de thons capturables. Ils représentent moins de 5% des enregistrements.

- **Prospection** ($hPêche > 0$ et $nSennes = 0$)

Le pêcheur n'a observé aucun thon susceptible d'être pêché, ou les bancs observés étaient trop petits pour valoir le coût d'une mise à l'eau du filet. Ces observations sont identifiées par une absence de coups de sennes ($nSennes = 0$). On considérera qu'elles désignent des régions sans thon capturable, bien qu'en réalité une petite quantité de thon était peut-être présente⁷⁹. Elles concernent 40% des données de pêche (c'est-à-dire après avoir ignoré les enregistrements des journées sans pêche).

- **Coups nuls** ($nSennes > 0$ et $nCalées = 0$)

Le pêcheur a fait au moins une tentative de capture ($nSennes > 0$), mais aucune n'a réussi ($nCalées = 0$). Puisque seules les espèces capturées sont répertoriées dans la base de données, nous ne pouvons pas connaître les espèces que le pêcheur avait vues. Ces cas concernent 12% des données de pêche⁸⁰ et seront ignorés.

- **Captures** ($nCalées > 0$)

Au moins un coup de senne a réussi ($nCalées > 0$), ce qui est le cas de 48% des données. Pour être strict, on pourrait exiger qu'il n'y ait aucun coup nul ($nCalées = nSennes$) afin d'augmenter les chances que les captures répertoriées englobent toutes les espèces vues. Nous avons toutefois émis l'hypothèse qu'une calée réussie le même jour et à la même position qu'un coup nul s'est faite sur le même banc. Ces données ont donc été acceptées. Ce dernier cas représente environ 2% des données de pêches.

Un ensemble de paramètres environnementaux ont été évalués pour chaque donnée de pêche. Les types de paramètres sont énumérés dans la table Paramètres et comprennent la température de surface (SST), la chlorophylle (CHL), la hauteur de l'eau (SLA) et le pompage d'Ekman (EKP). Ces paramètres correspondent à ceux de la table du même nom dans la base de métadonnées satellitaires (tableau 11 à la page 77), à la différence près que nous énumérons explicitement les séries à utiliser ainsi que leur ordre de priorité. Par exemple dans le cas de la température nous utiliserons de préférence les données les plus fines, mais nous rabattrons en second choix sur les données mondiales lorsqu'aucune donnée fine n'est disponible à la coordonnée spatio-temporelle d'une pêche. Dans le cas de l'anomalie de la hauteur de l'eau, nous préférerons les données combinées de Topex/Poseidon et ERS-1/2 plutôt que les données de Topex/Poseidon seul. Enfin, les courants géostrophiques ne seront quantifiés par les calculs en temps quasi-réel (*Near Real Time*) qu'en second recours.

Tableau 13: Paramètres environnementaux dans la base de données des captures

Paramètre	Série préférée	Seconde série
CHL	CHL (Monde - hebdomadaires)	CHL (Monde - mensuelles)
SST	SST (Réunion - synthèse)	
SLA	SLA (Monde - TP/ERS)	SLA (Monde - TP)
U	SLA (Réunion)	SLA (Réunion - NRT)
V	SLA (Réunion)	SLA (Réunion - NRT)
EKP	EKP (Monde)	
FLR	Bathymétrie de Sandwell (Réunion)	

Chaque paramètre est évalué à plusieurs positions relativement à une même pêche. Ces positions relatives sont énumérées dans la table Positions et comprennent les instants à 0, 5, 10, 15, 20, 25 et 30 jours

⁷⁹ De fait, dans 14% des cas où il y a eu prospection sans pêche, le pêcheur a quand même noté une association (banc libre, banc associé à un cétacé, etc.). Cela suggère qu'il a aperçu quelques poissons, mais nous ignorons les espèces.

⁸⁰ Cette proportion de 12% est basée sur le nombre d'enregistrements (prospection comprise), puisque la structure de la base de données nous oblige à accepter ou rejeter un enregistrement en bloc. Mais par rapport au nombre de coups de sennes (ce qui exclut la prospection), la proportion de coups nuls est de 25%.

avant la pêche, ainsi que 5 jours après. La décision d'utiliser des multiples de 5 jours découle du mode de calcul des données de base : les synthèses de température autour de l'île de la Réunion sont faites sur 5 jours ; les synthèses mondiales de chlorophylle sur 8 jours ; les anomalies de la hauteur de l'eau sur 10 jours. Un intervalle plus petit que 5 jours apporte peu de nouvelles informations ; il ne fait qu'augmenter le nombre d'interpolations. Notons enfin que bien que cette organisation soit conçue de manière à accepter des décalages spatiaux en plus des décalages temporels, cette dernière fonctionnalité n'a pas encore été exploitée.

La donnée de pêche (référencée par son numéro ID), le paramètre évalué (référencé dans la colonne paramètre) et la position relative à laquelle il a été évalué (référencé dans colonne position) forment conjointement la clé primaire de la table Environnements. Chacune des colonnes restantes représente une valeur obtenue par différentes chaînes d'opérations. Par exemple la colonne sobel3 contient non pas la valeur du pixel, mais plutôt une estimation du gradient calculée d'après l'opérateur de Sobel. Le tableau suivant énumère ces colonnes ainsi que les chaînes d'opérations qui servent à calculer leurs valeurs. Les identificateurs tels que « NodataFilter » et « Interpolate » désignent des opérations qui seront exécutées en Java via l'interface GridCoverageProcessor spécifiée par OpenGIS (voir page 66).

Tableau 14: Chaînes d'opérations pour chaque colonne de la table Environnements.

Colonne	Chaîne d'opérations
pixel	Valeur du pixel sans interpolation ni filtrage de données manquantes.
valeur	NodataFilter; Interpolate Valeur du pixel interpolée dans l'espace et dans le temps.
isotrope	NodataFilter; GradientMagnitude; Interpolate Gradient calculé par l'opérateur isotropique (3×3 pixels).
sobel3	NodataFilter; GradientMagnitude; Interpolate Gradient calculé par l'opérateur de Sobel classique (3×3 pixels).
sobel5	NodataFilter; GradientMagnitude; Interpolate Gradient calculé par un opérateur de Sobel modifié (5×5 pixels).
sobel9	NodataFilter; GradientMagnitude; Interpolate Gradient calculé par un opérateur de Sobel modifié (9×9 pixels).

La figure suivante donne un aperçu du contenu de la table Environnements. Sa structure permet un stockage relativement compact des données (peu de cellules vides), ce qui importe vu que le nombre de lignes se chiffre en plusieurs millions. Elle est aussi pratique à manipuler grâce aux capacités de regroupement des logiciels de bases de données relationnelles. En général, les analyses statistiques n'accéderont pas directement à cette table, mais exploiteront plutôt plusieurs vues (requêtes SQL) adaptées aux différents types d'analyses qui nous souhaitons faire.

Date et marée	dt (journ)	Paramètre	Valeur pixel	Valeur	Isotrope 3x3	Sobel 3x3	Sobel 5x5	Sobel 9x9
15/04/1994 12:00:00 - 1279	-15	SLA	-4,30	-3,54	5,575	5,591	5,009	3,862
15/04/1994 12:00:00 - 1279	-20	SST	30,10	30,06	17,719	17,729	15,516	12,770
15/04/1994 12:00:00 - 1279	-20	SLA	-11,20	-7,34	8,500	8,615	7,412	5,110
15/04/1994 12:00:00 - 1279	-25	SST	27,00	26,81	43,502	42,225	33,979	17,180
15/04/1994 12:00:00 - 1279	-25	SLA	-11,20	-9,88	10,449	10,632	9,013	6,076
15/04/1994 12:00:00 - 1279	-30	SST	29,60	29,46	5,268	5,590	7,916	5,054
15/04/1994 12:00:00 - 1279	-30	SLA	-9,80	-9,15	9,519	9,599	8,249	5,775
15/04/1994 12:00:00 - 1279	+05	SST	29,80	29,75	3,783	3,800	1,458	0,838
15/04/1994 12:00:00 - 1279	+05	SLA	-1,40	-1,56	14,390	14,576	12,948	10,232
07/10/1998 12:00:00 - 3764	+00	CHL	0,20	0,21	0,117	0,133	0,059	
07/10/1998 12:00:00 - 3764	+00	SST	28,40	28,35	4,885	4,906	3,175	1,800
07/10/1998 12:00:00 - 3764	+00	SLA	6,50	6,49	0,469	0,495	0,496	0,979
07/10/1998 12:00:00 - 3764	-05	CHL	0,20	0,19	0,196	0,210	0,088	

Figure 23: Exemple d'enregistrements de la table Environnements

3.5.3.1 Vues des données physico-halieutes

Le chapitre 3.5 décrivait la structure des tables Captures (données halieutiques) et Environnements (données physiques). Les analyses statistiques auront besoin de vues combinant ces deux tables dans une forme plus simple : une ligne par donnée de pêche, qui aligne à la suite des captures tous les descripteurs physiques sélectionnés. Nous nommons ces tables Physico-halieutes. Si toutes les données étaient enregistrées de cette façon dès le départ, nous aurions une table de plusieurs dizaines de millions de cellules dont les trois quarts seraient vides. De plus, chaque ajout d'un nouveau paramètre physique, d'une nouvelle opération ou d'une nouvelle position relative nécessiterait l'ajout d'une dizaine de colonnes. C'est pour cette raison que les données sont enregistrées tel qu'ilustré à la figure 23 et que les tables Physico-halieutes ne sont générées qu'à la volée pour différents sous-ensembles des données disponibles. Ces tables placent sur chaque ligne les informations suivantes :

- les coordonnées spatio-temporelles d'une donnée (date, longitude, latitude) ;
- un drapeau indiquant si la position a été visitée par un senneur ou s'il s'agissait d'une position générée aléatoirement ;
- un drapeau indiquant s'il y avait présence ou absence de thons susceptibles d'être capturés en quantité suffisante (pertinent seulement si le drapeau précédent indiquait une position visitée) ;
- un drapeau indiquant si le banc était libre ou associé (pertinent seulement si le drapeau précédent indiquait une présence de thons) ;
- les quantités capturées pour les espèces et/ou les catégories sélectionnées ;
- les descripteurs sélectionnés du paysage océanique.

Les trois drapeaux apparaissent dans les colonnes visitée, présence et libre respectivement. La figure suivante illustre leur agencement :

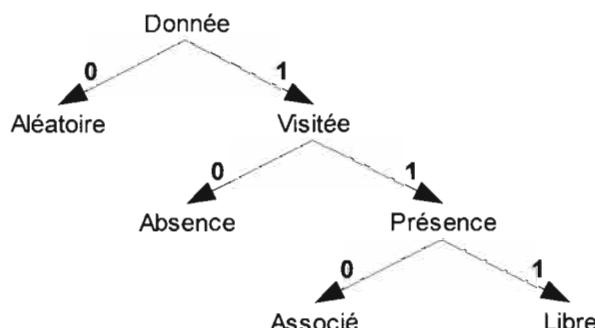


Figure 24: Agencement des drapeaux visité, présence et libre dans les tables physico-halieutes

Le caractère « libre » ou « associé » d'un banc est déterminé par la table Associations. Les données qui nous ont été fournies ne permettent pas de distinguer par exemple les bancs associés à une épave de ceux qui sont associés à un mont sous-marin. Le tableau 15 énumère les types d'associations spécifiées dans la base de données, ainsi que leur classement pour la figure 24 et les analyses du chapitre .

Tableau 15: Associations des bancs notées dans la base de données

Type d'association	Classement	Occurrence
Banc libre, requins, billfishes, tortues	Libre	16,46%
Cétacés	Libre	1,21%
Requin baleine	Libre	0,98%
Épaves, mont sous-marins, charognes	Associé	54,88%
Sous un bateau (thonier, skiff, canneur)	Associé	0,04%
Inconnu, code non-spécifié	Inconnu	27,43%

Les analyses sur les bancs libres n'utiliseront que 15 000 données de captures, soit moins de 19% des données disponibles. Les analyses sur les bancs associés utiliseront 55% des données. Ces restrictions ne s'appliquent pas aux données de prospection.

La création d'une vue Physico-halieutes à partir de la base de données du chapitre 3.5 est effectuée automatiquement par un programme en Java. Cette procédure est pratique lorsque les données halieutiques et environnementales du chapitre 3.5 sont centralisées dans une base de données multi-utilisateurs (par exemple PostgreSQL), et des extractions créées localement pour chaque chercheur (par exemple dans des bases Access) en fonction de l'étude qu'il souhaite effectuer. Des routines statistiques en langage S peuvent puiser directement leurs données dans cette vue via un lien ODBC⁴¹ à chaque exécution, ce qui permet de régénérer l'ensemble des graphiques en cas de changement dans les données. L'existence sur le poste local de cette vue Physico-halieutes formatée selon les besoins de l'étude, et qui peut être elle-même régénérée automatiquement à partir de la base centrale du chapitre 3.5 en cas de besoin, accélère considérablement les analyses statistiques.

3.5.4 Interfaces de programmation accédant aux bases de données

Les chapitres 3.3 et 3.4 décrivaient des interfaces d'intérêt général développées au sein du projet collaboratif Geotools 2, indépendamment de toute base de données. La présente section décrit des interfaces accédant aux bases de données d'images et de pêches présentées dans les sections 3.5.2 et 3.5.3 respectivement. Ces interfaces étant propres à notre problématique, elles ont été développées dans un projet séparé (boîte jaune de la figure 11, page 43).

En utilisant comme brique de base le standard JDBC⁶⁷ pour les connections aux bases de données et le standard OpenGIS pour les représentations spatiales (page 63), nous avons bâti des interfaces de programmation facilitant l'exploitation des données environnementales et de pêches. Chaque base de données dispose d'un jeu d'interfaces dans un paquet qui lui est propre (`fr.ird.database.coverage` et `fr.ird.database.sample`⁸¹ pour les données d'images et de pêches respectivement). L'utilisateur spécifie les coordonnées spatio-temporelles des données qui l'intéressent, l'opération qu'il veut appliquer, etc. Le programme construit des requêtes SQL⁵⁰ en fonction de ces demandes, procède automatiquement à la lecture des images lorsque c'est nécessaire et construit lui-même la succession d'objets requis jusqu'à GridCoverage.

3.5.4.1 Obtention des données environnementales

Toutes les données environnementales sont accessibles par la méthode `evaluate(...)` décrite à la page 63. Cette méthode évalue toujours une valeur provenant d'un objet `Coverage` donné. De cet objet dépend le paramètre évalué (température, courants, etc.), l'opération (valeur interpolée, gradient, etc.) ainsi que la date des données. Les classes décrites dans ce chapitre ont pour but ultime de produire des objets `Coverage` en fonction de critères simples et en puisant les informations nécessaires dans la base de métadonnées d'images. Ces classes s'enchaînent dans une suite d'opérations qui font intervenir plusieurs tables, mais la séquence principale peut être résumée comme suit :

`CoverageDataBase` → `CoverageTable` → `CoverageEntry` → `GridCoverage`

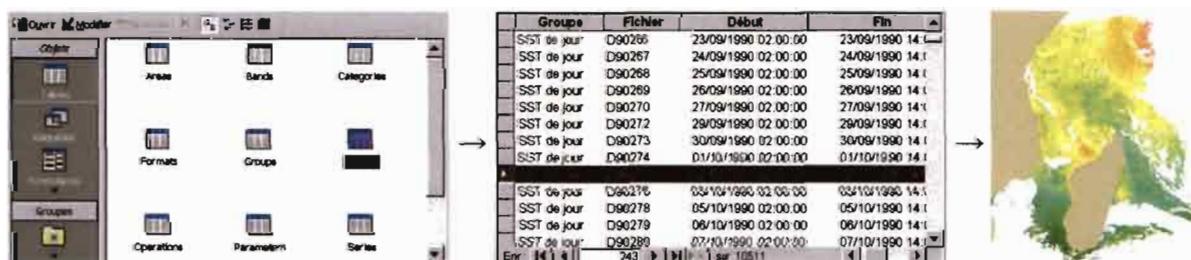


Figure 25: Illustration du séquençage des classes `CoverageDataBase` → `CoverageEntry`

⁸¹ Nous utilisons le nom « sample » plutôt que « fishery » car les mêmes interfaces pourraient être utilisées pour tout type d'échantillonnage ponctuel. La pêche n'est considérée ici que comme une méthode d'échantillonnage parmi d'autres.

L'interface `CoverageDataBase` représente une connexion à la base de métadonnées d'images. Une seule instance de cette classe suffit pour une session de travail. Un objet `CoverageDataBase` peut produire un ou plusieurs objets `CoverageTable`. Chaque objet `CoverageTable` est une vue sur la table `GridCovages` de la base de données (section 3.5.2). Cette vue peut être modifiée en fonction de différents arguments, par exemple les coordonnées spatio-temporelles de la région d'intérêt. Les objets `CoverageTable` peuvent eux-mêmes produire un ou plusieurs objets `CoverageEntry`, qui représentent chacun une entrée (ou ligne) de la table `GridCovages`. On peut considérer un objet `CoverageEntry` comme une référence vers une image, plutôt que l'image elle-même. Les données de l'image ne sont chargées par `CoverageEntry` que sur demande, et retournées sous forme d'objets `GridCoverage`. Les paragraphes suivants expliquent pas à pas les lignes de code qui doivent être écrites pour accéder à des données à partir d'un programme Java. La première étape est d'obtenir une connexion avec la base de données. Cette étape n'a besoin d'être faite qu'une fois pour toutes au début d'une session de travail.

```
CoverageDataBase database = new CoverageDataBase();
```

Les connexions établies de cette façon utilisent une configuration spécifiée dans la préférence du système. Si cette configuration par défaut ne convient pas, `CoverageDataBase` dispose d'un autre constructeur permettant de spécifier explicitement le nom de la base de données, le nom de l'utilisateur et un mot de passe. À partir de cette connexion, on obtient une vue de la table `Covages` pour une série d'images donnée.

```
CoverageTable table = database.getCoverageTable("SST (Monde)");
```

La ligne précédente construisait une table pour les images de température. Le texte spécifié en argument doit être un des noms qui apparaissent dans la table `Series` de la base de données. Typiquement, on répétera cette ligne plusieurs fois pour construire plusieurs tables, une pour chaque paramètre environnemental d'intérêt. Par défaut, ces tables couvrent toute la plage de coordonnées spatio-temporelles disponibles avec la résolution maximale disponible. Il est possible de se restreindre à une région d'intérêt plus réduite. L'exemple suivant sélectionne une région autour de l'île de la Réunion allant de 50 à 60°E et 15 à 25°S pour la période du 24 au 29 août 1998. Les quatre premières lignes ne font intervenir que des classes de la bibliothèque standard du Java.

```
DateFormat parser = DateFormat.getDateInstance(Locale.FRANCE);
Date startTime = parser.parse("24/08/1998");
Date endTime = parser.parse("29/08/1998");
Rectangle2D area = new Rectangle2D.Double(50, -25, 10, 10);
table.setTimeRange(startTime, endTime);
table.setGeographicArea(area);
```

Si l'on souhaite non pas la valeur directe du paramètre environnemental, mais plutôt le résultat d'une opération (par exemple le gradient du paramètre), on peut spécifier le nom de l'opération désirée. Cette opération sera transmise automatiquement à `GridCoverageProcessor` (section 3.4.2) le moment venu. L'exemple ci-dessous demande un calcul de gradient en utilisant les opérateurs horizontaux et verticaux de Sobel.

```
table.setOperation("GradientMagnitude");
```

Maintenant que la table est configurée, on peut demander les images qui interceptent les coordonnées spatio-temporelles spécifiées. Deux méthodes existent à cet effet : `getEntries()` retourne la liste de toutes les images qui interceptent la région et la plage de temps, tandis que `getEntry()` sélectionne une image parmi les candidates. La sélection se fait en privilégiant l'image dont la plage de temps couvre le mieux la plage demandée, et n'examinera l'enveloppe spatiale que si le critère précédent n'a pas permis de les départager. Cette politique est appropriée lorsque les images couvrent à peu près la même région et que leur date est le principal facteur qui varie (ce qui est le cas de nos données). Plus précisément, les critères de comparaison sont dans l'ordre :

- Pour chaque image, la quantité [*temps à l'intérieur de la plage de temps demandée*] - [*temps à l'extérieur de la plage de temps demandée*] sera calculée. Si une des images a une quantité plus grande, elle sera choisie.
- Sinon, si une image se trouve mieux centrée sur la plage de temps demandée, cette image sera choisie.

- Sinon, pour chaque image, l'intersection entre la région de l'image et la région demandée sera obtenue, et la superficie de cette intersection calculée. Si une des images obtient une valeur plus grande, cette image sera choisie.
- Sinon, la superficie moyenne des pixels des images sera calculée. Si une image a des pixels d'une meilleure résolution (couvrant une surface plus petite), cette image sera choisie.

Ces règles comparent toujours les images par paires. Mais elles ont été choisies de manière à définir une relation transitive, c'est-à-dire que $A > B$ et $B > C$ implique que $A > C$ (où $>$ signifie « image préférable à »). L'utilisation de relations transitives permet de classer un nombre arbitraire d'images en ordre de préférence. Ces règles sont implémentées par défaut dans la classe `CoverageComparator`, que le développeur peut redéfinir s'il souhaite utiliser un autre ensemble de règles. L'exemple suivant les utilise :

```
CoverageEntry entry = table.getEntry();
```

Cette méthode est souvent utilisée en couple avec `setTimeRange(...)` : spécifier une courte plage de temps (correspondant par exemple à une pêche), demander la « meilleure » image correspondante, changer la plage de temps, redemander une nouvelle image, etc. Les autres propriétés de `CoverageTable` n'ont pas besoin d'être re-spécifiées à chaque fois si elles ne changent pas.

`CoverageEntry` n'est qu'une référence vers une image. Elle permet de connaître quelques propriétés de l'image (dates exactes, nom de fichier, format, etc.) sans procéder immédiatement à sa lecture. Le chargement des données, s'il est nécessaire, n'est effectué qu'au stade suivant (code ci-dessous). En outre, un système de caches permet à `CoverageEntry` d'éviter de faire des lectures redondantes si l'image se trouve déjà en mémoire. Le programmeur n'a donc pas besoin de maintenir un bilan des images qu'il a déjà lues; c'est fait automatiquement.

```
GridCoverage coverage = entry.getGridCoverage(null);
```

L'argument (null dans cet exemple) permet à ceux qui le désirent de suivre les progrès d'une éventuelle lecture. Si la lecture a réussi, l'utilisateur peut maintenant extraire les données en utilisant la méthode `evaluate(...)` introduite dans le chapitre 3.4. Lorsqu'elle est appliquée sur un `GridCoverage` obtenu par le code précédent, la méthode `evaluate(...)` tente d'interpoler les valeurs en utilisant une interpolation bicubique. Si l'interpolation bicubique échoue parce qu'un nuage masquait des pixels environnants, la méthode se rabat sur une interpolation bilinéaire. Si cette dernière échoue aussi, alors cette méthode retourne la valeur du plus proche voisin⁸². L'exemple ci-dessous utilise une boucle pour afficher les valeurs de la première bande (numérotée 0) à différentes longitudes le long de la latitude 22°30'S. Le même tableau est réutilisé à chaque passage dans la boucle pour plus d'efficacité, écrasant les anciennes valeurs.

```
float[] values = null;
Point2D coord = new Point2D.Double();
for (double x=55.0; x<=60.0; x+=0.5) {
    coord.setLocation(x, -22.5);
    values = coverage.evaluate(coord, values);
    System.out.println(values[0]);
}
```

Dans cet exemple, nous ignorons la troisième dimension (le temps) puisque nous travaillons sur une seule image. Nous pouvions utiliser un objet `Point2D` plutôt qu'un objet `DirectPosition` de dimension arbitraire. Une interpolation a donc été faite dans l'espace, mais pas dans le temps. La dernière étape consiste à ajouter l'axe temporel. Plutôt que de manipuler une série de `Coverages` bidimensionnels (un `CoverageTable`), nous créons une vue à trois dimensions de l'ensemble des données sélectionnées. On créera typiquement une vue 3D par séries d'images, d'où le nom de `SeriesCoverage3D` :

```
Coverage3D coverage3D = new SeriesCoverage3D(table);
```

`Coverage3D` est une extension de l'interface `Coverage` d'OpenGIS, adaptée à l'API décrit dans ce chapitre. Il a l'apparence d'une fonction continue : l'extraction des valeurs se fait en spécifiant les coordonnées spatio-temporelles désirées (dans le système de référence de coordonnées décrit à la page 58), sans se soucier de

⁸² La pertinence de ces interpolations pour nos données halieutiques sera discutée dans la section 4.1.3.

savoir sur quelles images elles se trouvent ni de connaître la taille des voxels⁸³. Les valeurs sont interpolées à la fois dans l'espace et dans le temps, toujours à la condition qu'il n'y ait pas de trous dans les données. L'exemple ci-dessous interpole la donnée du 27 août 1998 à la coordonnée 55°E, 20°S. Nous réutilisons les mêmes objets parser et values que dans les exemples précédents.

```
Date time = parser.parse("27/08/1998");
Point2D point = new Point2D.Double(55,-20);
values = coverage3D.evaluate(point, time, values);
```

Cette classe Coverage3D sert de base à tout ce qui servira à obtenir des données environnementales aux positions de pêches, dans tous les chapitres suivants de cette thèse.

3.5.4.2 Obtention des données de pêches

Les données de pêches s'obtiennent d'une façon tout à fait similaire aux trois premières étapes de l'obtention des données environnementales (section 3.5.4.1). Il faut d'abord obtenir une connexion avec la base de données des captures (indépendamment de la base de données d'images), obtenir une table des captures et enfin les données des captures. Les classes impliquées s'enchaînent comme suit :

```
SampleDataBase → SampleTable → SampleEntry
```

Comme pour les images, la première étape consiste à obtenir une connexion avec la base de données. Cette étape n'a besoin d'être faite qu'une fois pour toutes au début d'une session de travail.

```
SampleDataBase database = new SampleDataBase();
```

D'autres constructeurs sont disponibles si l'on souhaite spécifier un nom d'utilisateur et un mot de passe. L'étape suivante consiste à utiliser cette connexion pour obtenir une vue de la table Captures.

```
SampleTable table = database.getSampleTable("YFT");
```

Le texte spécifié en argument nomme la ou les espèces à prendre en compte (tableau 12 à la page 76). Si cet argument est omis, toutes les espèces sont retenues. Par défaut, la table couvre toute la plage de coordonnées spatio-temporelles disponibles. Il est possible de se restreindre à une région d'intérêt plus réduite en procédant de la même façon que pour les images, notamment en appelant la méthode setTimeRange(...). Une fois la table configurée, on peut demander les captures qui interceptent les coordonnées spatio-temporelles spécifiées. La méthode getEntries() retourne la liste de toutes les captures qui répondent aux conditions, comme dans l'exemple suivant :

```
Collection<SampleEntry> entry = table.getEntries();
```

Chaque objet SampleEntry représente des captures à une position spatio-temporelle définie. Dans le cas des données palangrières, c'est une ligne de palangre mise à l'eau et récupérée. Pour les données des senneurs, c'est un ou quelques coups de senne (par « quelques », on entend au maximum les coups de senne d'une journée). Ces objets contiennent entre autres les méthodes suivantes :

- **getCoordinate()**
Retourne une coordonnée géographique représentative de la capture.
- **getShape()**
Retourne une forme géométrique représentative de la zone de pêche. Par exemple pour les données palangrières, il s'agit d'une ligne approximant la trajectoire de la palangre. Cette méthode ne s'applique pas aux données des senneurs utilisées dans la présente étude. Mais elle deviendra pertinente si une version future de la base de données s'enrichit des positions des bateaux par satellites GPS⁶⁸.

⁸³ Un voxel est l'équivalent 3D d'un pixel, c'est-à-dire le plus petit élément d'un espace tridimensionnel. Le terme *voxel* est l'acronyme anglais de *volume cell element*.

- **getTime()**

Retourne une date représentative de la capture.

- **getValue(Species)**

Retourne les captures pour une espèce donnée (spécifiée en argument). La méthode `getUnit()` doit être interrogée pour connaître les unités. Pour les palangres, les captures se mesurent souvent en nombre d'individus par 1000 hameçons. Pour les données des senneurs, elles se mesurent en tonnes.

A partir d'une donnée de pêche (un objet `SampleEntry`), on peut facilement obtenir les données environnementales à la position correspondante en utilisant un objet `Coverage3D` par paramètre (section 3.5.4.1). La classe `EnvironmentTableFiller` est dédiée à ce croisement de données. Elle extrait les données environnementales aux positions de pêches 30, 25, 20, 15, 10 et 5 jours avant, le jour de la pêche et 5 jours après. Les informations trouvées sont enregistrées dans la table `Environnements` de la base de données des pêches, tel que décrit dans la section 3.5.3.

3.5.5 Enregistrements des activités dans un journal

Les méthodes décrites dans le chapitre 3.5 génèrent des événements qui peuvent être enregistrés dans un journal. Ce journal remplit deux fonctions. Il sert d'abord d'accusé de réception pendant le calcul, nous confirmant que chacune des instructions décrites dans les chapitres précédents a bien été exécutée. Il sert aussi d'archive après le calcul, au cas où l'on voudrait vérifier les opérations qui ont été effectuées. Les événements se déclinent en plusieurs niveaux. En allant du plus sévère (et rare) au plus détaillé (et abondant), on compte les niveaux `SEVERE`, `WARNING`, `INFO`, `CONFIG`, `FINE`, `FINER` et `FINEST`. Les méthodes qui modifient l'état d'un objet `Table` (par exemple pour spécifier des coordonnées spatio-temporelles) génèrent des événements de niveau `CONFIG`. Les méthodes qui extraient des données d'une table génèrent des événements de niveau `FINE` (ces événements étant généralement plus abondants que la configuration d'une table, ils doivent être d'un niveau inférieur). Le journal des événements peut être enregistré dans un fichier XML; son paramétrage fait partie des services standards du Java.

Chapitre 4

Corrélations et modèles

4.1 Variables utilisées et hypothèses

Nous cherchons à établir des corrélations entre les paramètres environnementaux et une quantité se rapprochant autant que possible de la probabilité de présence des thons. Mais cette probabilité étant estimée à partir des données de pêches, elle est soumise à des contraintes mentionnées dans le chapitre 1.3. Ces contraintes ont un impact sur l'utilisation que l'on peut faire de ces données :

- A l'exception des corrélations de Spearman (chapitre) et de quelques figures occasionnelles, nous ne prendrons pas en compte les tonnages capturés. Seule la présence ou l'absence de thons **visibles** par le pêcheur sera prise en compte, peu importe que le dit pêcheur ait réussi à les capturer ou pas. Nous réduisons ainsi les biais dus aux différences d'habiletés des pêcheurs, ainsi que l'impact des facteurs qui modifient la capturabilité des thons (par exemple les changements de technologie ou les variations de la tactique de pêche).
- L'échantillonnage n'est pas aléatoire, puisque la distribution de l'effort de pêche ne l'est pas (section 4.2.1 à la page 96). Les valeurs sont donc corrélées entre elles et l'estimation de l'écart type devrait exiger un modèle de la distribution des bancs de thons (section 1.3.2 à la page 22). Ce problème n'a donc pas été pris en compte pour cette étude, bien que l'ignorer ne soit pas rigoureusement correct.

Ces contraintes impliquent que les résultats statistiques obtenus à partir de ces données ne fournissent pas d'indication sur l'abondance de la ressource, et ne peuvent pas être généralisés à l'ensemble de l'océan Indien. Dans le chapitre 4.4, nous allons tenter de prédire à l'aide de modèles linéaires la **présence** ou **l'absence** de thons **visibles** par le pêcheur (c'est-à-dire près de la surface) **aux positions prospectées** seulement. Ces modèles s'appuieront sur les hypothèses suivantes, qui découlent en partie de contraintes liées à la base de données décrite dans la section 3.5.3 (page 78) :

- la détectabilité des thons reste constante ou ne varie qu'en fonction des paramètres environnementaux que nous mesurons ;
- tous les pêcheurs ont la même habileté à détecter le thon (indépendamment de leur habileté à le capturer) et notent leurs observations dans leurs cahiers de pêche avec la même discipline ;
- le fuseau horaire est présumé égal à celui de l'île de la Réunion ;
- la coordonnée spatiale est présumée notée à midi lorsque l'heure n'est pas spécifiée ;
- les calées et coups nuls combinés dans un même enregistrement sont présumés faits sur le même banc de thons ;
- on peut ignorer les enregistrements qui ne sont composés que de coups nuls (environ 12% des données, voir page 80) ;
- toutes les valeurs d'un même paramètre environnemental ont le même poids, indépendamment de leur incertitude.

Ces hypothèses s'appliquent à l'ensemble du chapitre . Certaines d'entre elles (notamment toutes les hypothèses relatives aux coordonnées spatio-temporelles) sont intervenues avant même ce chapitre, lors de l'extraction des valeurs des paramètres environnementaux (section 3.5.3). La dernière hypothèse découle des modèles de régression linéaire utilisés au chapitre 4.4, qui ne prennent pas en compte l'incertitude sur les données. Or, de nombreux facteurs font varier cette incertitude :

- une mince couche nuageuse peut altérer des données plus que d'autres ;
- certaines coordonnées spatio-temporelles peuvent s'approcher plus que d'autres du centre des pixels, ce qui a un impact sur les interpolations ;
- une moyenne hebdomadaire peut avoir été remplacée par une moyenne mensuelle, ou une valeur finale par une valeur en temps quasi-réel (tableau 13 à la page 80).

D'autres techniques que la régression linéaire offrent une prise en compte des incertitudes, par exemple les réseaux bayésiens (section 5.2.2). L'exploration de ces techniques a été remise à une date ultérieure.

4.1.1 Variables à expliquer

Les variables disponibles sont la présence ou l'absence de poisson des espèces énumérées dans le tableau 16. Ce tableau reprend les espèces ombragées du tableau 12 (page 79), augmentées des attributs « petits / gros poissons » (catégories 1 et 3) lorsque cette information est disponible. La colonne « présence » donne la proportion d'enregistrements dans lesquels des individus de cette espèce ont été observés. Les valeurs de cette colonne sont représentées dans la figure 26 par des histogrammes bleus.

Tableau 16: Codes des espèces et catégories de taille utilisés comme variables expliquées

Code	Nom scientifique	Français	Présence
SKJ	<i>Katsuwonus pelamis</i>	Listao; Bonite à ventre rayé	44,52%
YFTc1	<i>Thunnus albacares</i>	Petit albacore (moins de 10 kg)	44,04%
YFTc3	<i>Thunnus albacares</i>	Gros albacore (plus de 10 kg)	27,65%
BETc1	<i>Thunnus obesus</i>	Petit thon obèse (moins de 10 kg)	42,79%
BETc3	<i>Thunnus obesus</i>	Gros thon obèse (plus de 10 kg)	24,56%
ALB	<i>Thunnus alalunga</i>	Germon	1,31%

Dans 42,76% des enregistrements, on note une présence simultanée de SKJ, YFTc1 et BETc1, tandis que la présence d'au moins une de ces trois espèces est notée dans 44,55% des enregistrements. Les valeurs de la colonne « présence » du tableau 16 pour ces trois espèces sont forcément comprises entre ces deux extrêmes. De la même manière, on trouve l'intervalle 24,59% à 27,65% pour le groupe formé par YFTc3 et BETc3. Ces intervalles sont représentées par des lignes horizontales bleues dans la figure 26. Le caractère étroit de ces intervalles indique que les espèces SKJ, YFTc1 et BETc1 (désignées « groupe C1 » dans les pages qui suivent) sont pratiquement toujours observées ensemble, et de même pour les espèces YFTc3 et BETc3 (désignées « groupe C3 »). Cette observation traduit le fait déjà connu que les thons tendent à se regrouper par taille davantage que par espèce (Petit, 1991). Le listao (skj) étant une espèce de petite taille (section 1.1.1), il n'est pas surprenant de retrouver ces individus parmi les petits albacores et les petits thons obèses. D'ailleurs, ces espèces sont souvent confondues dans les données de pêches.

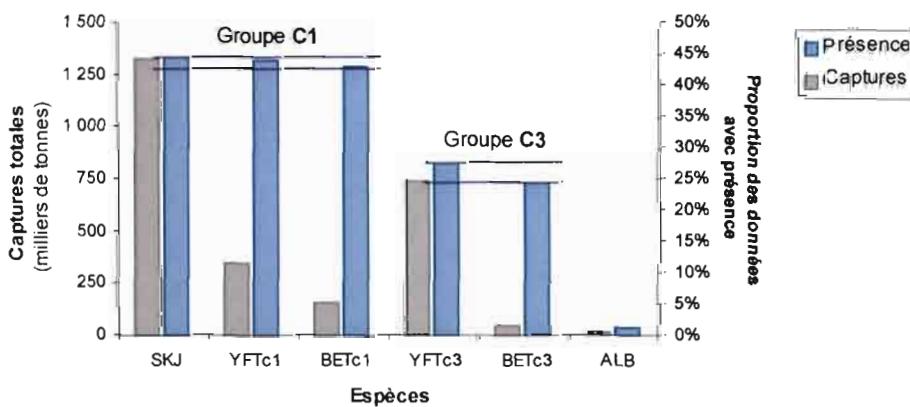


Figure 26: Captures par espèces et proportions de coups de sennes non-nuls

L'analyse en composantes principales⁸⁴ confirme l'existence des groupes C1 et C3, ces deux axes expliquant 98% de la variance. Le cas du germon (ALB) est particulier et sera traité séparément des autres groupes. Les analyses des pages qui suivent porteront donc principalement sur les groupes C1 et C3, sans distinguer les espèces qui les composent. Nous utiliserons l'opérateur ou, c'est-à-dire que nous considérerons qu'il y a présence d'une catégorie dès qu'au moins une de ses espèces constitutantes est présente. En outre, pour chaque groupe C1 ou C3, les données sont réparties entre bancs libres et bancs associés tel que spécifié dans le tableau 15 (page 82). On obtient donc un total de 4 variables à expliquer.

⁸⁴ Cette analyse a été appliquée sur les quantités de poissons pêchés, plutôt que sur les valeurs booléennes (présences / absences) utilisées dans la figure 26.

4.1.2 Variables explicatives

Les variables explicatives sont des *paramètres environnementaux* (tableau 13 à la page 80) évalués à l'aide de diverses *opérations* (tableau 14 à la page 81) à plusieurs *positions relatives* aux positions prospectées (page 80). Chaque combinaison de ces trois critères forme un *descripteur du paysage océanique*. Pour chaque prospection p_i ($i = 1 \dots n$, le nombre de données de prospection), la valeur x_i du descripteur pour une combinaison donnée peut être représentée par :

$$x_i = g(E) \Big|_{\vec{p}_i + \vec{r}} \quad \text{où} \quad \begin{aligned} E &\text{ est le paramètre environnemental sous forme d'images ;} \\ g(\dots) &\text{ est l'opération à appliquer sur l'image (souvent un gradient) ;} \\ \vec{p}_i &\text{ est une des positions de prospection pour } i = 1 \dots n ; \\ \vec{r} &\text{ est la position relative aux prospections.} \end{aligned}$$

Équation 2: Évaluation d'un descripteur du paysage océanique

Cette expression est évaluée dans un espace tridimensionnel, selon le système de coordonnées de la page 58. Selon cette définition, le paramètre environnemental E est en soi une forme de descripteur du paysage océanique lorsque :

- l'opération $g(\dots)$ est l'opération identité ;
- la position relative \vec{r} est un vecteur nul.

On peut facilement créer des milliers de descripteurs ; la figure 23 (page 81) illustre les valeurs de quelques-uns. Ces différents descripteurs sont fortement corrélés entre eux. Les analyses statistiques se feront donc d'abord sur les paramètres environnementaux, utilisés comme la forme la plus simple des descripteurs. D'autres types de descripteurs seront introduits progressivement au gré des besoins.

La pertinence de l'opération $g(\dots)$ dans la définition d'un descripteur du paysage océanique se comprend bien si on compare avec les paysages terrestres. Par exemple pour un marcheur, le gradient de l'altitude (c'est-à-dire la pente) importe souvent plus que l'altitude absolue du paysage dans lequel il évolue.

4.1.3 Pertinence des interpolations des données environnementales

La résolution de nos données environnementales varie de 1 à 15 minutes d'angle, soit environ 2 à 28 kilomètres à l'équateur. En comparaison, la *Résolution* des coordonnées de pêches est généralement d'une minute d'angle. Cette résolution ne doit pas être confondue avec la *Précision*, difficile à déterminer puisqu'elle dépend de la discipline du pêcheur. Les coordonnées ne sont souvent notées qu'à midi, alors que les bateaux peuvent parcourir plusieurs centaines de milles nautiques par jour. On note toutefois que dans notre base de données, 25% des coordonnées prospectées sont situées à moins de 30 milles nautiques de la position de la veille, et 50% à moins de 80 milles nautiques (figure 27).

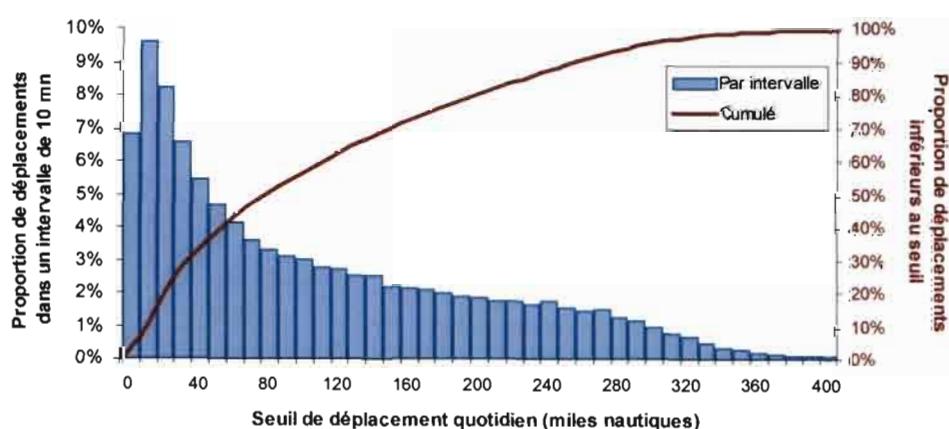


Figure 27: Déplacements maximums des bateaux par rapport à la position de la veille

Ces déplacements sous-estiment les distances réellement parcourues, puisque des bateaux peuvent prospecter sur de grands arcs de cercle avant de revenir à proximité de leur position initiale. Toutefois, l'information recherchée n'est pas la distance parcourue, mais plutôt l'éloignement de la coordonnée notée à midi. Considérant que midi tombe au milieu de la période d'activité (les pêches étant rares la nuit dans l'océan Indien) et qu'une calée dure plusieurs heures pendant lesquelles le bateau se déplace peu, nous supposerons que l'écart entre la position de la pêche et la position notée dépasse rarement la distance par rapport à la position de la veille. Si on admet cette hypothèse, alors la courbe de la figure 27 donne une estimation de l'incertitude sur les positions de pêche ou de prospection. On observe qu'environ 12% des coordonnées ont une précision plus fine que la dimension à l'équateur d'un pixel d'anomalie de la hauteur de l'eau (SLA). Cette proportion augmente à près de 18% si l'on ignore les journées de prospection sans pêche. Il est possible que la proportion réelle soit encore plus élevée du fait que certains pêcheurs notent la coordonnée réelle de la pêche plutôt que la position de midi.

Pour une quantité non négligeable de données de pêches, la position à l'intérieur d'un pixel pourrait donc avoir une signification. L'échantillonnage à un degré de précision inférieur au pixel est un problème classique en traitement d'images, et les bibliothèques telles que *Java Advanced Imaging* (utilisée dans cette thèse) proposent depuis longtemps plusieurs méthodes d'interpolations spatiales. Pour éviter de perdre la donnée lorsqu'un pixel voisin – nécessaire au calcul – est masqué par un nuage, nous avons appliqué non pas un seul type d'interpolation, mais plutôt la chaîne ci-dessous. Si la première interpolation ne réussit pas, nous essayons le type suivant et ainsi de suite :

- bicubique (4×4 pixels) utilisant le coefficient a recommandé par Rifman (Wolberg, 1990) ;
- bilinéaire (fenêtre de 2×2 pixels) ;
- plus proche voisin, ce qui équivaut à une absence d'interpolation.

En plus d'une interpolation spatiale, nous avons aussi appliqué une interpolation temporelle linéaire sur les données (section 3.5.4.1 à la page 86). Ces interpolations suscitent les réserves habituelles ; elles supposent notamment qu'il n'y a pas de signal fort à une fréquence trop rapide par rapport à la fréquence d'échantillonnage. Si on exige que les longueurs d'ondes ne soient pas inférieures à la taille de deux voxels⁸⁵ de nos données les moins fines, alors la longueur d'onde et la période d'oscillation minimales sont de 30 milles nautiques et 20 jours respectivement. Les interpolations ne sont donc pas appropriées lorsque nos données sont affectées principalement par les ondes internes, inertielles, les petites tempêtes ou les tornades. Mais elles restent appropriées lorsque dominent les ondes de Rossby, de Kelvin, les dépressions aux latitudes moyennes, les cyclones tropicaux, les tourbillons à mésoscale et les fronts océaniques.

L'aptitude des interpolations à améliorer la description de l'environnement dépend donc des données disponibles. Nos données de température permettent un test simple, puisqu'elles sont disponibles à deux résolutions différentes issues de chaînes de traitements distinctes : **LAC** (*Local Area Coverage*, 1×1 degré d'angle par jour) et **GAC** (*Global Area Coverage*, environ 5×5 degrés d'angle par 8 jours). La figure 28 illustre les données GAC en fonction des données LAC aux positions et aux dates prospectées. Les lignes oranges représentent des droites de pente 1. La dispersion des nuages de points autour de cette droite théorique s'explique par les facteurs suivants :

- Les points sans interpolation (figure de gauche) apparaissent alignés sur un quadrillage. Cet artefact découle du mode de stockage des données LAC et GAC, qui les arrondit au dixième de degré.
- Le mode de stockage des données LAC ne supporte pas de température supérieure à 31°C , d'où le caractère tronqué des deux nuages de points.
- Malgré l'absence de température supérieure à 31°C , la majorité (70%) des données LAC sont tout de même plus chaudes que les données GAC (en dessous de la ligne orange dans la figure 28). Cette différence, déjà observée par les pêcheurs, s'explique par la différence d'algorithme de calcul de la température de surface entre le mode GAC et le mode LAC. Le premier fait l'hypothèse d'une atmosphère globale homogène, tandis que le second possède ses propres coefficients calculés en fonction des informations locales.
- Certaines données LAC sont beaucoup plus froides (au moins de 3°C) que les données GAC. Cette différence s'observe dans la figure 28 par des points décalés vers la gauche dans la partie au-dessus de

⁸⁵ Ce critère correspond à la fréquence critique de Nyquist utilisée en analyse spectrale, soit $1/2\Delta$ où Δ est la période d'échantillonnage.

la ligne orange. Elle s'explique par un filtre de nuages plus permissif dans le cas des données LAC, alors que l'algorithme produisant les données GAC est plus conservateur.

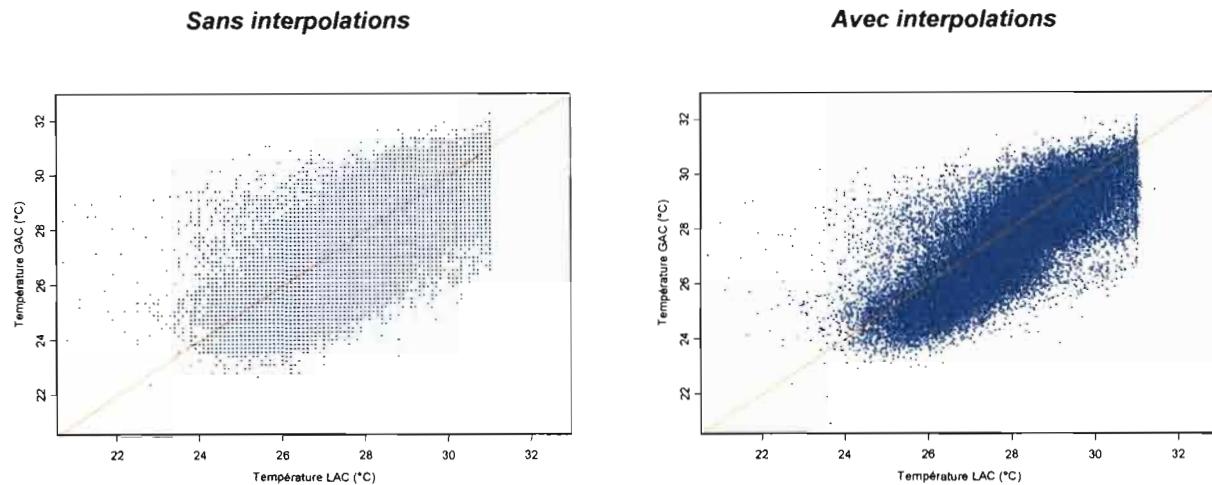


Figure 28: Températures GAC en fonction des températures LAC, avec et sans interpolations

La moyenne des écarts (en valeurs absolues) entre les données LAC et GAC est de 0,76°C pour les données interpolées, contre 0,80°C sans interpolation. De même, on trouve un coefficient de corrélation de Spearman légèrement plus élevé avec les données interpolées (0,816 contre 0,804), ainsi qu'une erreur type résiduelle plus faible après application d'un modèle linéaire.

Si l'on admet que les interpolations améliorent la description de l'environnement plus souvent qu'elles ne la détériorent⁸⁶, il reste encore à déterminer si cela vaut la peine par rapport à la précision des données. Les figures ci-dessous comparent les valeurs avec et sans interpolation dans le cas de la température (une image par jour avec des pixels d'une minute d'angle) et de l'anomalie de hauteur de l'eau (une image par période de 10 jours avec des pixels de 15 minutes d'angle). Ce sont les deux extrêmes du spectre de résolution de nos données. Les valeurs affichées sont les valeurs environnementales aux positions prospectées seulement.

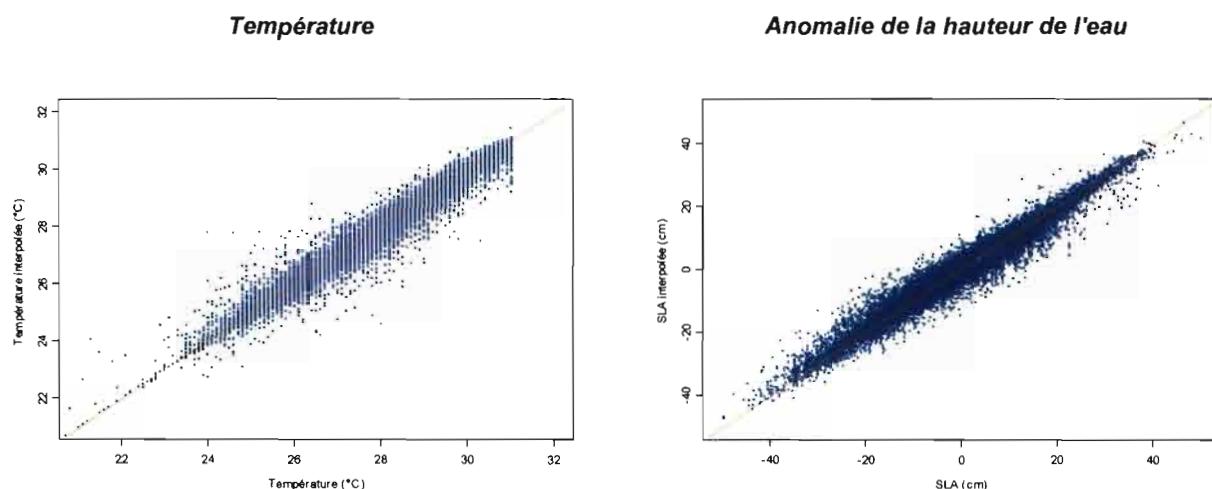


Figure 29: Comparaison des valeurs avec et sans interpolation

Les lignes oranges de la figure 29 représentent des droites de pente 1. Les différences entre les valeurs avec et sans interpolation (les distances verticales à la droite) se distribuent selon une loi centrée semblable à une gaussienne. Environ 85% des différences sont inférieures à la précision de ces deux types de données

⁸⁶ Cette hypothèse appliquée à l'ensemble des données environnementales est une extrapolation de la figure 28, qui ne concernait que la température aux positions et aux dates des prospections.

(0,2°C et 2 cm respectivement) ; ces ajustements risquent d'être dénués de valeur réelle. Le 15% restant se trouve dans des régions à plus forts gradients spatiaux ou temporels.

La précision des coordonnées de prospection et des données environnementales, lorsque l'on prend en compte ces deux contraintes simultanément, suggère que les interpolations ne se justifient guère plus que dans 3% des cas. Malgré ces limitations, les résultats exposés dans le reste de ce chapitre montrent que les interpolations améliorent quand même légèrement la plupart des corrélations. La figure ci-dessous illustre les améliorations qu'elles apportent aux corrélations de Spearman de la figure 35 (page 102). Les gains vont jusqu'à 17%, en excluant les corrélations non significatives.

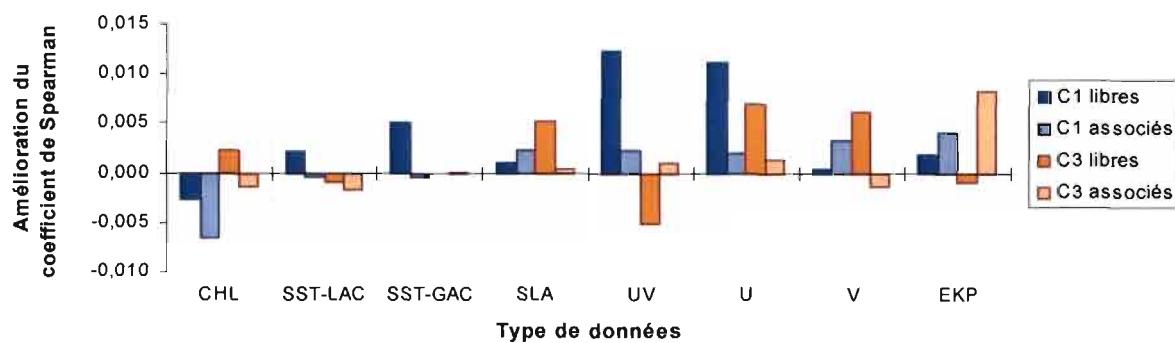


Figure 30: Améliorations des corrélations de Spearman apportés par les interpolations

4.2 Distribution des données

Environ 60% des positions prospectées ont donné lieu à au moins un coup de senne. La figure 31 donne la composition moyenne approximative des coups de sennes, incluant les coups nuls (25% des cas). Cette figure est approximative du fait que, contrairement à la figure 26 (page 90), elle ne prend en compte que les enregistrements pour lesquels le caractère « libre » ou « associé » des bancs était connu. Cet état étant plus rarement spécifié dans le cas des coups nuls (21% des cas) que dans le cas des calées (79% des cas), les proportions calculées s'en trouvent artificiellement augmentées. Les valeurs de la figure 31 ont subi une correction visant à compenser cette sous-représentation des coups nuls afin d'obtenir des valeurs plus proches de celles de la figure 26. La correction consiste à supposer que lorsque la composition des coups nuls n'est pas spécifiée, la proportion moyenne des composantes est la même que lorsqu'il y a eu des calées.

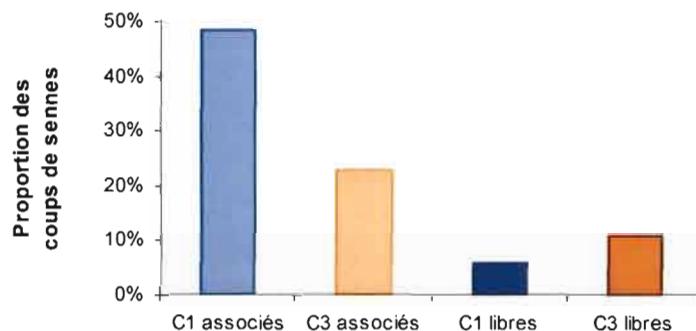


Figure 31: Proportions des coups de sennes contenant les catégories étudiées

Dans une calée, la quantité de poissons suit une distribution exponentielle (figure 32). On remarque que les bancs libres capturés contiennent plus souvent de gros poissons (C3) que de petits poissons (C1), alors que les bancs associés ont le rapport inverse. Cette remarque est conforme aux observations rapportées dans la littérature (section 1.4.4 à la page 27).

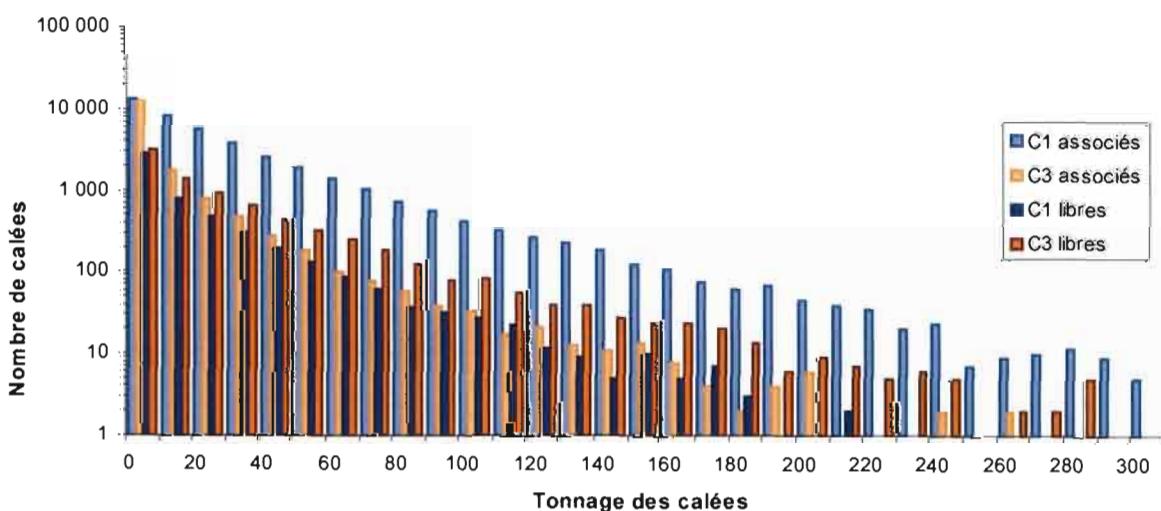


Figure 32: Distribution du tonnage des calées

NOTE: Pour estimer le nombre de calées en ordonnée, nous avons compté le nombre d'enregistrements plutôt que de faire la somme de la colonne « nCalées » de la base de données (section 3.5.3 à la page 78). La différence entre les deux méthodes est d'environ 5%. La raison pour laquelle nous n'avons pas fait la somme réside dans le fait que, lorsqu'un seul enregistrement représente deux calées (ou plus), nous ne savons pas comment répartir les tonnages entre ces deux calées. Nous avons supposé que les deux calées se sont faites sur le même banc et les considérons comme si la totalité du banc avait été capturé en une seule calée.

4.2.1 Distribution géographique des données de prospection

La figure 30 illustre le nombre de données de prospection dans chaque carré d'un demi-degré d'angle, toutes saisons confondues, sur les dix années de données disponibles. La région couverte va de 35°S à 30°N, et 30°E à 90°E. Cette figure inclut les bancs libres et associés.

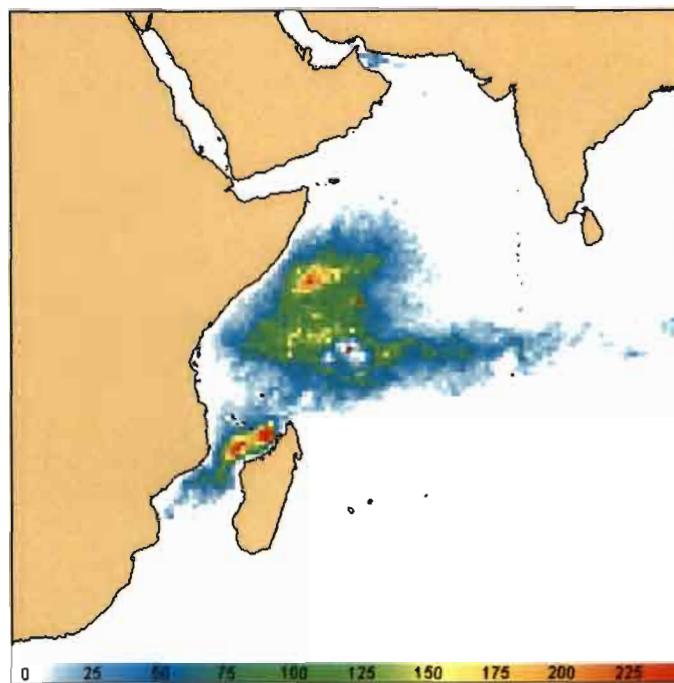


Figure 33: Nombre de données de prospection par demi-degré d'angle

Environ 34% de la surface océanique de cette zone a reçu la visite d'au moins une prospection. On reconnaît les zones de pêches privilégiées : le nord du canal du Mozambique, les eaux entre la Somalie et les Seychelles ainsi que la bande équatoriale.

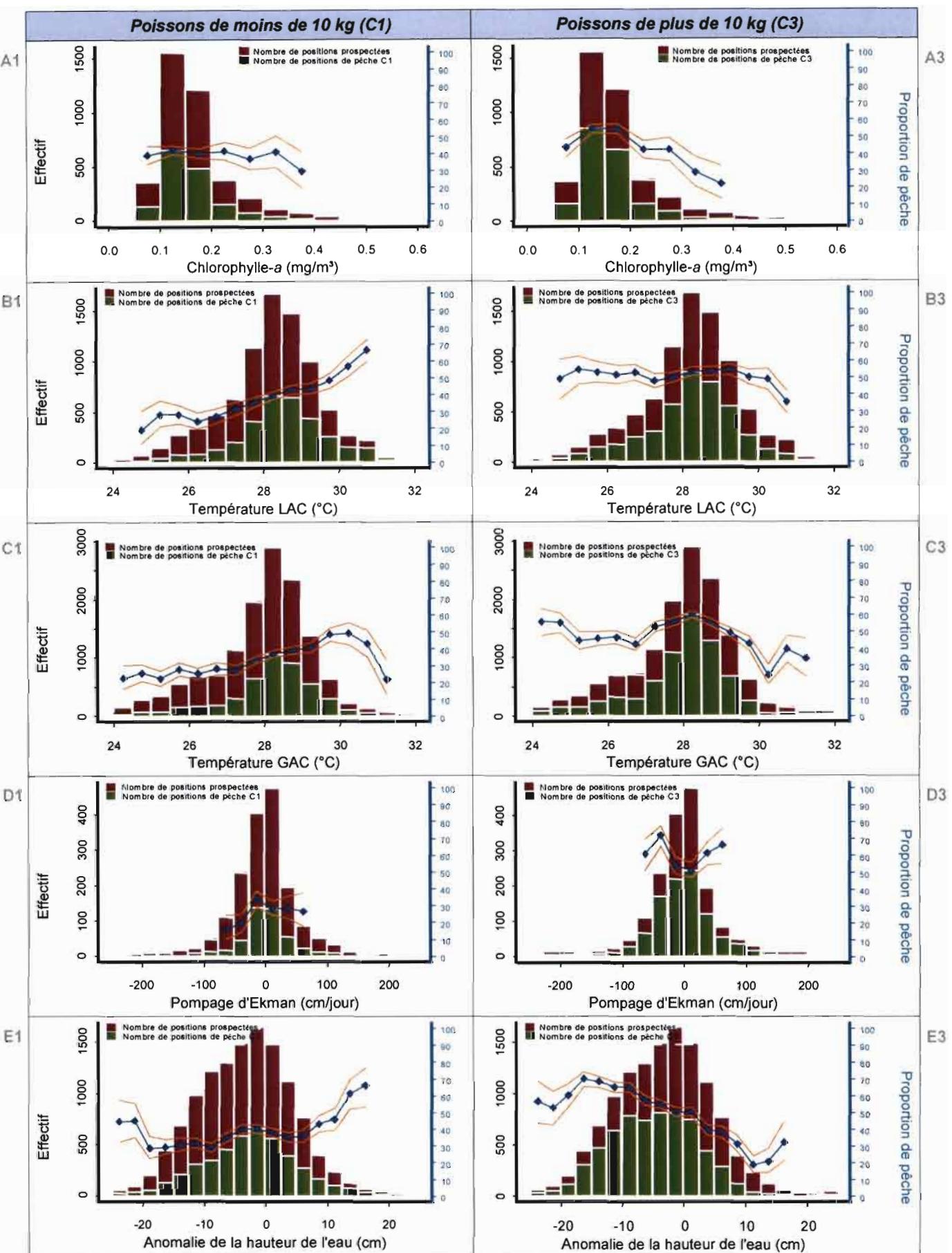
4.2.2 Distribution des données environnementales aux positions de prospection

Les valeurs de chaque paramètre environnemental (température, chlorophylle-*a*, etc.) ont été évaluées par la méthode evaluate (page 64) à toutes les positions prospectées, et leurs distributions illustrées dans la figure 34. Les barres bourgognes de ces histogrammes donnent le nombre de positions prospectées, tandis que les barres vertes donnent le nombre de ces positions où il y a eu au moins une calée. Seuls les bancs libres ont été pris en compte. Les coups nuls ont été ignorés, faute d'informations adéquates dans la base de données (section 3.5.3, page 80). Les lignes bleues représentent le rapport entre les barres vertes et les barres bourgognes des histogrammes. Elles correspondent à l'axe « proportion de pêche » et peuvent être interprétées comme une sorte de « rendement » moyen sous différentes conditions environnementales, en supposant que la pêche n'a pas diminué significativement l'abondance. Les lignes oranges représentent l'intervalle de confiance⁸⁷ des lignes bleues, pour un risque d'erreur de 5%. Ces lignes ne sont pas tracées pour les plages de valeurs contenant moins de 60 observations, car l'intervalle de confiance devient alors trop grand.

⁸⁷ La proportion ayant été calculée sur des données appartenant à un échantillon, les résultats obtenus diffèrent de ceux que l'on aurait obtenus en considérant la population entière. Lorsque l'effectif de l'échantillon est élevé, l'échantillonnage de la proportion suit une distribution normale avec un écart type estimé par :

$$s_p = \sqrt{\frac{p(1-p)}{n-1}} \quad \text{où } n \text{ est le nombre de positions prospectées et } p \text{ la proportion de pêche (entre 0 et 1).}$$

Cette équation ne s'applique que de façon approximative, puisque la population de thon n'est pas infinie et que l'échantillonnage ne s'est pas effectué avec remise (les poissons n'ont pas été remis à l'eau).



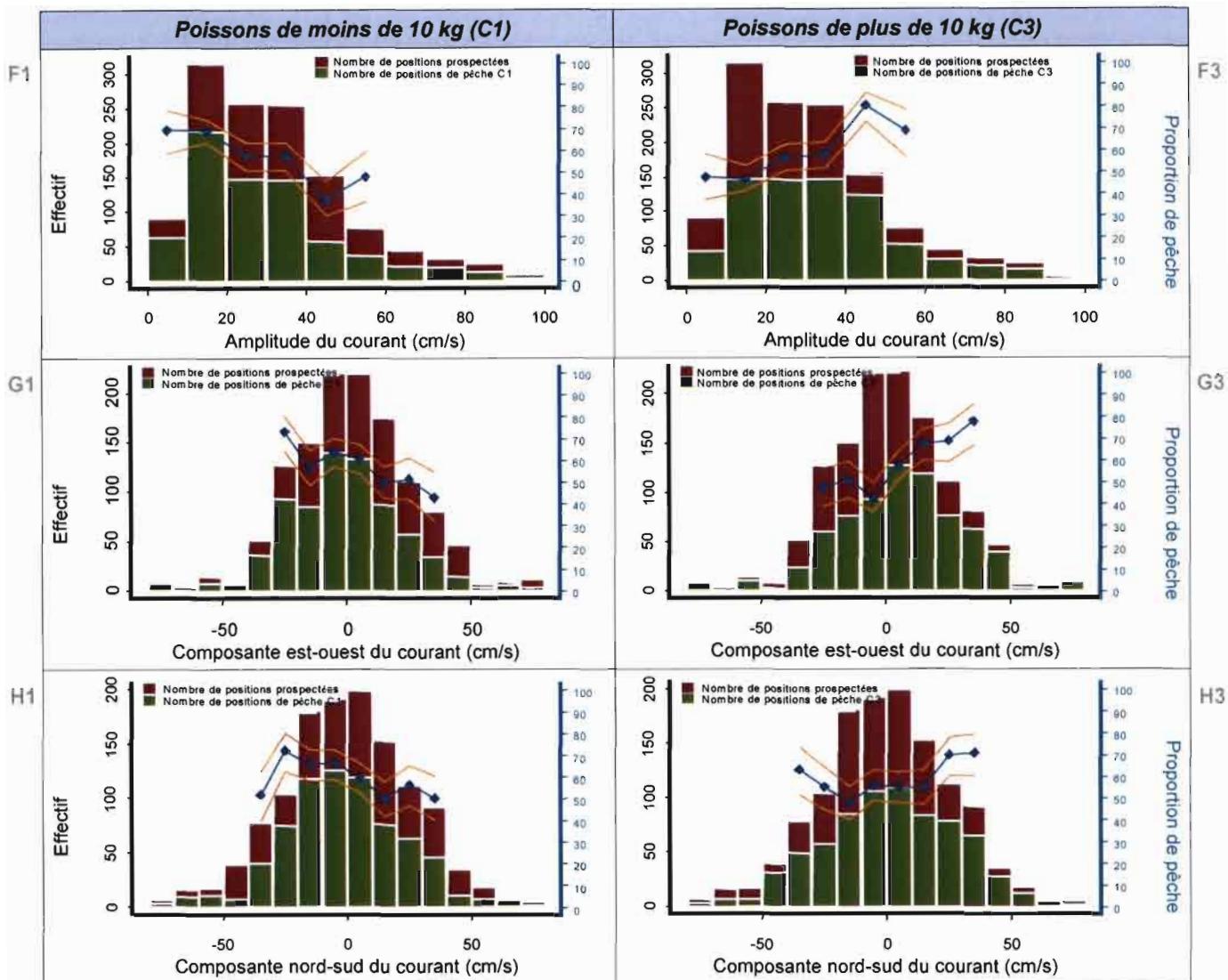


Figure 34: Distribution des paramètres environnementaux des bancs libres le jour des prospections

Les histogrammes de la figure 34 n'exploitent que les données environnementales du jour de la pêche. L'introduction d'un décalage de quelques jours produit des histogrammes de même aspect, qui n'ont pas été représentés ici. Ces graphiques ont l'apparence de distributions normales, sauf la chlorophylle- α , l'amplitude du courant géostrophique et l'amplitude de tous les gradients (non-illustrées) qui ont l'apparence de distributions log-normales. Les modèles linéaires du chapitre 4.4 utiliseront donc les logarithmes de ces dernières variables.

Les concentrations en chlorophylle- α suivent une distribution à peu près log-normale, tout comme plusieurs autres types de concentrations tels que les polluants atmosphériques. D'ailleurs, les données fournies par la Nasa étaient déjà sous forme de logarithme de la chlorophylle. Néanmoins, les opérations sur les images (tableau 14 à la page 81) n'ont pas été appliquées sur les logarithmes des concentrations ; elles ont été appliquées sur les valeurs de concentration elles-mêmes. Dans une étape ultérieure, il conviendrait d'effectuer un test statistique pour comparer ces deux approches.

Dans le cas de l'amplitude du courant géostrophique, on aurait pu s'attendre à une distribution semblable à celle de χ^2 à deux degrés de liberté. Par définition, χ^2 est la somme des carrés de variables aléatoires normales centrées réduites et indépendantes. Si les composantes U et V du courant géostrophique ont une distribution normale et si Z_0 et Z_1 sont les variables centrées réduites de U et V (après une simple soustraction de la moyenne et division par l'écart-type), alors $Z_0^2 + Z_1^2$ aurait pu suivre une loi de χ^2 . Toutefois, les composantes du courant géostrophique ne sont pas indépendantes. Ces courants sont calculés à partir d'un scalaire, l'anomalie de la hauteur de l'eau, et sont souvent associés à des tourbillons dans lesquels U et V varient ensemble. Il en résulte une distribution qui s'éloigne de χ^2 pour ressembler davantage à une log-normale.

L'amplitude du gradient se calcule d'une manière similaire à celle du courant géostrophique : on calcule les composantes nord-sud et est-ouest du gradient, puis on extrait le module à l'aide de la relation de Pythagore. Les gradients sont souvent associés à des tourbillons, et la distribution de leurs amplitudes ressemble aussi à une log-normale.

Aucune des distributions de la figure 34 ne correspond réellement au modèle théorique d'une distribution normale ou log-normale ; le test de χ^2 retourne une probabilité de zéro (aux erreurs d'arrondissement près) pour tous nos descripteurs.

Parmi les courbes de proportion de pêches de la figure 34, certaines affichent une tendance claire à augmenter ou diminuer sur quasiment toute la largeur de l'axe des x (par exemple **B1**, **E3**, **G1**, **G3**). Les maximums locaux au milieu de la distribution (comme dans la figure 34-**C3**) ne sont pas très prononcés. Ces tendances rendent plausible la pertinence des corrélations de Spearman (section 4.3) et des modèles linéaires (section 4.4). Des interprétations océanographiques de la figure 34 seront proposées après les corrélations de Spearman (figure 35).

4.3 Corrélations entre paramètres environnementaux et captures

Les graphiques des sections précédentes suggèrent que la probabilité de capturer des thons varie en fonction des paramètres environnementaux. Certaines corrélations apparentes peuvent n'être dues qu'à la quantité de variations que l'on observerait si les valeurs avaient été tirées au hasard. Pour vérifier si les tendances observées sont significatives ou pas, nous utiliserons un test non-paramétrique. Le test paramétrique de Pearson n'a pas été utilisé car il suppose que les valeurs X et Y sont échantillonnées à partir de populations qui suivent une distribution normale (gaussienne), au moins approximativement. Or, ce n'est pas le cas de nos données de pêches qui suivent soit une loi exponentielle (quantités pêchées, figure 32), ou soit une distribution binomiale (présence ou absence de pêches). Nous avons plutôt utilisé le test de Spearman, qui est basé sur les rangs des variables X et Y plutôt que leurs valeurs. Ce test peut être un meilleur indicateur que le coefficient de Pearson lorsque la relation entre deux variables est non-linéaire. Il nous dispense d'effectuer certaines transformations de variables (par exemple le logarithme des valeurs de chlorophylle), puisqu'elles ne changent rien au rang des valeurs.

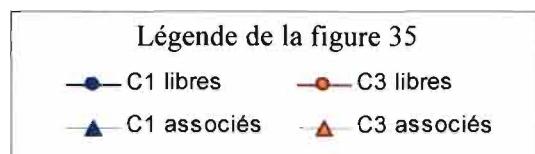
Le classement d'une liste ne contenant que des 0 et des 1 soulève le problème des *ex-aequo* : on ne peut pas ordonner toutes les valeurs d'absence (0) entre elles par exemple. Dans ces conditions, il faut parfois introduire un facteur de correction à l'équation de corrélation (Scherrer, 1984). Une autre approche est de donner le même rang à toutes les valeurs *ex-aequo*. Cette approche a été appliquée dans cette thèse sur trois jeux de données différents :

1. Sur les valeurs d'absence ou de présence (0 ou 1).
2. Comme l'approche précédente, mais en remplaçant les valeurs de présence (1) par les quantités pêchées (> 0). Ainsi, la suite des valeurs de présence reste ordonnée. Toutefois, cette approche n'évite les *ex-aequo* (autres que fortuits) que pour environ la moitié des données, puisque pour toutes les prospections sans pêche la quantité reste exactement égale à 0 tonne⁸⁸.
3. Comme l'approche précédente, mais en n'utilisant que les données où il y a eu pêche. En éliminant les valeurs de prospection (c'est-à-dire tous les zéros), les *ex-aequo* restants sont peu nombreux et dus au hasard. Cette approche n'utilise que la moitié des données des deux premières approches.

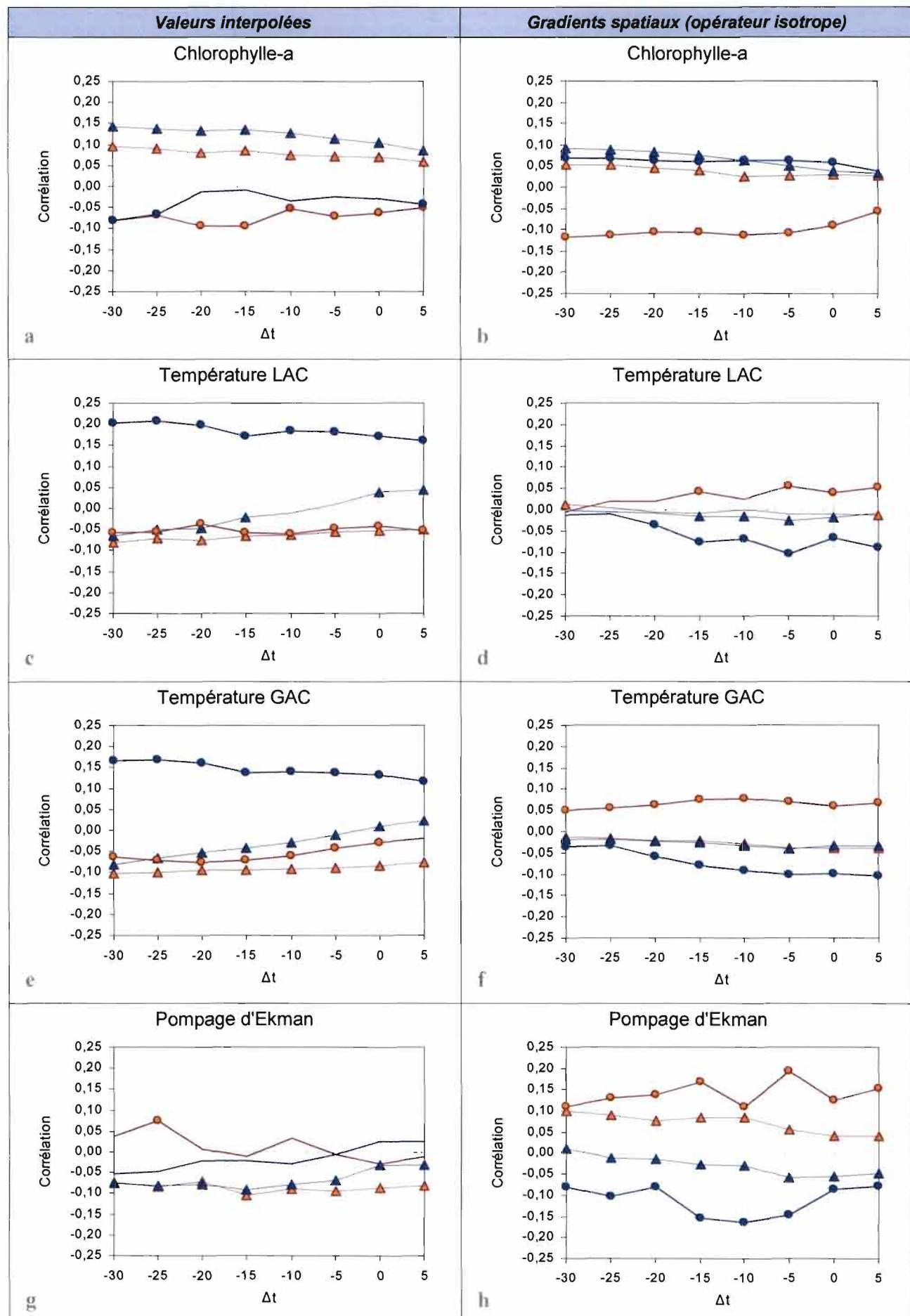
Les graphiques des corrélations de Spearman selon les méthodes 1 et 2 sont pratiquement identiques. Cela suggère que l'utilisation des quantités pêchées plutôt que simplement la valeur 1 pour les présences n'introduit pas beaucoup plus d'information que du bruit. Cette hypothèse est confortée par les corrélations de Spearman obtenues selon la méthode 3, qui sont proches de zéro. Pour les corrélations illustrées dans la figure 35, nous avons retenu la méthode 2. Ces corrélations sont donc les seuls tests statistiques de cette thèse utilisant les quantités pêchées plutôt que la présence ou l'absence de captures, mais nous avons vu que cela ne change que peu de choses. Les corrélations calculées sont cohérentes avec les tendances observées dans les « proportions de pêches » de la figure 34.

Les figures 35 reportent en abscisse l'écart de temps (Δt) entre le jour de la prospection et celui de la donnée environnementale. Elles illustrent donc la variation de la corrélation en fonction de l'écart de temps, qui varie de 30 jours avant la prospection jusqu'à 5 jours après. La décision d'examiner les valeurs environnementales après la prospection s'appuie sur l'hypothèse que les phénomènes océanographiques détectés par satellites ne sont pas nécessairement en phase avec les signaux perçus par le thon. Par exemple une remontée de la thermocline ne se manifestera sur une image de température que lorsqu'elle atteint la surface, alors que le thon vivant sous la surface peut l'avoir perçue avant.

Toutes les figures utilisent la légende ci-dessous. Une marque absente indique qu'il y a au moins 1% de risque que la corrélation correspondante ne soit pas significative.



⁸⁸ La prospection ne peut être confondue avec les coups nuls, du fait que nous excluons ces dernières de l'analyse pour les raisons exposées à la section 3.5.3 (page 80).



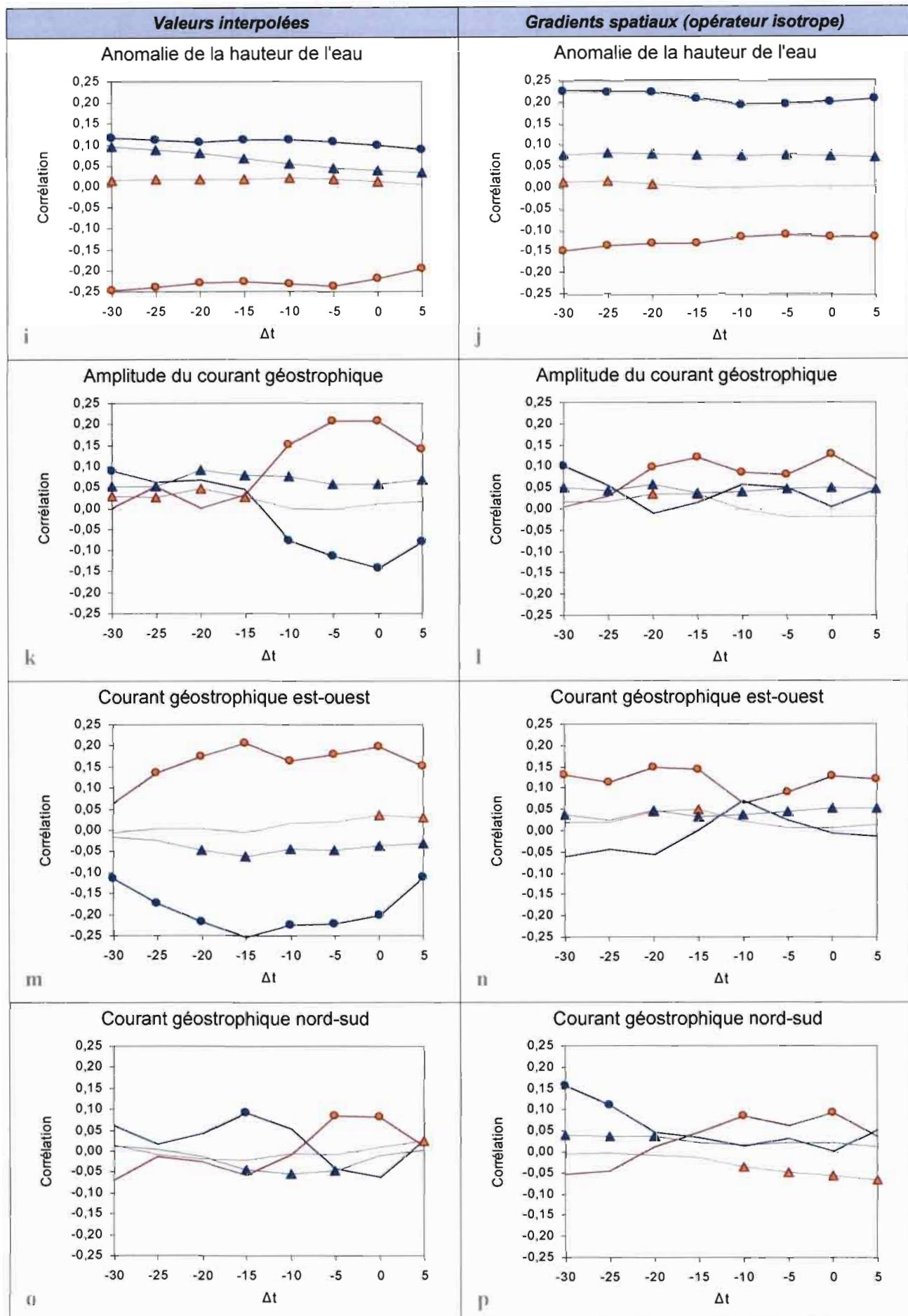


Figure 35: Corrélations de Spearman entre les captures et différents paramètres environnementaux

Les figures 34 et 35 sont complémentaires. La figure 34 illustre la distribution des descripteurs et permet de vérifier si la « proportion de pêche » passe par un extremum local pour certaines valeurs du dit descripteur. De tels extrema diminuent la corrélation de Spearman ; une tendance qui aurait la forme d'une cloche symétrique donnerait une corrélation nulle. En revanche, la figure 35 synthétise des informations qui nécessiteraient un trop grand nombre d'histogrammes dans la figure 34 : ces derniers n'incluent pas d'information sur les gradients (colonne de droite de la figure 35), ne concernent que les bancs libres et n'utilisent que les données du jour de la pêche. De l'observation des corrélations de la figure 35, complétées avec les « proportions de pêches » de la figure 34, on en tire les remarques suivantes (les références commençant par une lettre majuscule, telles que « A1 », se réfèrent à la figure 34 tandis que les références écrites en minuscules, telles que « a », se réfèrent à la figure 35) :

Remarques générales

- Le succès des pêches est généralement corrélé aux descripteurs du paysage océanique. Cette corrélation est plus forte pour les bancs libres que pour les bancs associés, à l'exception notable des concentrations en chlorophylle-*a* (**a**). Dans le cas particulier des gros poissons (plus de 10 kg), la température est une autre exception (**e**).
- Les bancs libres de petits poissons (moins de 10 kg) et des gros poissons ont rarement des corrélations significatives de même signe. Leurs corrélations avec les descripteurs sont de signe opposé dans le cas de la température (**e**), de l'anomalie de la hauteur de l'eau (**i**), de la composante est-ouest du courant géostrophique (**m**) et pour tous les gradients excepté ceux du courant géostrophique (**b, d, f, h, j**).

Concentrations en chlorophylle-*a* et température

- La probabilité de capturer des bancs associés (quel que soit le poids des individus) augmente avec les concentrations en chlorophylle-*a* (**a**). Les bancs libres ont la tendance opposée, mais de manière moins significative. Si les objets flottants se retrouvent plus souvent dans des zones de convergences, on peut supposer que cette même convergence a aussi contribué à l'accumulation de plancton passif (section 1.4.4 à la page 26). En l'absence de dispositif agrégatif, le thon peut préférer des zones à plus faibles concentrations puisqu'il chasse à vue.
- Les corrélations avec la chlorophylle-*a* et la température sont un peu plus prononcées si l'on utilise les valeurs au moins 15 jours avant la prospection (**a, e**). Cette tendance peut s'expliquer s'il faut en moyenne le même délai avant qu'un accroissement de la population locale de phytoplancton ne soit suivi d'un accroissement de la population locale de zooplancton, puis des proies des thons (Stretta, 1991), à la condition qu'il n'y ait pas de transport net de la production secondaire et tertiaire entre la zone de pêche et les régions avoisinantes. Cette dernière condition est rarement respectée, ce qui explique peut-être en partie pourquoi la tendance observée n'est pas plus prononcée.
- La probabilité de capturer des bancs libres de petits poissons (moins de 10 kg) augmente à mesure que la température de l'eau s'accroît entre 26 et 30°C (**C1**). Cette corrélation n'est pas due à un effet saisonnier (un accroissement des pêches en été), puisque les pêches sont bien réparties à toutes époques de l'année.
- Les bancs libres de gros poissons semblent préférer les eaux froides ou la proximité de fronts thermiques, tandis que les petits poissons semblent préférer les eaux chaudes ou de température plus uniforme (**e, f**). Les bancs associés ont un comportement qui se rapproche de celui des bancs de petits poissons.
- La température LAC (*Local Area Coverage*) semble légèrement plus significative que la température GAC (*Global Area Coverage*) dans le cas des petits poissons (**c, e**). Cet avantage est moins clair, voire contredit dans le cas des gros poissons (figure 36 et section 4.4.3.4).

Anomalies de la hauteur de l'eau et courants géostrophiques

- La probabilité de capturer des bancs libres de gros poissons (plus de 10 kg) est plus élevée lorsque l'anomalie de la hauteur de l'eau est négative (-10 cm) plutôt que positive (+10 cm) (**i, C3**). Les anomalies négatives correspondent souvent à des tourbillons cycloniques. Ils sont le siège de divergences qui peuvent favoriser une remontée de la thermocline. Cette remontée, si elle se maintient

suffisamment longtemps, peut favoriser un accroissement local de la production primaire et secondaire. La physiologie des gros thons leur permettant de passer plus de temps en eau froide que les petits (section 1.4.1, page 25), il est possible qu'ils soient plus aptes à exploiter cette production.

- Les corrélations des bancs libres (petits ou gros poissons) avec la température de surface d'une part (e), et l'anomalie de hauteur de l'eau d'autre part (i), sont cohérentes avec l'hypothèse des tourbillons cycloniques favorisant un refroidissement.
- Pour les gros poissons, le paramètre le plus corrélé est l'anomalie de la hauteur de l'eau (i). Pour les petits poissons, le paramètre le plus corrélé est le gradient de cette hauteur (j).
- Les corrélations sont relativement stables dans le temps, excepté celles entre les bancs libres et l'amplitude du courant géostrophique. Cette dernière est la seule à afficher un maximum évident proche du jour de la pêche, et diminue rapidement jusqu'à devenir non significative pour un écart de temps supérieur à 10 jours (k). Il est possible que les courants aient un effet plus immédiat sur le comportement du poisson, puisqu'ils affectent sa nage. Par exemple les gros poissons profitent peut-être des courants pour accroître leurs déplacements.

Ces corrélations de Spearman permettent d'effectuer un certain tri entre les paramètres, ainsi qu'entre les opérateurs appliqués sur les paramètres. Dans le cas de la température de surface, la figure 36 compare les coefficients de corrélations des données LAC et GAC, qui diffèrent principalement par leur résolution spatio-temporelle. Les données les plus fines (LAC) se révèlent les plus avantageuses dans le cas des bancs libres de petits poissons, mais l'avantage est moins clair pour les autres catégories. Dans le cas des bancs libres de gros poissons, les données GAC semblent légèrement plus avantageuses. Les modèles linéaires présentés dans la section 4.4.3.4 appuient cette observation.

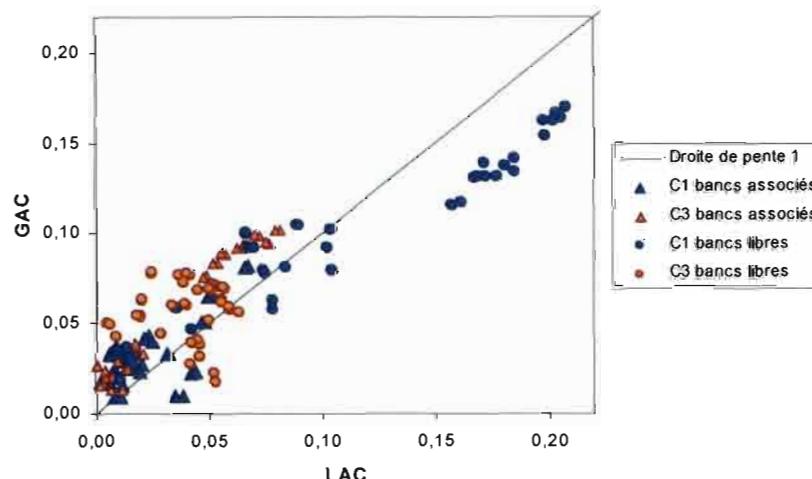


Figure 36: Comparaison des corrélations de Spearman entre températures LAC et GAC

4.4 Analyses factorielles discriminantes

Les corrélations du chapitre précédent mesuraient les degrés de liaison entre deux groupes de variables : les captures de différentes catégories de thons et les descripteurs du paysage océanique. Les analyses de ce chapitre chercheront à prédire les présences ou absences de thons capturables (variable à expliquer y) en fonction des descripteurs (variable explicative x). Lorsque les variables x et y sont toutes deux continues, la méthode d'analyse habituelle est une *régression multiple*. Lorsque la variable y est nominale (dans notre cas « présence » ou « absence »), l'analyse est analogue à la régression multiple et se dénomme *analyse factorielle discriminante*. Mais dans le cas particulier où la réponse y ne comprend que deux classes, des simplifications apparaissent et l'analyse factorielle devient équivalente à la régression multiple (Lebart et al., 1995).

La régression linéaire est un cas particulier du modèle linéaire statistique. Le terme linéaire n'est pas associé à l'idée d'une droite rectiligne, mais plutôt à celle de combinaisons linéaires de paramètres. Par exemple dans l'équation suivante :

$$\hat{y} = c_0 + c_1 x + c_2 x^2$$

La combinaison des paramètres c_0 , c_1 , et c_2 est linéaire, même si la relation entre \hat{y} et x est quadratique. Les observations y ne comprennent que deux catégories, que l'on représentent par les valeurs +1 ou -1 selon que des thons capturables aient été aperçus ou pas. Les valeurs prédictes \hat{y} sont continues, le plus souvent (mais pas obligatoirement) dans la plage -1 à +1. Les valeurs positives seront considérées comme une prédiction de présence de thons capturables, tandis que les valeurs négatives seront interprétées comme une prédiction d'absence.

Affecter à y les valeurs -1 et +1 plutôt que 0 et +1 réduit le risque de confusion avec les probabilités : une valeur prédictée \hat{y} de +1 ne représente pas une certitude (probabilité de 100%) de présence de poisson. De même, une échelle de 0 à 1 aurait rendu plus surprenantes en apparence les valeurs négatives de \hat{y} , lorsque les prédictions du modèle sortent de la plage de valeurs de y . Pour transformer les sorties d'un modèle linéaire en probabilités, il faudra utiliser une table de conversion dérivée de tests empiriques comme la figure 52 (page 131).

Puisque plusieurs descripteurs quantitatifs sont utilisés dans le modèle, on parle de régression multiple. Les résultats de la régression multiple sont instables si des variables explicatives sont fortement corrélées (Lebart et al., 1995). Aussi, nous tenterons d'améliorer la qualité des modèles en partant d'un jeu de descripteurs limité, puis en supprimant encore autant de descripteurs que possible.

4.4.1 Choix des descripteurs du paysage océanique (variables explicatives)

Les descripteurs du paysage océanique ont été définis dans la section 4.1.2 (page 91). Nous avons calculé les valeurs de 576 descripteurs, construits à partir de 8 paramètres utilisant 9 opérations à 8 positions relatives. Chaque descripteur contient entre 10 000 et 120 000 données, à l'exception de ceux qui sont dérivés du pompage d'Ekman (EKP) ainsi que les composantes est-ouest (U) et nord-sud (V) du courant géostrophique. Ces derniers ont été écartés à cause du faible nombre de données disponibles.

Cet ensemble de 576 descripteurs effleure à peine les possibilités; un nombre considérablement plus grand de descripteurs auraient pu être construits. Il ne sera généralement pas possible de les donner tous à un modèle linéaire ou à un réseau de neurones artificiels. En outre, certaines opérations sont très similaires (par exemple les différents opérateurs de gradients du tableau 14, page 81). Une forme de sélection préalable est requise, et les corrélations de Spearman (figure 35) constituent un des outils que nous avons utilisés à cette fin. Cette présélection a été effectuée sur la base de quatre critères :

- Le descripteur reste relativement proche de la prospection (nous évitons les descripteurs remontant à plus d'un mois, même s'ils semblent encore significatifs).
- Le descripteur est l'un des plus significatifs au sein d'un groupe de descripteurs fortement corrélés (par exemple les données à -5, -10 et -15 jours).
- Le descripteur partage un maximum de caractéristiques communes avec les autres descripteurs retenus. Par exemple si interpoler les valeurs améliore les corrélations dans le cas de la plupart des

paramètres environnementaux excepté la chlorophylle-*a*, nous interpolons quand même ce dernier dans un souci d'uniformité. Ce critère a pour but de faciliter les interprétations.

- Pour chaque paramètre, on retient au moins deux descripteurs à des positions temporelles relatives différentes. On donne ainsi une opportunité au modèle linéaire de prendre en compte les gradients temporels, que nous savons importants.

Le tableau 17 énumère les paramètres ainsi que les modes de calcul des gradients retenus pour ces paramètres. La colonne « symboles » donne quelques exemples de descripteurs, avec les étiquettes qui seront utilisées dans les tableaux 19 et suivants. Le nombre en indice indique le décalage temporel, en nombre de jours. Les interpolations dans l'espace et dans le temps (section 4.1.3) sont implicites pour tous les descripteurs.

Tableau 17: Exemples de descripteurs retenus pour les modèles linéaires

Type d'opération et paramètre	Opération	Exemples de symboles
Logarithmes des concentrations en chlorophylle- <i>a</i>		CHL CHL ₋₁₅
Températures de surface LAC (Local Area Coverage)		SST ₋₁₅ SST ₋₂₅
Températures de surface GAC (Global Area Coverage)		SST _{G-15} SST _{G-25}
Anomalie de hauteur de l'eau		SLA SLA ₋₃₀
Logarithmes des gradients de chlorophylle	Sobel 9×9	∇CHL ₋₁₅ ∇CHL ₋₃₀
Logarithmes des gradients de température LAC	Sobel 5×5	∇SST ∇SST ₋₁₅
Logarithmes des gradients de température GAC	Sobel 3×3	∇SST _G ∇SST _{G-15}
Logarithmes des gradients d'anomalies de hauteur de l'eau	Sobel 5×5	∇SLA ∇SLA ₋₃₀

Exemple : dans le cas des gradients de température LAC, les gradients sont calculés en utilisant un opérateur similaire à Sobel mais de taille 5×5 (tableau 8 à la page 68), les logarithmes seront calculés tels que dans le tableau 18, et les valeurs le jour même et 15 jours avant la pêche seront utilisées. Ces valeurs seront désignées par les symboles ∇SST et ∇SST₋₁₅ respectivement.

4.4.1.1 Transformation des variables ayant une distribution log-normale

Les distributions des amplitudes des gradients spatiaux ont toutes l'apparence de courbes log-normales. Pour les analyses qui vont suivre, nous appliquerons sur ces valeurs les transformations énumérées dans le tableau 18. Dans ce tableau, la variable *m* représente n'importe laquelle des colonnes sobel3, sobel5, sobel7, sobel9, isotrope, prewitt et kirsch de la table Environnements de la base de données (figure 23 à la page 81). Autrement dit, nous appliquons la même transformation quelle que soit la méthode de calcul du gradient spatial.

Tableau 18: Transformations de variables appliquées sur les gradients

Paramètre	Symbol	Transformation
Gradient de température LAC (Local Area Coverage)	∇SST	log(1 + 0,25 × <i>m</i>)
Gradient de température GAC (Global Area Coverage)	∇SST	log(1 + 5,00 × <i>m</i>)
Gradient de concentration en chlorophylle- <i>a</i>	∇CHL	log(1 + 100 × <i>m</i>)
Gradient d'anomalie de hauteur de l'eau	∇SLA	log(1 + 6,25 × <i>m</i>)
Gradient des composantes <i>U</i> et <i>V</i> du courant géostrophique	∇U et ∇V	log(1 + 0,05 × <i>m</i>)
Gradient du pompage d'Ekman	∇EKP	log(1 + 0,20 × <i>m</i>)

Tous les logarithmes du tableau 18 sont en base naturelle. Les valeurs du facteur multiplicatif de *m* ont été déterminées empiriquement, en traçant des courbes de même moyenne et de même écart-type par dessus les histogrammes des valeurs transformées. Un choix était effectué visuellement après l'essai de plusieurs facteurs. Les tests statistiques classiques tels que chi-carré ou Kolmogorov-Smirnov ne nous ont pas été utiles, car ils produisent une probabilité de normalité de zéro (aux erreurs d'arrondissement près) pour presque tous les paramètres.

Outre les gradients spatiaux, les amplitudes du courant géostrophique (et non pas ses composantes *U* et *V*) ont été transformées en logarithmes, ainsi que les concentrations en chlorophylle-*a*.

4.4.1.2 Mode de représentation des gradients temporels

Contrairement aux gradients spatiaux (page 68), le calcul des gradients temporels ne nécessitent pas l'application d'opérateurs sur les images. Ils peuvent se calculer directement à partir des valeurs d'un paramètre à différentes dates. Les modèles linéaires peuvent inclure explicitement un gradient temporel comme suit :

$$\hat{p} = C_0 T_i + C_1 (T_i - T_j) \quad \text{où} \quad \hat{p} \quad \text{est la prédiction de présence ou d'absence de thons ;}$$

C_0 et C_1 sont les coefficients du modèle ;
 T_i et T_j sont les températures aux jours i et j .

ou l'inclure implicitement si on réécrit l'expression précédente comme suit :

$$\hat{p} = C'_0 T_i + C'_1 T_j \quad \text{où} \quad \begin{aligned} C'_0 &= C_0 + C_1 \\ C'_1 &= -C_1 \end{aligned}$$

Tous les modèles linéaires de cette section incluront un gradient temporel exprimé selon la deuxième forme. Par rapport à la première, elle nous évite de choisir *a priori* la variable qui sera remplacée par un gradient (T_j dans l'exemple précédent). Ce choix explique que nous avons retenu pour certains modèles (par exemple la chlorophylle, section 4.4.3.2) un paramètre *a priori* non-significatif.

4.4.2 Choix des données de pêches (variables expliquées)

Les données expliquées seront les présences (+1) ou absences (-1) de thons, en analysant séparément les bancs constitués de petits et de gros individus (catégorie C1 et C3 respectivement, voir figure 26 à la page 90). Notre analyse se concentre principalement sur les plus gros poissons. Les données de prospections des bancs de catégorie C3 ont l'avantage de répartir les présences et absences de manière à peu près égale : les bancs de catégorie C3 sont absents dans $(50 \pm 3)\%$ des données (figure 37). La proportion exacte dépend du modèle linéaire, car le jeu de données sans valeurs manquantes varie légèrement en fonction des descripteurs utilisés. En comparaison, les bancs de catégorie C1 sont absents dans $(63 \pm 3)\%$ de nos données, de sorte qu'un modèle linéaire pourrait affirmer qu'il n'y a de thon nulle part dans l'océan Indien et prétendre jusqu'à 66% de succès dans ses prédictions.

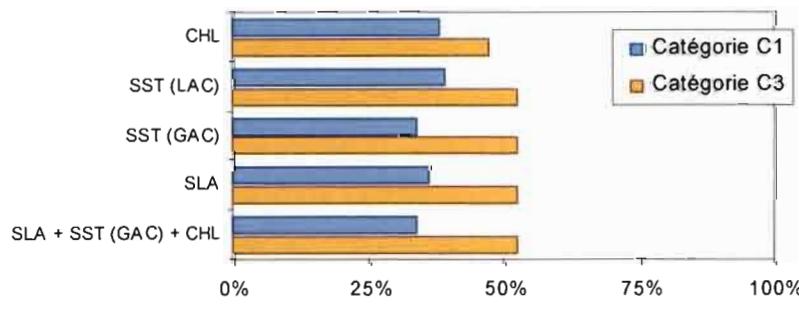


Figure 37: Proportion des présences en fonction des modèles linéaires

Pour tous les modèles linéaires décrits dans les sections qui vont suivre, nous avons systématiquement calculé des prédictions aléatoires ayant la même distribution que le modèle testé. Nous avons ainsi testé que le taux de succès des prédictions aléatoires pour la catégorie C3 restait toujours autour de 50%. Ce seuil de 50% constituera notre référence ; la qualité des modèles linéaires sera mesurée par l'écart entre leurs taux de succès et ce seuil.

4.4.3 Résultats des modèles linéaires pour les thons > 10 kg

Avant de créer un modèle combinant plusieurs paramètres environnementaux, des modèles ont d'abord été testés sur les paramètres pris individuellement. Les pages suivantes donnent les coefficients de ces modèles (tableaux 19 à 28). Dans chaque cas, on a utilisé la totalité des lignes pour lesquelles tous les descripteurs requis par le modèle sont disponibles sans données manquantes. Les descripteurs ont été choisis

en partant d'un modèle comprenant une vingtaine de descripteurs (le tableau 17 donnait quelques exemples), et en retirant progressivement ceux qui apparaissent les moins significatifs. Ce filtrage s'est effectué en examinant à la fois la probabilité que le coefficient ne soit pas significativement différent de zéro (colonne « Probabilité $> |t|$ » des tableaux 19 à 28), et les résultats de la procédure ANOVA de S+. Après chaque retrait, nous avons calculé les prédictions du modèle (telles que définies à la page 105) à chacune des positions prospectées, et les avons comparées avec les observations de présence ou d'absence de pêches. Le pourcentage de prédictions correctes (appelé « taux de succès ») a été estimé, et nous avons cessé de retirer des paramètres au modèle lorsqu'il n'était plus possible de le faire sans provoquer une baisse significative du taux de succès.

Lorsque le modèle comprend une vingtaine de descripteurs, les premiers retraits augmentent souvent le taux de succès. Ces descripteurs apportent manifestement plus de bruit que d'information dans le modèle. Ce dernier ne se met à diminuer que lorsque l'on approche de la dizaine de descripteurs. De légères baisses du taux de succès sont tolérées si elles permettent de retirer plusieurs descripteurs, ou si elle permettent de mieux remplir les critères énumérés dans la section 4.4.1 (par exemple en remplaçant un descripteur rarement utilisé par un autre plus courant). Mais le seuil à partir duquel on ne retire plus de descripteurs est arbitraire. Souvent, nous arrêtons lorsque tout nouveau retrait entraîne une baisse du taux de succès beaucoup plus grande que les retraits précédents. Cette « baisse significative » varie typiquement entre 0,5 et 2%.

Malgré l'aide des outils statistiques tels que ANOVA, la simplification des modèles linéaires comprend une part de tâtonnement. Le résultat peut dépendre du jeu de descripteurs initial, de l'ordre dans lequel on retire des descripteurs, de compromis entre taux de succès et critères de la section 4.4.1, etc. De plus, le jeu de données peut augmenter à chaque retrait d'un descripteur, puisque les risques qu'une ligne contienne au moins une donnée manquante diminuent à mesure que l'on supprime des colonnes. Un léger changement du taux de succès peut provenir de l'insertion de nouvelles données plutôt que d'un apport réel du descripteur retiré au modèle. Aussi n'avons-nous pas figé le jeu de données en éliminant toutes les lignes incomplètes. Nous avons plutôt tenté d'estimer l'intervalle de confiance du taux de succès en fonction du nombre de lignes utilisées pour chaque modèle.

Si l'on représente les prédictions correctes et erronées par 1 et 0 respectivement, le taux de succès est la moyenne de ces nombres. La moyenne ayant été calculée sur des données appartenant à un échantillon, le résultat obtenu (\bar{x}) diffère de celui que l'on aurait obtenu en considérant la population entière (μ). Dans le cas des grands échantillons, l'intervalle pour une probabilité d'erreur de α est :

$$Pr[\bar{x} - \epsilon_\alpha < \mu < \bar{x} + \epsilon_\alpha] = 1 - \alpha \quad \text{où} \quad \epsilon_\alpha \approx z_{\alpha/2} \frac{\sigma_x}{\sqrt{n}} \quad (89)$$

où σ_x est la déviation standard des succès, n est le nombre de données et z est l'aire sous la courbe d'une distribution normale. Pour un risque d'erreur de 5% ($\alpha = 5\%$), on a $z_{\alpha/2} \approx 1,96$ Scherrer, 1984.

Élaborer un modèle à partir de la totalité des données, et vérifier sa qualité en utilisant les mêmes données, s'écarte de l'approche généralement recommandée : ne calculer le modèle qu'à partir d'une fraction des données (par exemple les deux-tiers), et vérifier le taux de succès sur les données restantes. Mais l'objectif de cette thèse n'est pas de prédire le potentiel de pêche de différentes zones géographiques. Nous essayons plutôt d'établir des modèles aussi simples que possible, et d'estimer la pertinence des données altimétriques dans ces modèles. De plus, la différence entre le nombre de descripteurs des modèles considérés (moins de 35) et le nombre de données d'apprentissage (plusieurs milliers) est suffisamment importante pour éviter l'écueil du sur-apprentissage. La séparation des données en un jeu d'apprentissage et un jeu de test ne sera faite que pour le dernier modèle (tableau 28 à la page 130), celui qui est le plus susceptible d'être sujet à un sur-apprentissage du fait du grand nombre de descripteurs (34) pour un nombre plus faible de lignes de données (2612).

4.4.3.1 Apport des positions spatio-temporelles

La figure 33 (page 96) illustre la distribution géographique des prospections. La figure 38 ci-dessous illustre la proportion de prospections dans lesquelles une présence de thon a été observée, pour les catégories C1 et C3 respectivement. On constate que le taux de présence varie géographiquement, avec un

⁸⁹ σ_x/\sqrt{n} est l'estimation de l'écart type de la distribution d'échantillonnage de la moyenne, appelée « erreur type » (« standard error » en anglais).

taux légèrement plus élevé dans le canal du Mozambique ainsi que près de la côte de Somalie. Les données de la figure 38 semblent couvrir une plus grande surface que dans la figure 33 du fait que la couleur de chaque pixel est indépendante du nombre de données ; un fort contraste apparaît dès qu'au moins une donnée est présente. Cette limitation implique aussi que les figures 38 et 39 ne donnent aucune indication sur les intervalles de confiance des taux de présence.

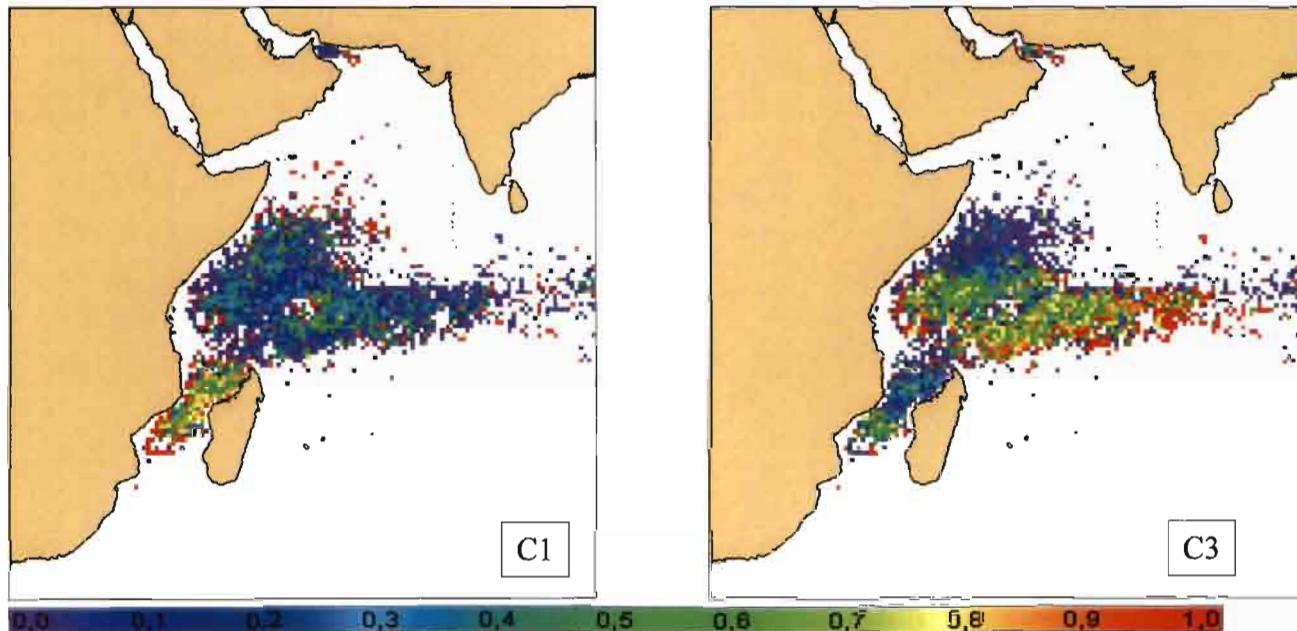
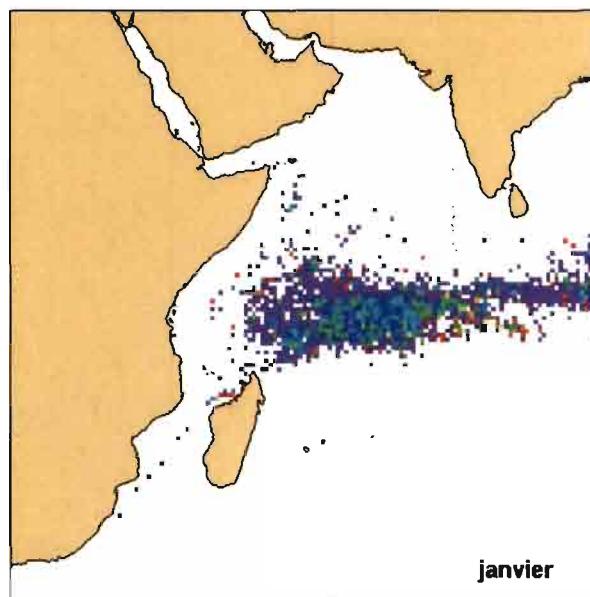


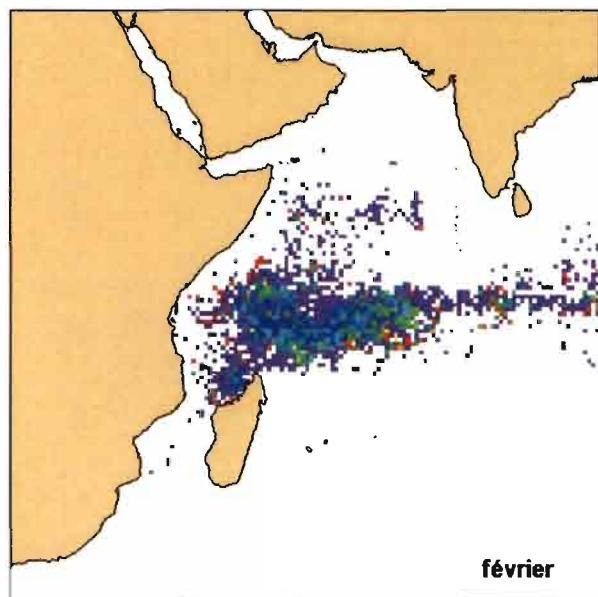
Figure 38: Distribution géographique du taux de présence par demi-degré d'angle (bancs libres)

La figure 39 dans les deux pages suivantes représente la distribution géographique du taux de présence par mois. Chaque mois est moyenné sur toutes les années disponibles. Contrairement à la figure 38, les figures 39 comprennent les bancs associés en plus des bancs libres. Cette cartographie rappelle celles des efforts de pêches que l'on trouve dans les rapports de la CTOI ou de la SFA :

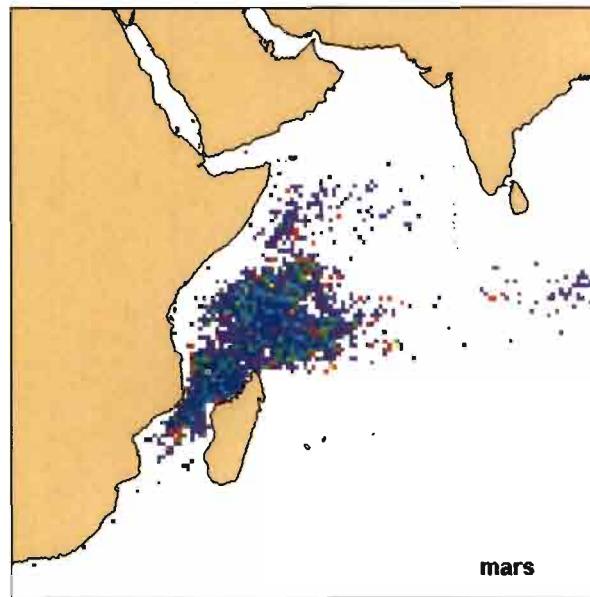
- **janvier – février :** effort concentré le long de l'équateur
- **mars – mai :** incursion dans le canal du Mozambique
- **juillet – septembre :** concentration le long de la côte de Somalie



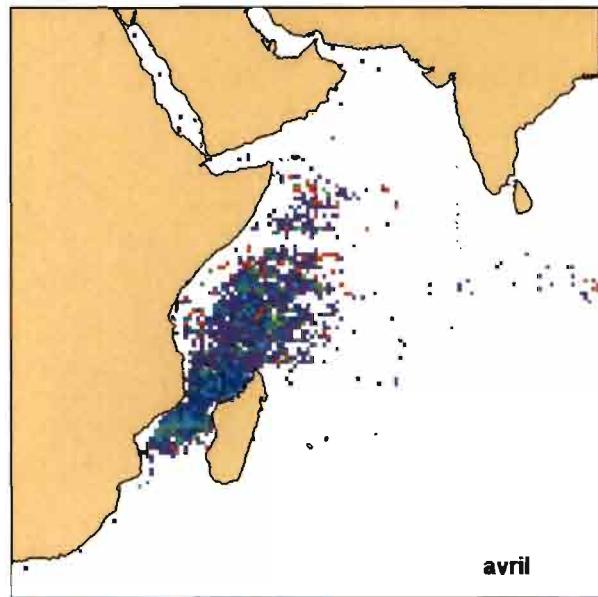
janvier



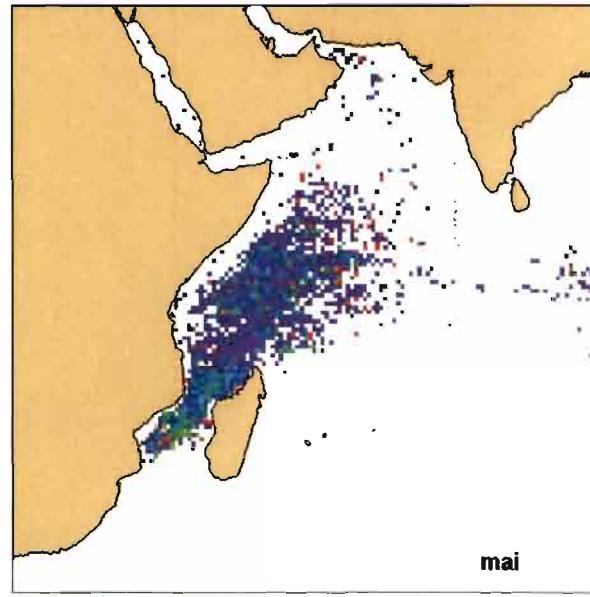
février



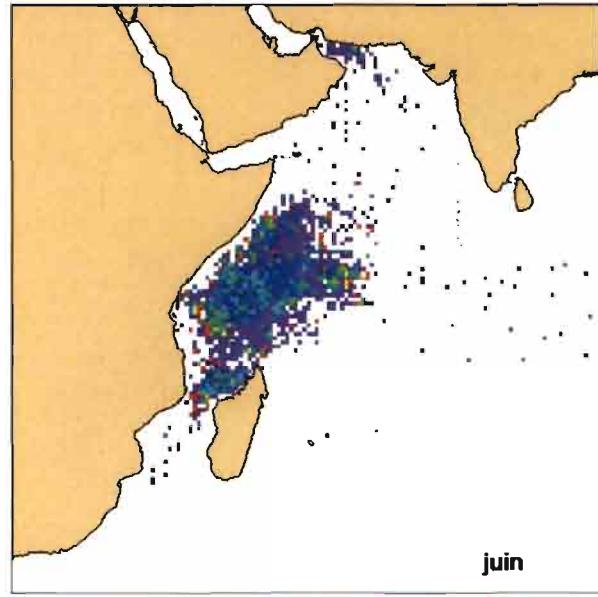
mars



avril



mai



juin



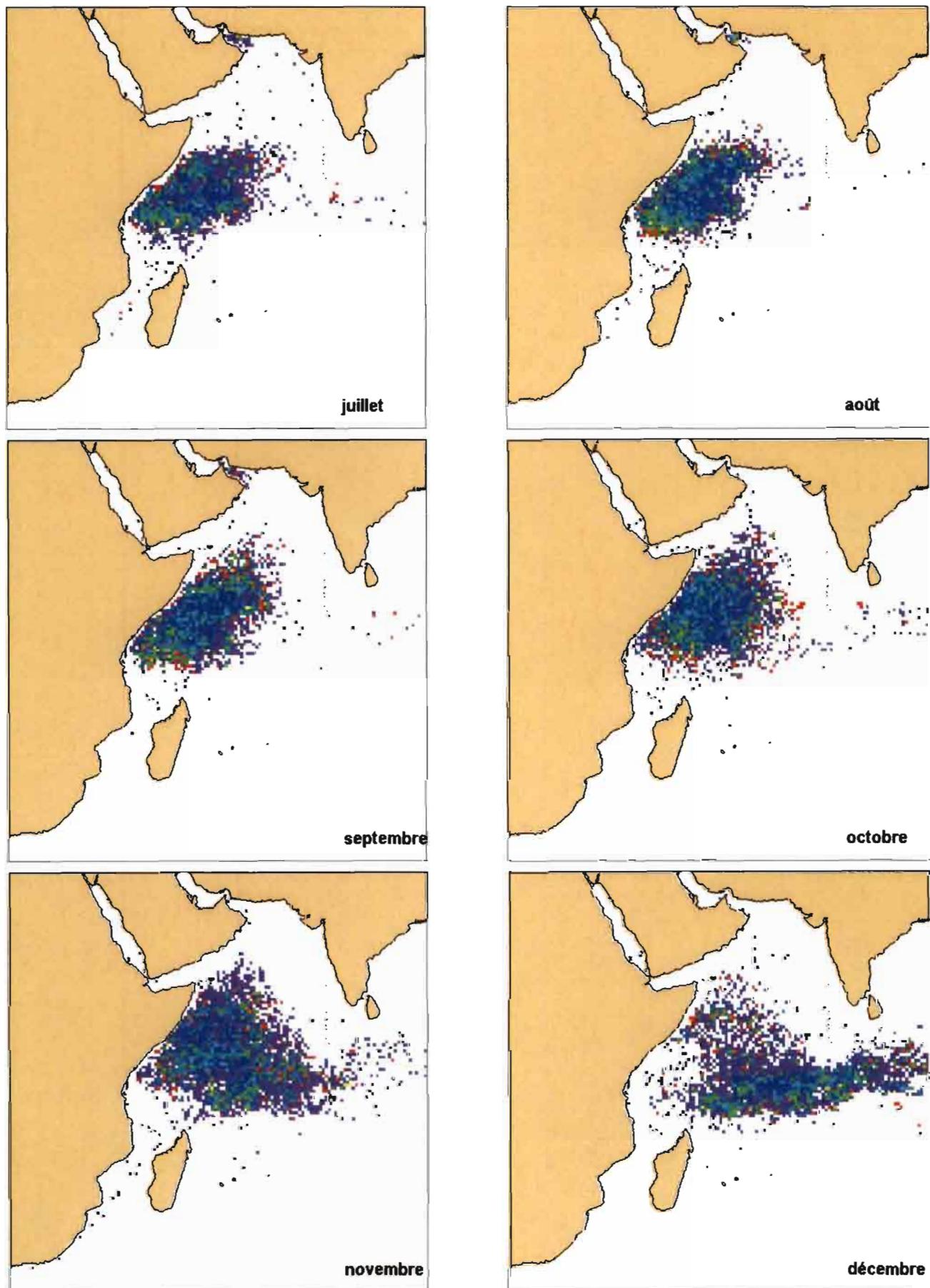


Figure 39: Distribution géographique du taux de présence par mois (bancs libres et associés)

La structure spatiale du taux de présence s'observe aussi dans les moyennes zonales et longitudinales. La figure 40 ci-dessous illustre les informations suivantes pour chaque bande d'un degré de latitude (première figure) ou de longitude (seconde figure) :

- Le taux de présence (histogramme jaune) pour les poissons de plus de 10 kg.
- Des paramètres environnementaux normalisés (courbes colorées) aux positions prospectées. La normalisation consiste à retrancher la moyenne et diviser par l'écart-type, ceci afin de comparer l'évolution des courbes sur un même axe.
- Les intervalles de confiance à 95%.

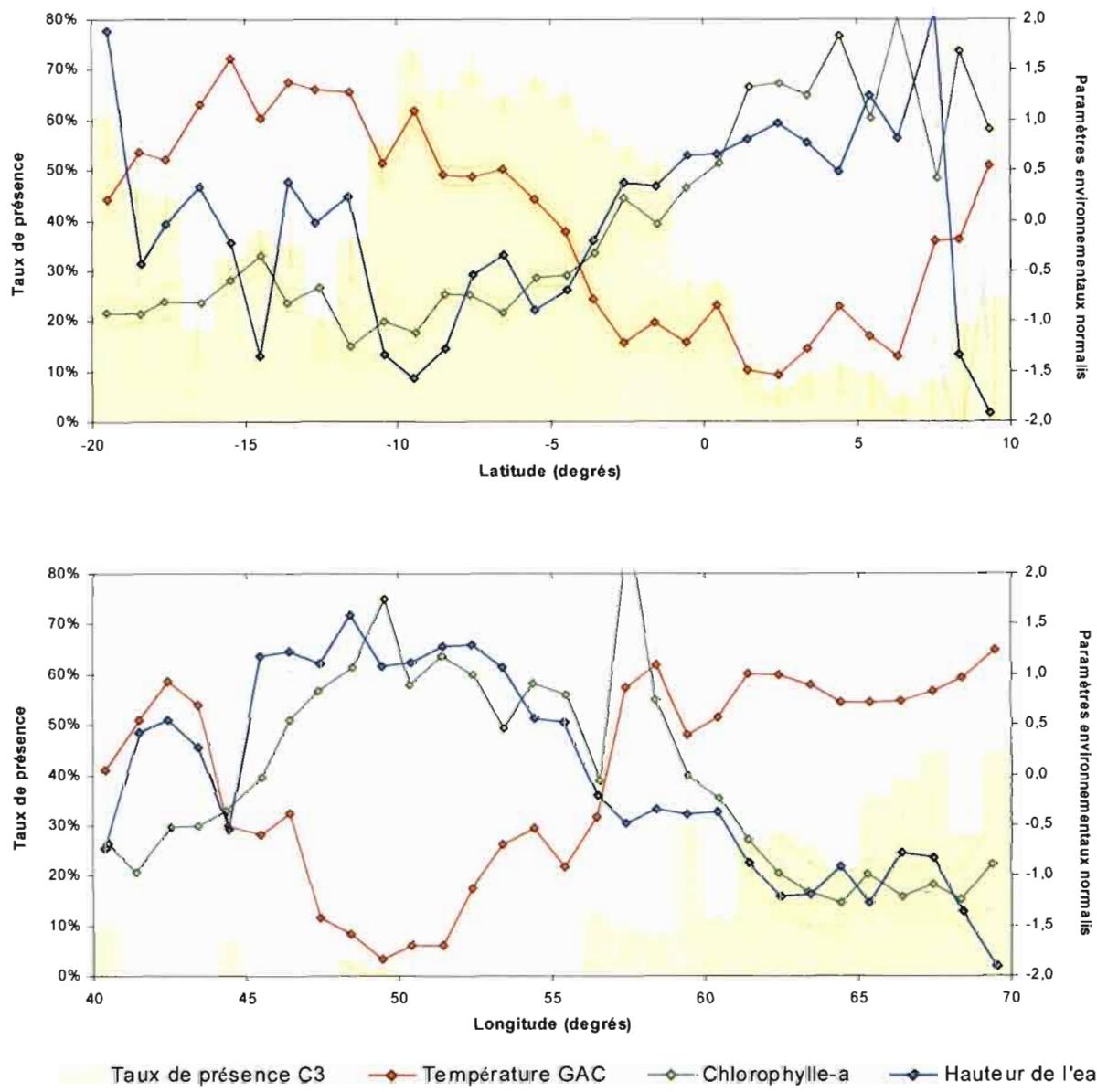


Figure 40: Variations spatiales du taux de présence et des paramètres environnementaux

L'histogramme de la figure 40 du haut suggère qu'un modèle linéaire prenant en compte uniquement la latitude (ϕ) et le cosinus de la latitude ($\cos \phi$) pourrait expliquer une partie de la variation du taux de présence. Un tel modèle sera discuté dans la section 4.4.3.3, où l'on testera aussi son apport à un modèle fondé sur la température de surface. Nous verrons que l'ajout de la latitude améliore significativement les modèles, alors que la longitude apporte peu d'amélioration aux modèles testés. La corrélation entre la latitude des prospections et le taux de présence peut être due à de multiples facteurs :

- **Latitude comme estimation des paramètres environnementaux :**

La latitude et le cosinus de la latitude (modèle de la forme $\varphi + \cos \varphi$) sont corrélés aux valeurs moyennes des paramètres environnementaux (figure 40), que l'on suppose déterminants pour le comportement du thon. Le coefficient de régression multiple R^2 est de 0,14 pour la température et 0,39 pour la chlorophylle.

- **Paramètres environnementaux comme estimation de la latitude :**

Certaines latitudes peuvent être plus « pêchantes » que d'autres pour des raisons indépendantes des paramètres environnementaux que nous mesurons. Par exemple les latitudes basses correspondent généralement à des prospections faites dans le canal du Mozambique (figure 33), une région au plancher océanique moins profond et soumise à des courants différents de ceux de la région des Seychelles. Un modèle linéaire pourrait utiliser les paramètres environnementaux comme un moyen de discriminer la région géographique de la prospection, plutôt que de refléter une réelle préférence des thons pour les conditions environnementales que nous mesurons. Dans ce cas, l'extrapolation des modèles vers de nouvelles régions géographiques risque d'être erronée.

- **Latitude comme estimation de biais dans l'échantillonnage :**

La discipline avec laquelle les pêcheurs notent les positions prospectées, avec et sans coups de senne, peut varier en fonction des flottilles ou des patrons. Or, différents patrons peuvent avoir différentes zones de pêches privilégiées. Un modèle linéaire pourrait utiliser la latitude (et donc, indirectement, les paramètres environnementaux) pour reproduire des biais dans l'échantillonnage.

Les informations dont nous disposons ne permettent pas de déterminer laquelle de ces hypothèses explique le mieux les corrélations observées dans la figure 40. La plupart des modèles présentés dans les sections 4.4.3.2 et suivantes n'incluront pas de termes de latitude, ce qui revient à privilégier la première hypothèse. Cette omission n'est pas motivée par la sélection d'une hypothèse jugée plus exacte, mais plutôt par le souci de faciliter les interprétations lors des comparaisons de modèles basés sur différents paramètres environnementaux.

La topographie du fond sous-marin influe probablement le caractère plus « pêchant » des zones discriminées par la latitude. On sait par exemple que les thons sont attirés par les monts sous-marins et les pentes récifales (section 1.4.4 à la page 26). Le système d'enrichissement en rapport avec les zones de guyots ou de hauts-fonds est mal connu des océanographes. L'enrichissement de la masse d'eau qui se produit près d'un guyot résulte d'une remontée générale des isothermes (Boehlert et Genin, 1987). Mais l'effet attractif d'un guyot sur les thonidés demeure même lorsqu'il est présent dans une zone ne comportant pas de structures d'enrichissement particulières (Petit et al., 1989). Des tests préliminaires ont suggéré que la bathymétrie est un descripteur pertinent pour les modèles, mais nous n'avons pas exploré davantage ce descripteur dans le cadre de cette thèse à cause de la faible couverture géographique des données dont nous disposons.

4.4.3.2 Concentration en chlorophylle-a

Le modèle décrit par le tableau 19 ne contient que des descripteurs dérivés des concentrations en chlorophylle-*a*. Ce modèle fut calculé à partir de 1591 lignes de données, soit le nombre le plus faible pour les modèles présentés dans cette thèse. La troisième colonne donne la probabilité qu'un descripteur ne soit pas significatif. Elle désigne le descripteur CHL comme peu significatif, ce qui s'explique par sa forte corrélation avec CHL₋₁₅. Le descripteur CHL reste toutefois nécessaire pour le calcul des gradients temporels (section 4.4.1.2). Son retrait provoque une baisse du taux de succès de 1,22% alors que le retrait de descripteurs non-significatifs n'entraîne habituellement qu'une baisse de l'ordre du dixième de pourcent, voire une hausse du taux de succès. Enfin, la dernière colonne donne à titre indicatif les coefficients de corrélation de Spearman (figure 35 à la page 102). Dans le cas du tableau 19, les signes de ces corrélations sont cohérents avec les coefficients du modèle linéaire. Cela ne sera pas toujours le cas, puisque les coefficients de corrélation de Spearman ont été calculés individuellement pour chaque descripteur, tandis que le modèle linéaire forme un tout. L'ajout ou la suppression d'un descripteur au modèle linéaire modifie les coefficients de tous les autres descripteurs, allant parfois jusqu'à inverser leur signe. Nos descripteurs étant tous corrélés, certains peuvent être interprétés comme une correction par rapport aux résultats obtenus par d'autres descripteurs. Or, le signe d'une correction n'est pas nécessairement lié au signe de la corrélation.

Tableau 19: Coefficients du modèle linéaire utilisant la concentration en chlorophylle-*a*

Descripteur	Coefficient	Probabilité > t	Spearman
Constante	1,2106	17,23%	
CHL	-0,1176	79,75%	-0,061
CHL ₋₁₅	-0,7927	0,89%	-0,093
∇CHL	-1,0067	0,00%	-0,108
∇CHL ₋₃₀	-0,4377	0,00%	-0,156
CHL × CHL ₋₁₅	-0,4186	0,81%	
CHL × ∇CHL	-0,2558	1,97%	
∇CHL × ∇CHL ₋₃₀	0,1455	0,00%	

On note que les concentrations en chlorophylle-*a* avant la prospection (CHL₋₁₅) semblent plus significatives que les concentrations du jour même (CHL), une information déjà suggérée par les corrélations de Spearman (figure 35) et commentée à la page 103.

Le taux de succès moyen pour ce modèle est de $(62,2 \pm 2,4)\%$ où le $\pm 2,4$ représente un intervalle de confiance de 95%. Rappelons que si l'on avait utilisé en guise de prédiction des nombres aléatoires, ou la même valeur présence/absence sur la totalité de la région d'étude, le taux de succès aurait été d'environ 50% pour les poissons de catégorie C3 (section 4.4.2).

Puisque les « prédictions » du modèle sont des valeurs continues d'environ -1 à +1, on peut s'attendre à ce que les classements en zones de présence ou d'absence de thons soient d'autant plus incertains que les valeurs sont proches de zéro. La figure 41 représente la distribution des valeurs de sorties (histogramme bleu) aux positions prospectées, ainsi que du taux de succès (points oranges ; les lignes pointillées représentent l'intervalle de confiance de 95%). Les zones classées « présence » semblent effectivement plus fiables à mesure qu'augmentent les valeurs de sorties du modèle. En revanche, les zones classées « absence » semblent conserver une plus grande part d'incertitude. Toutefois, l'intervalle de confiance est trop large pour confirmer cette dernière observation.

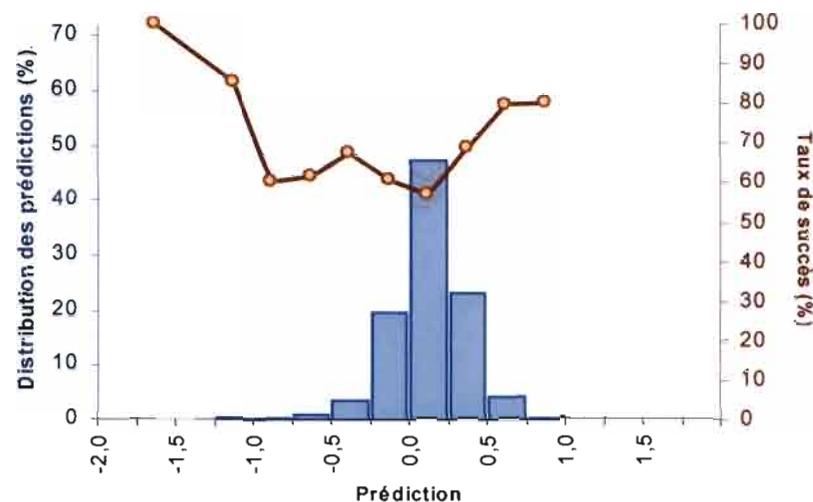


Figure 41: Distribution des prédictions du modèle fondé sur la chlorophylle-a

Des cartes de « potentiel de pêche » ont été générées à partir des sorties du modèle linéaire. Nous avons généré une carte tous les 5 jours sur toute la période pour laquelle nous disposons des données, soit du 12 octobre 1997 au 9 mai 2003 dans le cas de la chlorophylle-a. La région géographique couverte ne dépasse pas 30°S au sud, 30°N au nord, 30°E à l'ouest et 90°E à l'est, soit la même région que celle qui est illustrée dans la figure 33 (page 96) excepté la limite sud. Cette région englobe 96,7% des données de prospections, mais seulement 34% de sa surface océanique a reçu au moins une visite durant les 10 années de notre jeu de données. Le potentiel de pêche affiché dans la surface restante (la partie blanche de la figure 33) sera une extrapolation du modèle linéaire.

La figure 42 donne un exemple de carte de potentiel de pêche (à droite) pour la journée du 23 avril 2000. A titre de comparaison, la figure de gauche donne les concentrations en chlorophylle-a pour la période du 22 au 29 avril 2000 inclusivement. Les cercles rouges représentent les positions auxquelles des thons de plus de 10 kg ont été pêchés, tandis que les cercles gris représentent les prospections sans pêche.

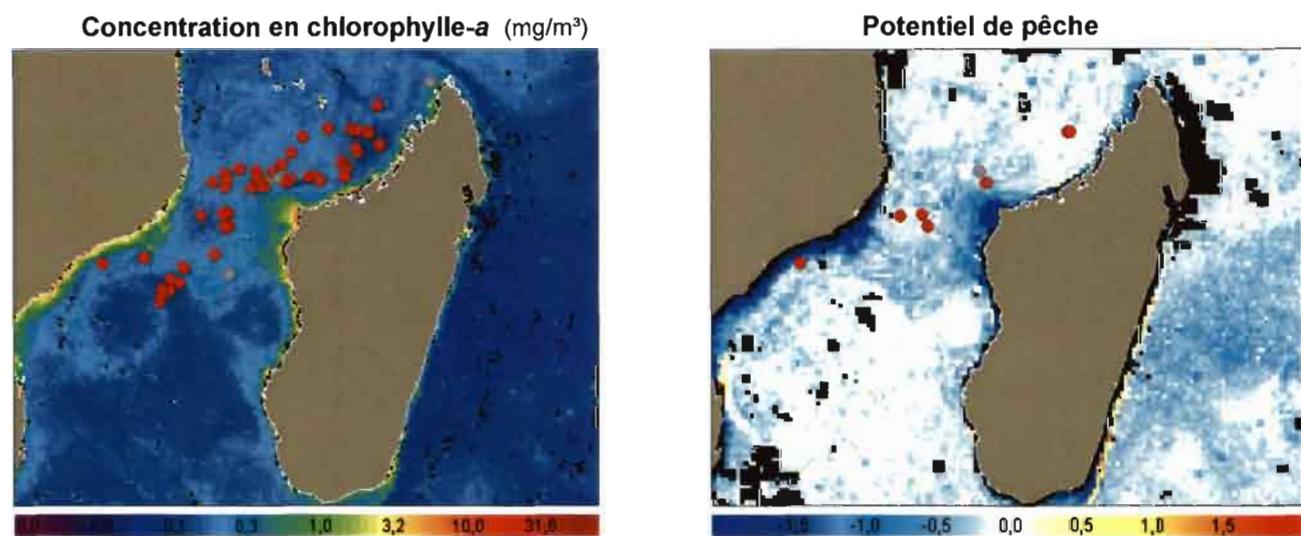


Figure 42: Exemple de carte de potentiel générée à partir des concentrations en chlorophylle-a

A partir d'un examen visuel de différentes cartes de potentiel basées sur le modèle du tableau 19, on observe que :

- le modèle prédit une absence de thon sur la majorité de la surface océanique (en bleu) ;
- les zones dans lesquelles une présence de thons est prédictive (en jaune) sont rares et étroites ;
- les prospections se font habituellement dans des régions proches du neutre (en blanc) ;

- la corrélation entre les nuances de bleus et le succès des prospections n'est pas évidente.

Même sur les cartes affichant des nuances de bleus plus foncées (valeurs tendant vers -1), les corrélations ne semblent guère plus convaincantes. Cette observation est appuyée dans la figure 41 par le caractère relativement constant du taux de succès entre les valeurs -1 et 0 du modèle. En revanche, les rares zones jaunes semblent réussir davantage, lorsqu'elles ont été prospectées.

4.4.3.3 Température de surface GAC (*Global Area Coverage*)

Le modèle décrit par le tableau 20 ne contient que des descripteurs dérivés de la température de surface GAC. La résolution spatiale et le pas de temps pour ce paramètre sont les mêmes que ceux de la chlorophylle-*a* (section 4.4.3.2). Ce modèle a été calculé à partir de 12 633 lignes de données, soit près de 8 fois plus que le modèle dérivé des concentrations en chlorophylle-*a*. Cette plus grande taille de l'échantillon s'explique par le fait que les données de température sont disponibles dès le début de nos données de prospection (janvier 1991), alors que les données de chlorophylle-*a* ne commencent qu'en septembre 1997. La troisième colonne (en grisé) donne les coefficients d'un modèle quasiment identique à la différence près que toutes les occurrences du descripteur ∇SST_{-15} ont été substituées par ∇SST .

Tableau 20: Coefficients du modèle linéaire utilisant la température GAC

Descripteur	Coefficient	$\nabla\text{SST}_{-15} \rightarrow \nabla\text{SST}$	Probabilité > t	Spearman
Constante	-36,2375	-36,4856	0,00%	
SST_{-15}	1,2374	1,2282	0,00%	-0,071
SST_{-30}	1,5570	1,5157	0,00%	-0,062
∇SST_{-15}	-2,4074	-1,4561	0,00%	0,077
$\text{SST}_{-15} \times \text{SST}_{-30}$	-0,0537	-0,0515	0,00%	
$\text{SST}_{-15} \times \nabla\text{SST}_{-15}$	0,0895	0,0545	0,00%	

Avant d'arrêter notre choix sur les descripteurs du tableau 20, quelques essais empiriques ont suggéré que les gradients les plus pertinents sont ∇SST et ∇SST_{-15} . Incorporer un seul de ces gradients dans le modèle donne un meilleur résultat que d'incorporer les deux. L'interprétation du rôle du gradient de température varie en fonction du décalage de temps choisi. Le gradient 15 jours avant la prospection (∇SST_{-15}) met l'accent sur l'aspect trophique : des conditions favorables à la production primaire entraînent une production secondaire et tertiaire qu'exploitera le thon. Le gradient le jour même (∇SST) met plutôt l'accent sur l'aspect comportemental : le thon est attiré par les gradients. Des expériences de marquages appuient cette dernière hypothèse. Les figures 43 et 44 (discutées plus bas) comparent ces deux modèles.

Le premier modèle utilise le gradient 15 jours avant la prospection (∇SST_{-15}). Le taux de succès moyen est de $(59,2 \pm 0,9)\%$ où le $\pm 0,9$ représente un intervalle de confiance de 95%. La figure 43 représente la distribution des valeurs de sorties aux positions prospectées, ainsi que du taux de succès. La légende est la même que celle de la figure 41 dans la section précédente.

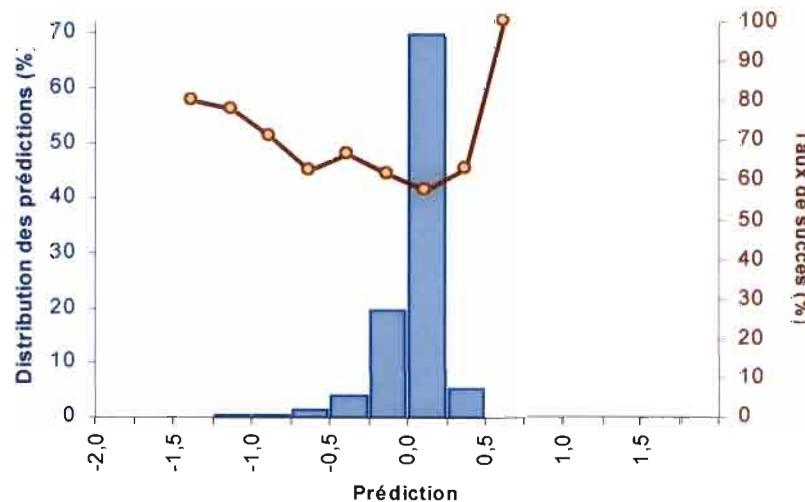


Figure 43: Distribution des prédictions du modèle basé sur la température de surface GAC

Le remplacement du gradient ∇SST_{-15} par ∇SST modifie peu le taux de succès moyen : l'augmentation ($0,15\%$) reste en deçà de l'intervalle de confiance. Il modifie en apparence la distribution (figure 44), mais là

encore les changements restent en deçà de l'intervalle de confiance. La comparaison de ces deux modèles ne permet donc pas de déterminer si une interprétation comportementale est plus appropriée qu'une interprétation trophique de l'effet des gradients de température. Pour cette section, nous privilégierons le gradient ∇SST_{-15} car il évite d'introduire un décalage temporel supplémentaire dans le modèle.

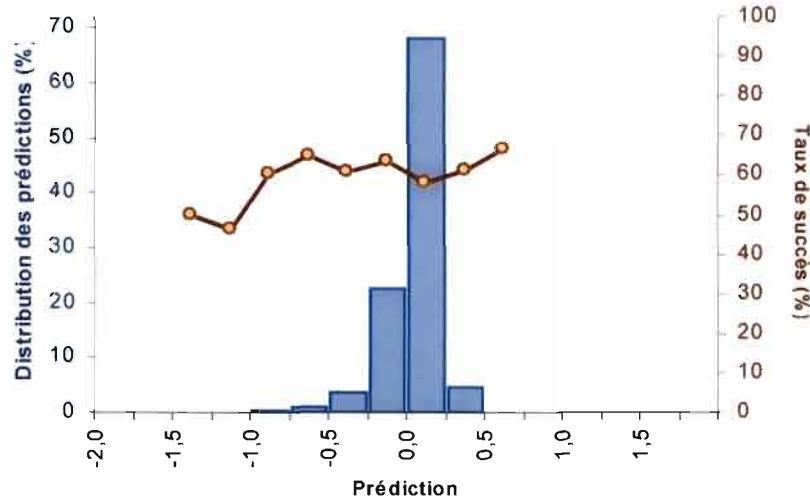


Figure 44: Version modifiée du modèle de la figure 43 après remplacement de ∇SST_{-15} par ∇SST

Pour ces deux modèles, environ 70% des valeurs de sorties aux positions prospectées sont comprises dans une seule barre d'histogramme entre 0 et 0,25. Cette observation se répète pour tous les modèles de cette thèse qui utilisent exclusivement des descripteurs dérivés de la température de surface (figures 43, 44 et 47). Ces modèles produisent les sorties les moins bien distribuées (donc des intervalles de confiance qui augmentent rapidement dès que l'on s'éloigne du mode) ainsi que les taux de succès les plus faibles. En comparaison, le modèle basé sur la chlorophylle-*a* ne produisait des valeurs de sorties entre 0 et 0,25 que dans 47% des cas (figure 41).

Des cartes de potentiel de pêche ont été générées à partir du modèle du tableau 20 pour la même région géographique et le même pas de temps que celles de la section 4.4.3.2, excepté que les cercles rouges (présences de thons) ont été remplacés par des cercles bleus dans la figure de gauche. La figure 45 à droite donne un exemple de carte pour la journée du 3 juin 1997. La figure de gauche illustre la température de surface pour la période du 2 au 9 juin 1997.

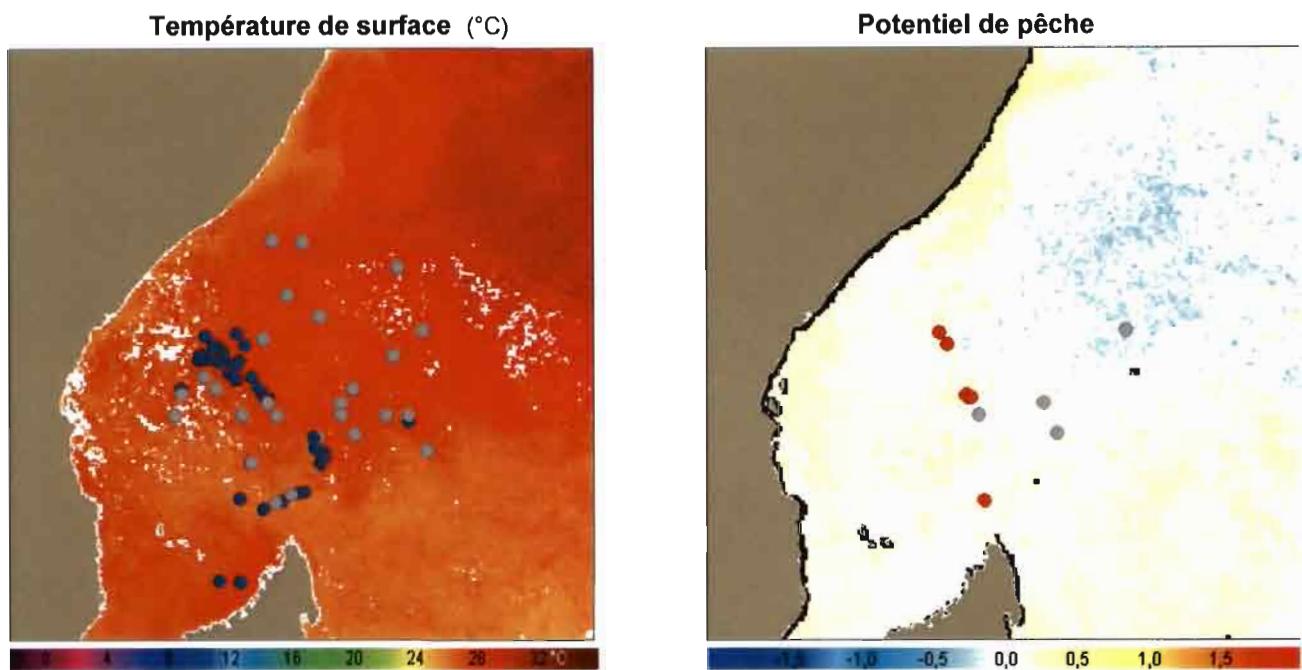


Figure 45: Exemple de carte de potentiel générée à partir des températures de surface

Contrairement au modèle basé sur les concentrations en chlorophylle-*a*, ce modèle-ci prédit une présence de thons sur de grandes surfaces océaniques. La superficie de ces zones varie en fonction de la latitude et des saisons, pour atteindre un maximum durant l'été austral. Cette variation saisonnière est cohérente avec le signe positif des coefficients SST₋₁₆ et SST₋₃₀ (tableau 26), mais n'est pas nécessairement représentative de la variation des zones favorables à la présence de thon. En effet, si l'on admet que le thon a une préférence pour des eaux plus chaudes ou plus froides, et que la température associée aux expressions « plus chaudes » et « plus froides » dépend fortement de la latitude et de la saison, alors en l'absence de descripteurs tels que la date ou une climatologie de la température, le modèle du tableau 26 peut difficilement produire des valeurs proches de -1 ou +1 en toute saison. On peut tenter de prendre en compte les variations saisonnières et géographiques de différentes façons :

- Construire des modèles distincts pour différentes régions géographiques et saisons. Cette approche aurait l'avantage de prendre en compte d'éventuelles différences de comportement du thon selon les régions (notamment le canal du Mozambique par rapport à la bande équatoriale), mais comporte l'inconvénient de diminuer le nombre de données disponibles pour chaque modèle. Cette approche n'a pas été explorée dans le cadre de cette thèse ; nous nous sommes limités à une analyse globale du sud-ouest de l'océan Indien en première approche.
- Utiliser les anomalies de température plutôt que les températures directes. Cette approche n'a pas été explorée dans le cadre de cette thèse, car elle implique l'introduction de nouveaux descripteurs dérivés d'une nouvelle opération (section 4.1.2 à la page 91). En effet, il faudrait sans-doute tester la pertinence de ces anomalies en construisant des modèles utilisant à la fois les températures directes et les anomalies.
- Inclure la latitude ou le cosinus de la latitude comme un des descripteurs du modèle. Ce descripteur est corrélé à l'intensité moyenne du rayonnement solaire incident. La corrélation entre la latitude et la température moyenne a été discutée dans la section 4.4.3.1 (page 112).

On note que le taux de succès d'un modèle basé uniquement sur le cosinus de la latitude, sans aucun paramètre environnemental, est de $(56,6 \pm 0,8)\%$. L'ajout de ce terme au modèle du tableau 20 n'améliore pas significativement le taux de succès. En revanche, l'ajout de la latitude (sans cosinus) ainsi que son interaction avec la température (c'est-à-dire l'ajout des termes φ et $\varphi \times \text{SST}_{-16}$) fait passer le taux de succès de 59 à 63%. La figure 46 illustre la distribution des sorties du modèle du tableau 20 augmenté de ces termes. Seulement 54% des valeurs de sorties sont comprises entre 0 et 0,25, contre 70% pour les figures 43 et 44. Cette meilleure distribution des valeurs de sorties se traduit par un intervalle de confiance plus serré. On peut ainsi vérifier que contrairement au modèle basé sur la chlorophylle (figure 41 à la page 115), ce modèle a plus de succès dans les prédictions d'absence de thon que dans les prédictions de présence.

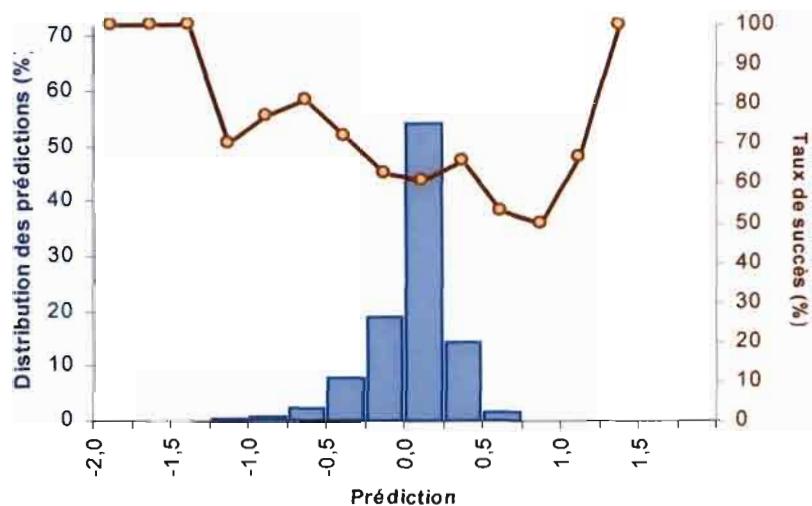


Figure 46: Version modifiée du modèle de la figure 43 avec ajout des termes ϕ et $\phi \times SST_{-15}$

Malgré la pertinence de la latitude dans les modèles basés sur la température, nous n'avons pas retenu ce descripteur dans les analyses qui vont suivre. Nous comparons plutôt des modèles n'utilisant rien d'autre que le paramètre étudié, car l'objectif de cette thèse se limite à estimer la pertinence d'un paramètre (l'anomalie de la hauteur de l'eau) par rapport aux autres. Il n'est pas certain toutefois que l'omission de la latitude permet de modéliser les seuls effets de la température. Comme mentionné à la page 113, il est possible qu'un modèle sans dépendance explicite envers la latitude utilise la température comme estimateur de phénomènes corrélés avec la latitude.

4.4.3.4 Température de surface LAC (Local Area Coverage)

Le modèle décrit par le tableau 21 ne contient que des descripteurs dérivés de la température de surface captés par l'antenne de l'IRD à l'île de la Réunion. Ce modèle a été calculé à partir de 6246 lignes de données, soit moitié moins que le modèle dérivé des températures GAC (section 4.4.3.3). Cette plus petite taille de l'échantillon s'explique par une couverture spatiale plus restreinte (8% des données de prospection sont en dehors de la région couverte par nos données LAC) et une couverture nuageuse plus importante (les synthèses des données LAC sont faites sur 5 jours, contre 8 pour les données GAC).

Tableau 21: Coefficients du modèle linéaire utilisant la température LAC

Descripteur	Coefficient LAC	Coefficient GAC	Probabilité $> t $	Spearman
Constante	-43,5621	-34,2373	0,00%	
SST_{-15}	1,5249	1,1682	0,00%	-0,057
SST_{-25}	1,5765	1,3001	0,00%	-0,054
$SST_{-15} \times SST_{-25}$	-0,0552	-0,0445	0,00%	
∇SST	0,0599	0,0725	2,07%	0,040

Les descripteurs apparaissant dans le tableau 21 sont légèrement différents de ceux qui apparaissaient dans le tableau 20 (modèle basé sur la température GAC). Ce choix résulte de la suppression d'un terme croisé non-significatif lorsque appliqué à la température LAC ($SST_{-15} \times \nabla SST_{-15}$), suivi d'ajustements empiriques pour tenter d'améliorer le taux de succès. La colonne en grisé du tableau 21 donne, à titre indicatif, les coefficients du même modèle déterminés à partir des températures GAC.

Le taux de succès moyen pour le modèle utilisant les températures LAC est de $(55,5 \pm 1,2)\%$ où le $\pm 1,2$ représente un intervalle de confiance de 95%. Il s'agit du taux de succès le plus faible de cette thèse ; il est à peine supérieur à ce que l'on aurait obtenu avec des prédictions aléatoires (environ 51%). À titre de comparaison, le même modèle utilisant les températures GAC (colonne en grisé du tableau 21) donne un coefficient de succès de 58,6%. La figure 36 (page 104) suggérait déjà que les données GAC sont plus pertinentes dans le cas des bancs libres de poissons de plus de 10 kg. La figure 47 représente la distribution des valeurs de sorties aux positions prospectées, ainsi que du taux de succès. Cette figure est soumise aux mêmes contraintes que les figures 43 et 44 (température GAC sans dépendance envers la latitude), c'est-à-dire un fort écart-type causé par la présence de 70% des valeurs dans une seule barre de l'histogramme.

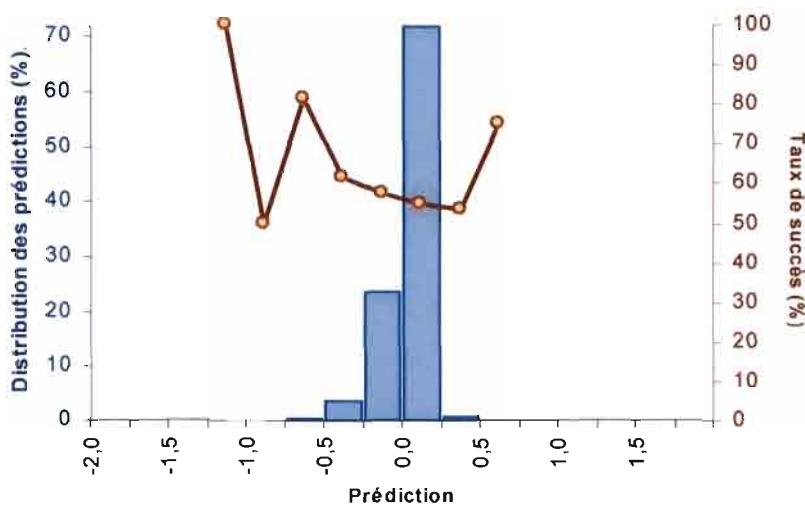


Figure 47: Distribution des prédictions du modèle fondé sur la température de surface

Des cartes de potentiel de pêche ont été générées à partir de ce modèle pour la même région géographique et le même pas de temps que celles de la section 4.4.3.2, excepté que les cercles rouges (présences de thons) ont été remplacés par des cercles orange pâle dans la figure de gauche. La figure 48 à

droite donne un exemple de carte pour la journée du 24 mai 1998. La figure de gauche illustre la température de surface pour la période du 21 au 26 mai 1998.

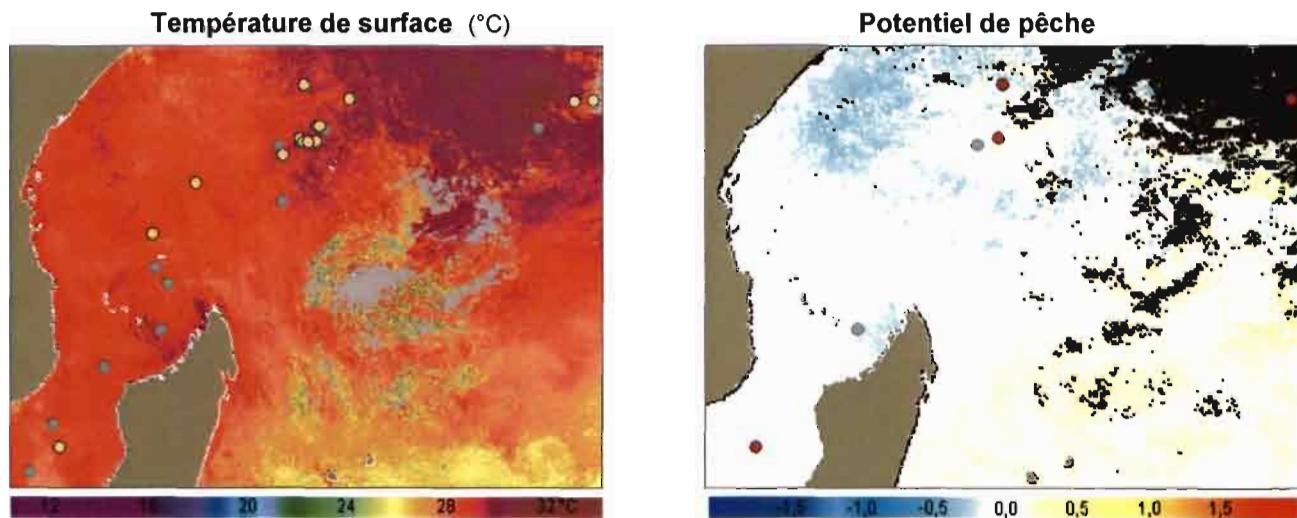


Figure 48: Exemple de carte de potentiel générée à partir des températures de surface

Comme pour le modèle de la section précédente (figure 45 à la page 119), ce modèle prédit une présence de thon sur de grandes surfaces océaniques. Bien que la figure 48 soit un exemple de carte de potentiel qui correspond bien aux résultats réels des prospections (du moins celles dont nous disposons des données), elle est loin d'être la norme. La plupart des cartes montrent de nombreux points rouges (position de pêche) au milieu de régions bleues, ou de points gris (position de prospection sans pêche) au milieu de régions jaunes.

4.4.3.5 Anomalie de la hauteur de l'eau

Le modèle décrit par le tableau 22 ne contient que des descripteurs dérivés des anomalies de la hauteur de l'eau (SLA). Ce modèle a été calculé à partir de 12 558 lignes de données, soit pratiquement autant que le modèle dérivé des températures GAC (section 4.4.3.3).

Tableau 22: Coefficients du modèle linéaire utilisant l'anomalie de la hauteur de l'eau

Descripteur	Coefficient	Probabilité $> t $	Spearman
Constante	-0,9315	0,00%	
SLA	-0,0163	0,00%	-0,219
SLA ₋₃₀	-0,0479	0,00%	-0,248
ΔSLA	0,3481	0,00%	-0,112
ΔSLA ₋₃₀	0,2610	0,00%	-0,146
ΔSLA × SLA ₋₃₀	0,0084	0,00%	
ΔSLA × ΔSLA ₋₃₀	-0,1021	0,00%	

Le taux de succès moyen pour ce modèle est de $(62,6 \pm 0,9)\%$, soit pratiquement la même valeur que pour le modèle basé sur la chlorophylle-*a* (page 114), mais avec près de 8 fois plus de données. La figure 49 illustre la distribution du taux de succès en fonction des valeurs des prédictions. On y observe une amélioration significative du taux de succès à mesure que les prédictions s'éloignent de 0.

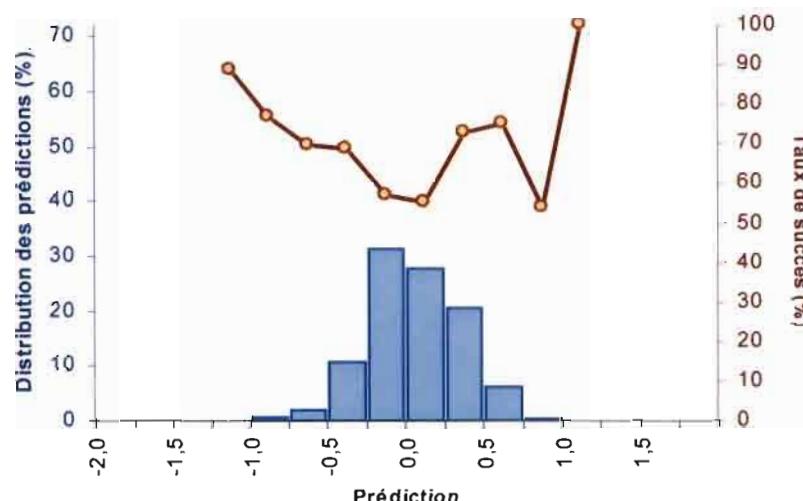


Figure 49: Distribution des prédictions du modèle fondé sur l'anomalie de hauteur de l'eau

Des cartes de potentiel de pêche ont été générées à partir de ce modèle pour la même région géographique et le même pas de temps que celles des sections précédentes. La figure 50 à droite donne un exemple de carte pour la journée du 19 février 1994. La figure de gauche illustre l'anomalie de hauteur de l'eau pour la période du 15 au 24 février 1994. Par rapport aux cartes de potentiel basées sur les autres paramètres, celles de la figure 50 ont l'avantage de n'être pratiquement pas masquées par la couverture nuageuse.

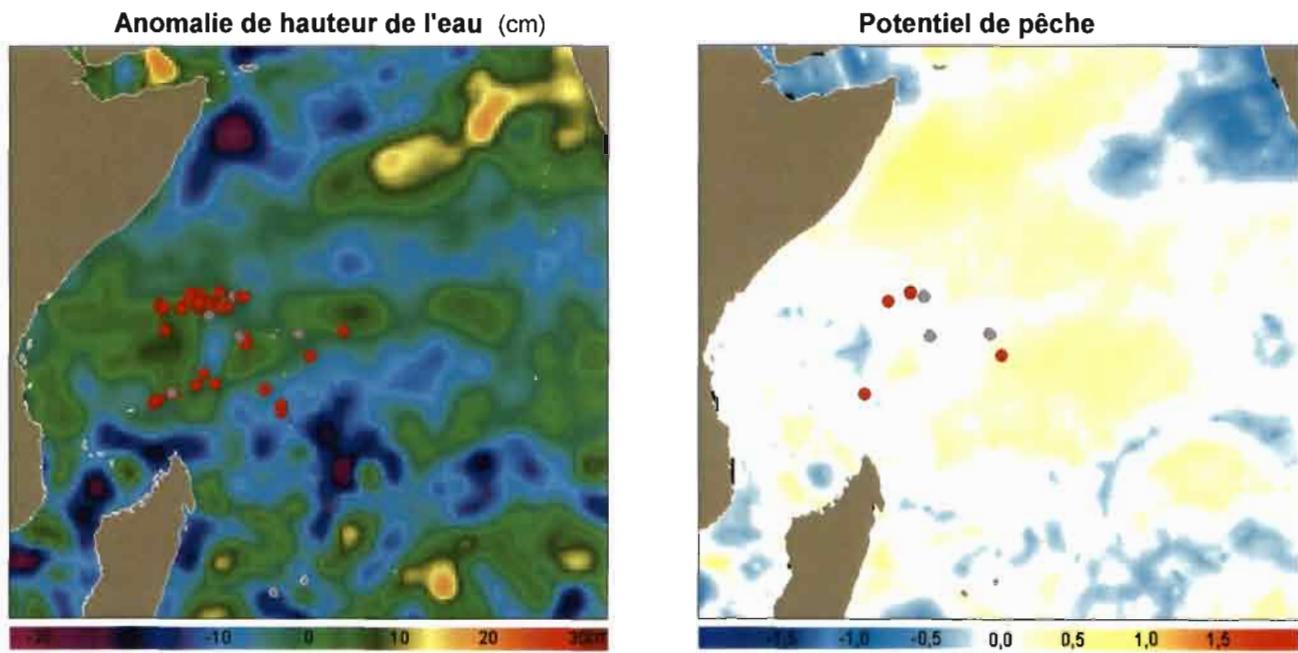


Figure 50: Exemple de carte de potentiel générée à partir des anomalies de hauteur de l'eau

Les cartes de potentiels basées sur la SLA prédisent souvent une probabilité de présence le long de l'équateur, ce qui correspond effectivement à une zone davantage prospectée par les pêcheurs (figure 38 à la page 109). Cet effet est peut-être à relier à la propagation d'ondes de Kelvin et de Rossby équatoriales, qui se propagent rapidement dans cette région (section 1.5.1 à la page 31).

4.4.3.6 Concentration en chlorophylle-a et anomalie de la hauteur de l'eau

Le modèle décrit par le tableau 23 ne contient que des descripteurs dérivés des concentrations en chlorophylle- a (dénommé « CHL ») et des anomalies de la hauteur de l'eau (« SLA »). Les termes apparaissant dans le tableau 23 sont l'union des termes apparaissant dans les tableaux 19 (modèle utilisant la CHL seulement) et 22 (modèle utilisant la SLA seulement). La colonne en grisé du tableau 23 donne les coefficients des tableaux 19 et 22 à titre de comparaison. Ce modèle a été calculé à partir de 1591 lignes de données.

Tableau 23: Coefficients d'un modèle linéaire utilisant les données de CHL et de SLA

Descripteur	Coefficient	Probabilité $> t $	Paramètre seul
Constante	0,2743	78,50%	
CHL	-0,1556	72,67%	-0,1176
CHL ₋₁₅	-0,7455	1,14%	-0,7927
∇ CHL	-0,6060	0,72%	-1,0067
∇ CHL ₋₃₀	-0,1144	17,93%	-0,4377
CHL × CHL ₋₁₅	-0,2783	7,56%	-0,4186
CHL × ∇ CHL	-0,2151	4,12%	-0,2558
∇ CHL × ∇ CHL ₋₃₀	0,0460	18,74%	0,1455
SLA	-0,0134	0,00%	-0,0163
SLA ₋₃₀	-0,0138	20,32%	-0,0479
∇ SLA	0,0371	81,15%	0,3481
∇ SLA ₋₃₀	-0,1322	36,21%	0,2610
∇ SLA × SLA ₋₃₀	0,0005	85,22%	0,0084
∇ SLA × ∇ SLA ₋₃₀	-0,0142	70,72%	-0,1021

Le taux de succès moyen pour ce modèle est de $(69,0 \pm 2,3)\%$. Cela représente une amélioration d'environ 6% par rapport aux modèles qui n'utilisaient que les concentrations en chlorophylle- a (section 4.4.3.2) ou que les anomalies de la hauteur de l'eau (section 4.4.3.5). Ces deux paramètres semblent donc offrir une certaine complémentarité. Notons que le modèle du tableau 23 ne prend pas en compte les interactions entre les anomalies de hauteur de l'eau et les concentrations en chlorophylle- a . Or, ces interactions existent probablement. Par exemple une concentration élevée de chlorophylle- a est susceptible d'être dispersée plus rapidement si elle se trouve dans une région de fort gradient de SLA, puisque les courants géostrophiques sont susceptibles d'y être plus forts. De tels termes croisés seront testés dans la section 4.4.3.9.

4.4.3.7 Concentration en chlorophylle-a et température de surface (CHL – SST)

Le modèle décrit par le tableau 24 ne contient que des descripteurs dérivés des concentrations en chlorophylle- a (« CHL ») et des températures de surface GAC (dénommées « SST » pour cette section). Les termes apparaissant dans le tableau 24 sont l'union des termes apparaissant dans les tableaux 19 (modèle utilisant la CHL seulement) et 20 (modèle utilisant la SST seulement). La colonne en grisé donne les coefficients des tableaux 19 et 20 à titre de comparaison. Ce modèle a été calculé à partir de 1547 lignes de données.

Tableau 24: Coefficients d'un modèle linéaire utilisant l'ensemble des paramètres

Descripteur	Coefficient	Probabilité $> t $	Paramètre seul
Constante	4,5784	0,72	
CHL	-0,4059	0,3856	-0,1176
CHL ₋₁₅	-1,2583	0,0001	-0,7927
∇CHL	-0,8437	0,0004	-1,0067
∇CHL ₋₃₀	-0,2822	0,0009	-0,4377
CHL × CHL ₋₁₅	-0,4970	0,0028	-0,4186
CHL × ∇CHL	-0,2547	0,0215	-0,2558
∇CHL × ∇CHL ₋₃₀	0,0989	0,0050	0,1455
SST ₋₁₅	-0,1552	0,7304	1,2374
SST ₋₃₀	0,3290	0,4378	1,5570
∇SST ₋₁₅	-3,7946	0,0023	-2,4074
SST ₋₁₅ × SST ₋₃₀	-0,0119	0,4208	-0,0537
SST ₋₁₅ × ∇SST ₋₁₅	0,1306	0,0028	0,0895

Le taux de succès moyen pour ce modèle est de $(65,2 \pm 2,4)\%$, ce qui représente une amélioration de 3% par rapport au modèle n'utilisant que la chlorophylle- a (section 4.4.3.2). L'apport de la température au modèle de la section 4.4.3.2 est donc moindre que l'apport de l'anomalie de la hauteur de l'eau à ce même modèle (voir section 4.4.3.6).

4.4.3.8 Température et anomalie de la hauteur de l'eau (SST – SLA)

Le modèle décrit par le tableau 25 ne contient que des descripteurs dérivés de la température de surface GAC (dénommé « SST ») et des anomalies de la hauteur de l'eau (« SLA »). Les termes apparaissant dans le tableau 25 sont l'union des termes apparaissant dans les tableaux 20 (modèle utilisant la SST seulement) et 22 (modèle utilisant la SLA seulement). La colonne en grisé du tableau 25 donne les coefficients des tableaux 20 et 22 à titre de comparaison. Ce modèle a été calculé à partir de 10 945 lignes de données, soit presque autant que les modèles utilisant la SST ou la SLA seulement. L'omission de la chlorophylle-*a* dans ce modèle permet d'éviter une perte importante de données, contrairement aux modèles des sections 4.4.3.6, 4.4.3.7 et 4.4.3.9.

Tableau 25: Coefficients d'un modèle linéaire utilisant la température GAC et la SLA

Descripteur	Coefficient	Probabilité > t	Paramètre seul
Constante	-19,0247	0,00%	
SST ₋₁₅	0,5652	0,00%	1,2374
SST ₋₃₀	0,9078	0,00%	1,5570
∇SST ₋₁₅	-2,5457	0,00%	-2,4074
SST ₋₁₅ × SST ₋₃₀	-0,0294	0,00%	-0,0537
SST ₋₁₅ × ∇SST ₋₁₅	0,0934	0,00%	0,0895
SLA	-0,0144	0,00%	-0,0163
SLA ₋₃₀	-0,0459	0,00%	-0,0479
∇SLA	0,2907	0,00%	0,3481
∇SLA ₋₃₀	0,2105	0,04%	0,2610
∇SLA × SLA ₋₃₀	0,0079	0,00%	0,0084
∇SLA × ∇SLA ₋₃₀	-0,0870	0,00%	-0,1021

Le taux de succès moyen pour ce modèle est de (63,3 ±0,9)%, ce qui représente une amélioration de 0,7% par rapport au modèle basé sur l'anomalie de hauteur de l'eau seulement (tableau 22), et une amélioration de près de 4,1% par rapport au modèle basé sur la température seulement (tableau 20). L'amélioration de 0,7% étant inférieure à l'intervalle de confiance, le modèle du tableau 25 ne permet pas de conclure à un apport significatif de la température de surface par rapport à un modèle basé seulement sur l'anomalie de la hauteur de l'eau.

Nous avons tenté d'obtenir un modèle plus efficace en ajustant les termes de manière empirique, et en introduisant des termes croisés entre la température (SST) et l'anomalie de la hauteur de l'eau (SLA). Le tableau 26 décrit un des modèles obtenus. Il fut calculé à partir de 10 011 lignes de données. Contrairement à ce qui était annoncé dans le tableau 17 (page 106), les gradients de SLA apparaissant dans ce modèle ont été calculés à partir d'un opérateur de taille 3×3 classique. Les lignes en jaune représentent des termes qui étaient présents dans le tableau 25. Les autres lignes sont des termes propres à ce modèle. Parmi les nouveaux termes, on note des termes au carré. Ces derniers avaient été essayés dans les modèles précédents, mais rarement retenus comme termes significatifs.

Tableau 26: Coefficients d'un modèle alternatif utilisant la température GAC et la SLA

Descripteur	Coefficient	Probabilité > t
Constante	-19,6724	0,00%
SLA ²	-0,0004	0,00%
∇SLA	-0,6217	7,27%
∇SLA ²	-0,0791	0,00%
SLA ₋₃₀	-0,2772	0,00%
SLA ₋₃₀ ²	-0,0002	1,40%
∇SLA ₋₃₀	0,3665	31,77%
SST	0,3390	9,54%
∇SST	-0,8167	4,78%
∇SST ₋₁₅	-2,2110	0,00%
SST ₋₃₀	1,3091	0,00%

Descripteur	Coefficient	Probabilité $> t $
SST_{-30}^2	-0,0188	0,33%
SLA \times SLA ₋₃₀	0,0005	0,00%
SLA \times ∇ SLA ₋₃₀	0,0026	9,51%
SLA \times SST	-0,0009	0,01%
SLA ₋₃₀ \times ∇ SLA	0,0010	55,68%
SLA ₋₃₀ \times ∇ SLA ₋₃₀	0,0068	0,01%
SLA ₋₃₀ \times SST	0,0027	2,22%
SLA ₋₃₀ \times SST ₋₃₀	0,0055	0,00%
∇ SLA \times SST ₋₃₀	0,0394	0,21%
∇ SLA ₋₃₀ \times SST	-0,0589	0,00%
∇ SLA ₋₃₀ \times SST ₋₃₀	0,0418	0,46%
∇ SLA ₋₃₀ \times ∇ SST ₋₁₅	0,0346	14,30%
SST \times SST ₋₃₀	-0,0157	2,74%
SST \times ∇ SST	0,0747	0,00%
SST \times ∇ SST ₋₁₅	0,0409	1,50%
SST ₋₃₀ \times ∇ SST	-0,0446	0,97%
SST ₋₃₀ \times ∇ SST ₋₁₅	0,0369	3,00%

Le taux de succès moyen pour ce modèle est de $(65,1 \pm 0,9)\%$, soit 1,8% de plus que le modèle du tableau 25. Ce modèle contient un certain nombre de termes considérés non-significatifs, mais leur retrait provoque tout de même une légère baisse du taux de succès. La figure 51 ci-dessous représente deux exemples de cartes de potentiel générées à partir de ce modèle.

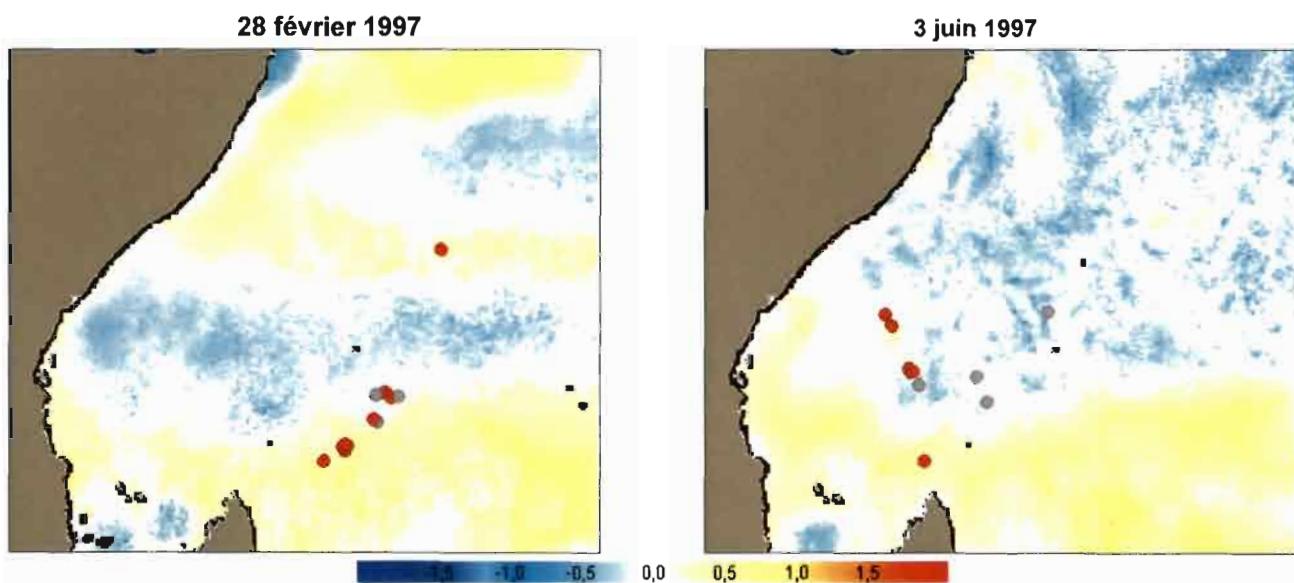


Figure 51: Exemple de carte de potentiel générée à partir de la température GAC et la SLA

4.4.3.9 Ensemble des paramètres (CHL – SST – SLA)

Le modèle décrit par le tableau 27 est similaire à celui du tableau 25, augmenté de termes dérivés des concentrations en chlorophylle-*a*. La colonne en grisé du tableau 27 donne les coefficients des tableaux 19, 20 et 22 à titre de comparaison. Ce modèle a été calculé à partir de 1547 lignes de données, soit 7 fois moins que les modèles de la section précédente.

Tableau 27: Coefficients d'un modèle linéaire utilisant l'ensemble des paramètres

Descripteur	Coefficient	Probabilité > <i>t</i>	Paramètre seul
Constante	20,6133	10,94%	
CHL	-0,3843	40,32%	-0,1176
CHL ₋₁₅	-1,0367	0,11%	-0,7927
∇CHL	-0,5974	0,98%	-1,0067
∇CHL ₋₃₀	-0,0626	46,42%	-0,4377
CHL × CHL ₋₁₅	-0,3633	2,79%	-0,4186
CHL × ∇CHL	-0,2246	3,78%	-0,2558
∇CHL × ∇CHL ₋₃₀	0,0354	31,24%	0,1455
SST ₋₁₅	-0,8417	6,62%	1,2374
SST ₋₃₀	-0,3945	36,31%	1,5570
∇SST ₋₁₅	-2,1289	8,26%	-2,4074
SST ₋₁₅ × SST ₋₃₀	0,0172	25,88%	-0,0537
SST ₋₁₅ × ∇SST ₋₁₅	0,0717	9,50%	0,0895
SLA	-0,0135	0,00%	-0,0163
SLA ₋₃₀	-0,0121	28,51%	-0,0479
∇SLA	0,0903	56,93%	0,3481
∇SLA ₋₃₀	-0,0693	63,80%	0,2610
∇SLA × SLA ₋₃₀	0,0003	89,81%	0,0084
∇SLA × ∇SLA ₋₃₀	-0,0265	49,24%	-0,1021

Le taux de succès moyen pour ce modèle est de (69,2 ±2,4)%. Bien que ce modèle désigne la plupart des termes de la SLA comme non-significatifs, leur retrait provoque une baisse de 4% du taux de succès (voir le modèle de la section 4.4.3.7). En revanche, le retrait de tous les termes de température ne diminue pas le taux moyen de manière significative (voir le modèle de la section 4.4.3.6). Le retrait de tous les termes de concentrations en chlorophylle-*a* diminue le taux de succès de 5,9% (voir le modèle de la section 4.4.3.8).

Nous avons tenté d'obtenir un modèle plus efficace en ajustant les termes de manière empirique, comme dans le cas du tableau 26. Le tableau 28 décrit un des modèles obtenus. Il fut calculé à partir de 2612 lignes de données. Contrairement à ce qui était annoncé dans le tableau 17 (page 106), les gradients de SLA et de CHL apparaissant dans ce modèle ont été calculés à partir d'un opérateur de taille 3×3 classique. La réduction de la taille de l'opérateur de gradient permet de réduire le nombre de données manquantes, ce qui nous a permis de bénéficier de 1000 données supplémentaires. Les lignes en jaune représentent des termes qui étaient présents dans le tableau 27.

Tableau 28: Coefficients du modèle linéaire utilisant l'ensemble des paramètres

Descripteur	Coefficient	Probabilité $> t $
Constante	-44,9449	0,02%
SLA	-0,2014	0,05%
SLA ²	-0,0001	26,57%
∇ SLA	2,1907	3,16%
∇ SLA ²	-0,0984	0,01%
SLA ₋₃₀	-0,2731	0,00%
SLA ₋₃₀ ²	-0,0001	56,72%
∇ SLA ₋₃₀	-0,9031	30,89%
SST	3,3136	0,00%
SST ²	-0,0497	0,07%
∇ SST	2,5027	0,53%
SST ₋₃₀	-0,3260	65,54%
SST ₋₃₀ ²	-0,0008	95,34%
CHL	0,1151	77,37%
CHL ²	-0,1285	12,37%
CHL ₋₁₅	2,7769	6,30%
∇ CHL ₋₃₀	0,1803	9,01%
SLA \times ∇ SLA	0,0016	66,24%
SLA \times ∇ SLA ₋₃₀	0,0025	48,42%
SLA \times SST	0,0068	0,17%
SLA \times CHL	0,0124	8,62%
SLA ₋₃₀ \times ∇ SLA	-0,0057	10,70%
SLA ₋₃₀ \times ∇ SLA ₋₃₀	0,0039	25,68%
SLA ₋₃₀ \times SST ₋₃₀	0,0106	0,00%
SLA ₋₃₀ \times CHL	0,0221	0,04%
∇ SLA \times SST	-0,0790	3,55%
∇ SLA \times SST ₋₃₀	0,0122	71,84%
∇ SLA \times CHL	-0,1788	2,87%
∇ SLA ₋₃₀ \times SST	-0,0939	0,33%
∇ SLA ₋₃₀ \times SST ₋₃₀	0,1216	0,01%
SST \times CHL ₋₁₅	-0,0484	37,16%
SST ₋₃₀ \times ∇ SST	-0,0779	0,98%
SST ₋₃₀ \times CHL ₋₁₅	-0,0574	24,68%
∇ SST \times ∇ CHL ₋₃₀	-0,1085	2,86%

Le taux de succès moyen pour ce modèle est de $(71,7 \pm 1,7)\%$, soit 2,5% de plus que le modèle précédent. Ce taux de succès a été obtenu en utilisant toutes les données disponibles pour le calcul de la régression, puis en utilisant ces mêmes données pour vérifier le succès des prédictions. Nous avons aussi testé le même modèle linéaire en n'utilisant que la moitié des données pour le calcul de la régression, et l'autre moitié pour vérifier le succès des prédictions. La séparation des données en un jeu d'apprentissage et un jeu de test est effectuée au hasard lors de chaque exécution de la vérification. Plusieurs exécutions ont été tentées. Elles ont toutes donné un taux de succès proche de 70%, ce qui suggère que le modèle a de bonnes capacités de généralisation.

La figure 52 donne la distribution du taux de succès en fonction de la valeur prédite. Elle montre que le modèle du tableau 28 fournit la distribution la mieux répartie parmi les modèles testés, et que l'intervalle de confiance est suffisamment étroit pour conclure que le succès des prédictions s'améliore bien à mesure que les valeurs s'éloignent de zéro.

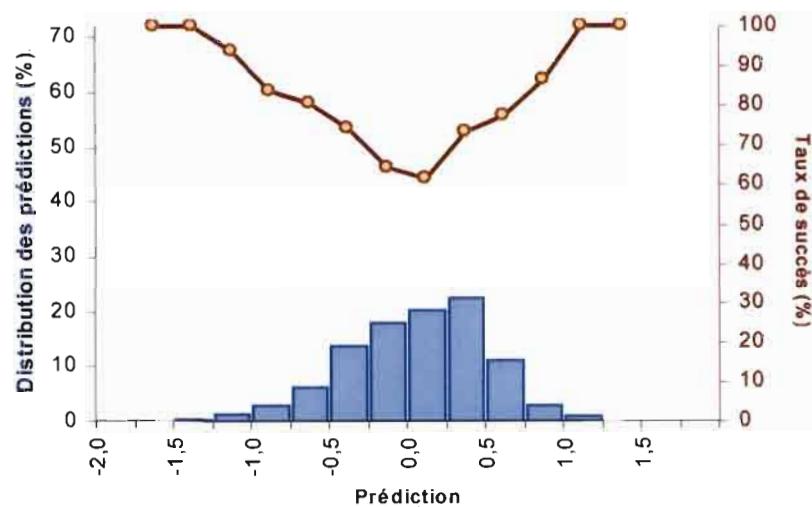
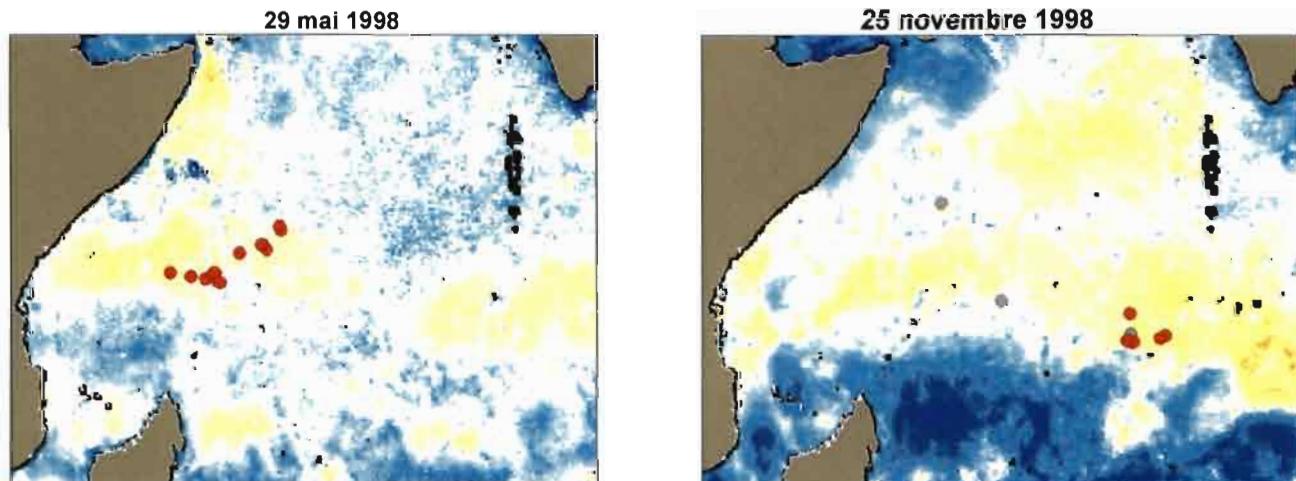


Figure 52: Distribution des prédictions du modèle fondé sur l'ensemble des paramètres

La figure 53 donne des exemples de cartes de potentiel générées à partir du modèle décrit par le tableau 28. Celle du 29 mai 1998 montre un cas où les pêches se sont concentrées dans une région où il y avait probabilité de présence, sans que ces régions jaunes ne soient majoritaires sur l'image. Les autres images présentent un mélange de prospections avec et sans pêches (points rouges et gris respectivement). Sur ces 4 images, les pêches (points rouges) se sont effectivement réalisées dans des régions où la présence de thon était considérée probable (régions jaunes), sauf pour une pêche le 2 juin 2000. Les prospections sans pêches se sont plus souvent réalisées dans des régions classées incertaines (régions blanches) ou avec une légère probabilité de présence (jaune pâle).



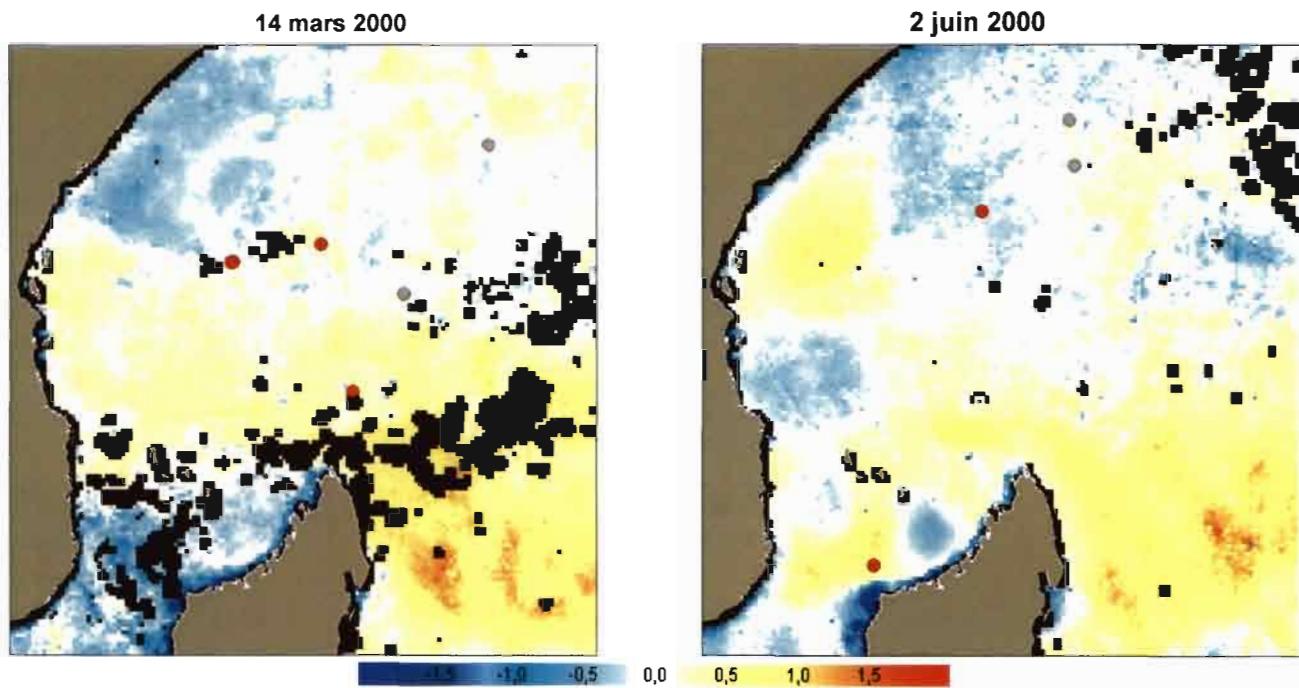


Figure 53: Exemple de carte de potentiel générée à partir de l'ensemble des paramètres

4.5 Modélisation non-linéaire

Le principe de la régression linéaire peut s'exprimer sous la forme d'une projection de la variable à expliquer (y) sur l'hyperplan formé par l'espace des variables explicatives ($x_1, x_2, x_3\dots$). Cependant, si la relation qui lie les variables entre elles s'avère non-linéaire, il est souvent intéressant d'utiliser une régression fonctionnelle adaptée au phénomène étudié. Il s'agit de trouver la fonction f dans l'équation 3 qui minimise l'erreur ε . Pour les régressions linéaires utilisées dans le chapitre 4.4, la forme de cette fonction était donnée à la page 105.

$$y = f(x_1, x_2, x_3, \dots) + \varepsilon \quad \begin{array}{l} y \text{ est fonction de la probabilité de présence} \\ \text{où } x_i \text{ sont des descripteurs du paysage océanique} \\ \varepsilon \text{ est la partie qui n'est pas expliquée par le modèle} \end{array}$$

Équation 3: Modélisation (pas nécessairement linéaire) de la probabilité de présence

Plusieurs types de régression non-linéaire apparaissent dans la littérature (modèles de régression à seuil, modèle bilinéaire, polynomiale...). Les réseaux de neurones artificiels en particulier possèdent des propriétés d'approximateur universel (Cybenko, 1989; Hornik et al., 1989). Les travaux de modélisation détaillés dans ce chapitre portent sur l'étude des capacités de tels modèles à produire une estimation du potentiel halieutique en fonction des variables environnementales retenues.

4.5.1 Les réseaux de neurones artificiels

Les recherches en intelligence artificielle visent à concevoir des systèmes capables de prendre des décisions en fonction de leur environnement. On confond parfois l'intelligence artificielle avec la vie artificielle, dont les recherches visent plutôt à concevoir des systèmes qui imitent certaines propriétés du vivant (par exemple se maintenir et se reproduire). Si, aux yeux du grand public, une certaine aura de magie inquiétante entoure parfois ces deux domaines de recherche, il ne s'agit pourtant que de structures mathématiques. Dans ce chapitre, nous passerons sous silence les questions théoriques ou philosophiques qui accompagnent ces recherches pour n'aborder que les aspects pratiques nécessaires à la programmation d'un réseau de neurones.

Chez les organismes biologiques, les cerveaux forment des réseaux de neurones bien plus élaborés que les réseaux dont il sera question dans ce chapitre. Les dendrites et les neurones se connectent entre eux d'une façon infiniment plus complexe. Les équations exposées ici s'appliquent au cas relativement simple d'un réseau idéalisé. Elles ne sont pas réalistes pour décrire le fonctionnement d'un réseau de neurones biologiques. Pour cette raison, nous n'utilisons le terme de « neurones » que parce qu'il est largement répandu. En aucun cas nos « neurones » mathématiques ne prétendent imiter de façon réaliste les neurones biologiques. Pour ne pas pousser plus loin la confusion, nous bannirons de notre discussion le terme de « synapses » ; nous préférerons parler de connections entre « neurones ».

Il existe une multitude de façons de connecter des neurones entre eux. Certains réseaux appartiennent à la catégorie des réseaux à propagation directe (*feed-forward*), une des catégories les plus étudiées dans la littérature (Hertz et al., 1991). Un des réseaux les plus simples de cette catégorie est le *perceptron multicouches* (Rosenblatt, 1962), dont la figure 54 illustre un exemple. Dans ce type de réseau, chaque neurone est connecté à tous les neurones de la couche précédente. On soumet un problème au réseau en donnant une valeur numérique à chaque neurone de la couche d'entrées. Ces neurones d'entrées ne faisant aucun calcul avec les valeurs qu'on leur donne, on ne considère pas vraiment qu'ils forment une couche du réseau. Dans la première « véritable » couche, chaque neurone utilisera les valeurs de tous les neurones de la couche d'entrées pour calculer sa propre valeur. Une fois cette étape complétée, chaque neurone de la seconde couche utilisera les valeurs de la première couche pour calculer leurs propres valeurs. Le processus se propage ainsi de couche en couche jusqu'à la sortie.

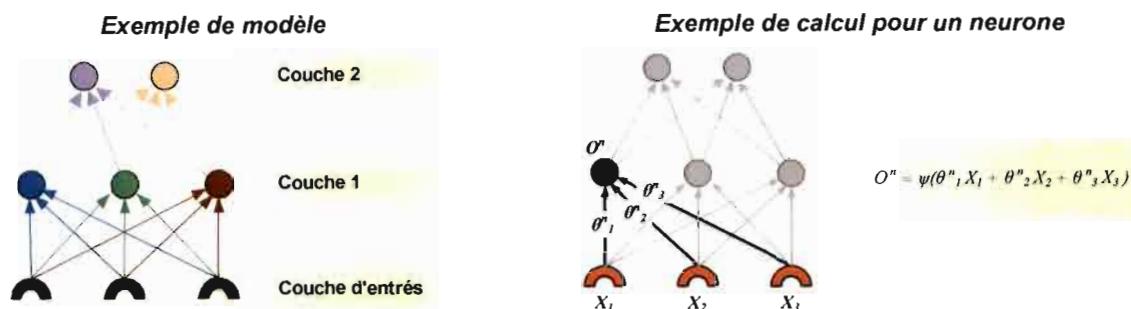


Figure 54: Exemple de réseau de neurones artificiels de type *feed-forward* à deux couches

On appelle couches cachées toutes les couches qui se trouvent entre la couche des entrées et la couche des sorties. Elles tiennent leur qualificatif de « caché » du fait qu'elles n'ont aucun contact avec l'extérieur du réseau. Si on utilisait le réseau comme une boîte noire, on ignorerait leur présence.

Toutes les équations exposées dans ce chapitre font intervenir une *fonction d'activation*, symbolisée par ψ . Cette fonction d'activation est généralement de forme sigmoïdale. De telles fonctions peuvent recevoir une valeur quelconque en entrée, mais retournent toujours une valeur comprise entre deux bornes en sortie. La figure 55 illustre deux sigmoïdes couramment utilisés. La première retourne toujours des valeurs comprises entre 0 et 1, tandis que la seconde (la tangente hyperbolique) retourne des valeurs comprises entre -1 et +1.

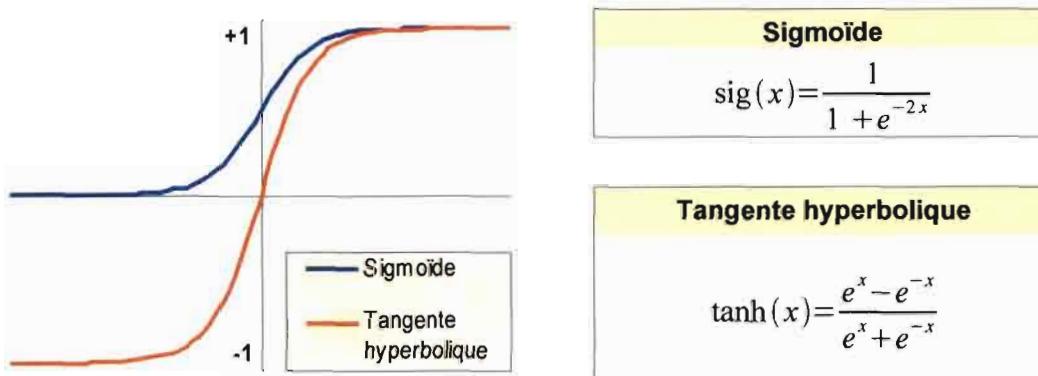


Figure 55: Illustration de deux sigmoïdes couramment utilisées comme fonction d'activation ψ

Il a été démontré que sans fonction d'activation non-linéaire, les réseaux ne peuvent résoudre qu'un nombre assez restreint de problèmes (Minsky et Papert, 1969). Ces réseaux linéaires ont donc peu d'utilité dans la pratique. Ajouter des neurones n'y change rien, car additionner des relations linéaires donne toujours une relation linéaire. Quel que soit le nombre de neurones, un réseau sans fonction d'activation non-linéaire équivaut toujours à un réseau avec une seule couche.

En introduisant un caractère non-linéaire, les fonctions d'activations augmentent les possibilités du réseau. Elles rendent utiles les couches cachées qui autrement auraient été superflues. Les fonctions de forme sigmoïdale ont aussi la propriété de présenter une pente plus abrupte pour les valeurs près de zéro. Lors de la phase d'apprentissage du réseau, cette propriété aura pour effet d'appliquer des corrections plus importantes sur les neurones qui « hésitent » encore sur les valeurs à produire. Ce type de fonction a de plus des propriétés mathématiques intéressantes qui permettent, avec un nombre restreint de paramètres, d'approximer des fonctions complexes (Leshno et al., 1993; Barron, 1993).

Parmi la multitude de façons de connecter des neurones entre eux, on pourrait imaginer un réseau où tous les neurones sont connectés avec tous les autres⁹⁰. Cependant, les techniques d'apprentissage des réseaux avec boucles (appelés aussi réseaux récurrents) posent encore des problèmes de robustesse et de convergence. Aussi nous contenterons-nous dans cette thèse d'étudier l'apport d'un certain type de réseau à propagation directe dont l'apprentissage est maintenant bien maîtrisé.

⁹⁰ Ce cas est effectivement étudié. Il représente une étape importante dans la compréhension des réseaux de neurones, un peu comme l'atome de Bohr en physique atomique.

Au sein de la famille des réseaux à propagation directe, les réseaux *sigma-pi* constituent une alternative intéressante pour modéliser des phénomènes fortement non-linéaires (Shin et Ghosh, 1991). Ils sont en particulier capables de prendre en compte les corrélations d'ordre supérieur entre les entrées. Les possibilités d'approximation s'en trouvent augmentées, tout en réduisant le nombre de neurones. La figure 56 illustre un exemple de réseau sigma-pi. Il comporte des neurones d'entrées plus un neurone dont l'entrée est toujours égale à 1. Ce dernier neurone d'entrée (aussi appelé biais) permet d'introduire une constante dans les combinaisons linéaires des variables. Les neurones des couches suivantes sont constitués de dendrites (Σ) qui opèrent une combinaison linéaire des sorties des neurones auxquels elles sont connectées. Ils effectuent ensuite un produit (Π) des sorties de leurs dendrites, puis appliquent une fonction de transfert ψ ou φ . Ces neurones sont symbolisés dans la figure 56 par $\Pi | \psi$ ou $\Pi | \varphi$, selon la fonction de transfert appliquée après le produit. Pour les neurones des couches cachées, la fonction d'activation ψ est généralement du type sigmoïdale (figure 55). Pour le neurone de sortie, cette fonction φ peut être linéaire ou sigmoïdale suivant le type de problème que l'on tente de résoudre. Dans le cas de la régression linéaire de variables continues, cette fonction est souvent prise linéaire. Dans le cas de la classification par exemple, cette fonction peut-être sigmoïdale.

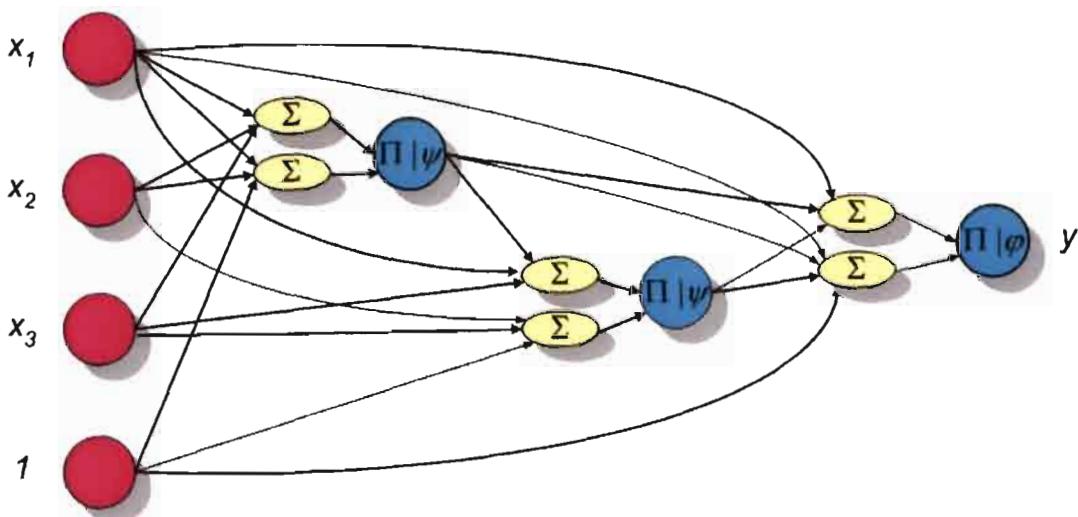


Figure 56: Exemple de réseau sigma-pi

A l'instar de son ancêtre le perceptron multicouches (figure 54), ce modèle a des propriétés d'approximation universelle. Il se distingue du perceptron notamment par le produit des dendrites (Π). La régression non-linéaire s'appuyant sur un modèle neuronal sigma-pi peut s'écrire sous la forme mathématique suivante :

$$Y = \varphi \left(\prod_{d=1}^D O_d^{N+1} \right) \quad \text{où}$$

$$O_d^n = \sum_{i=1}^P \theta_i^n X_i + \sum_{j=1}^{n-1} \theta_{P+j}^n \psi \left(\prod_{d=1}^D O_d^j \right)$$

- Y est la sortie du réseau (supposé ici scalaire) ;
- P est le nombre de variables d'entrées ;
- N est le nombre de neurones cachés ;
- ψ est une fonction d'activation du type sigmoïdale ;
- φ est une fonction identité ou sigmoïde ;
- θ est le poids d'une connexion. Le vecteur de paramètres est $(\theta_p^n)_{1 \leq n \leq N+1, 1 \leq p \leq P+n}$.

Équation 4: Sortie du réseau de neurones sigma-pi

4.5.2 Apprentissage supervisé du réseau

Pour obtenir du réseau qu'il produise les valeurs que l'on désire, il faut donner les bons poids aux connections entre les neurones (θ_p^n dans la figure 54). On y parvient en soumettant le réseau à une phase d'apprentissage pendant laquelle on lui spécifie à la fois les valeurs d'entrées et les valeurs de sorties qu'il devrait obtenir. Les poids appropriés sont alors calculés pour l'ensemble du réseau, y compris pour les

couches cachées. L'algorithme effectuant cette tâche est connue en anglais sous le nom de *back-propagation*. Il en existe de multiples variantes, mais cette section ne présente que la plus classique.

On peut initialiser le réseau en donnant à ses poids des valeurs aléatoires. A partir d'un vecteur d'entrée X qu'on lui spécifiera, le réseau calculera évidemment un vecteur de sortie Y qui sera très différent de celui qu'on attendait. La différence entre la valeur calculée et la valeur attendue constitue l'erreur du réseau. L'idée de la méthode d'apprentissage par *back-propagation* consiste à modifier les poids de chaque connexion par un facteur proportionnel à l'erreur. Plus spécifiquement, la correction à apporter au poids de la connexion entre un neurone n de la couche de sortie et un neurone p de la couche précédente est la suivante :

$$c_{n,p} = \alpha \psi'(O_n) \psi(t_n - O_n) X_p \quad \begin{array}{l} \text{où } \psi' \\ \text{est la dérivée de la fonction d'activation } \psi; \\ t_n \\ \text{est la valeur attendue pour le neurone } n; \\ \alpha \\ \text{est le facteur d'apprentissage.} \end{array}$$

Équation 5: Correction à apporter aux poids d'une connexion entre deux neurones.

Lorsque α est suffisamment petit, l'application de ces corrections ajuste les poids des connections vers des valeurs qui minimisent l'erreur (Widrow et Stearns, 1985). Ce type d'algorithme permet au modèle d'avoir une capacité d'apprentissage ; la précision des prédictions s'améliore au fur et à mesure que l'on présente des données en entrées et à la sortie du réseau.

4.5.3 Application d'un réseau de neurones artificiels aux données de pêches

Nous avons construit un réseau de neurones avec suffisamment d'entrées pour recevoir les descripteurs sélectionnés, et une sortie représentant le « succès » de la pêche. Comme dans le cas des modèles linéaires (chapitre 4.4), ce succès n'est pas représenté par le nombre d'individus capturés. Il est fonction de la probabilité de présence, mais n'est pas non plus une mesure directe de celle-ci.

Afin de sélectionner une configuration du réseau adaptée au problème de l'estimation du potentiel halieutique, une méthodologie d'identification de modèle a été utilisée (Mangeas, 1996). Cette méthodologie, appelée SSM (*Statistical Stepwise Method*), est basée sur des tests statistiques successifs de nullité des connections (non significativité des poids). Les connections les moins pertinentes sont automatiquement éliminées. Pour ce modèle, nous avons utilisé les mêmes descripteurs du paysage océanique que pour le modèle linéaire du tableau 27 (page 129), à l'exception des termes croisés (par exemple SST□□□ × □SST□□□). Nous avions donc 11 descripteurs plus un biais. Notre modèle initial comprenait 200 connections, et l'algorithme SSM en a éliminé jusqu'à obtenir au final un modèle de 98 connections. Ce dernier est mieux adapté au problème considéré et surtout plus robuste au sur-apprentissage. La figure 57 montre l'architecture finale.

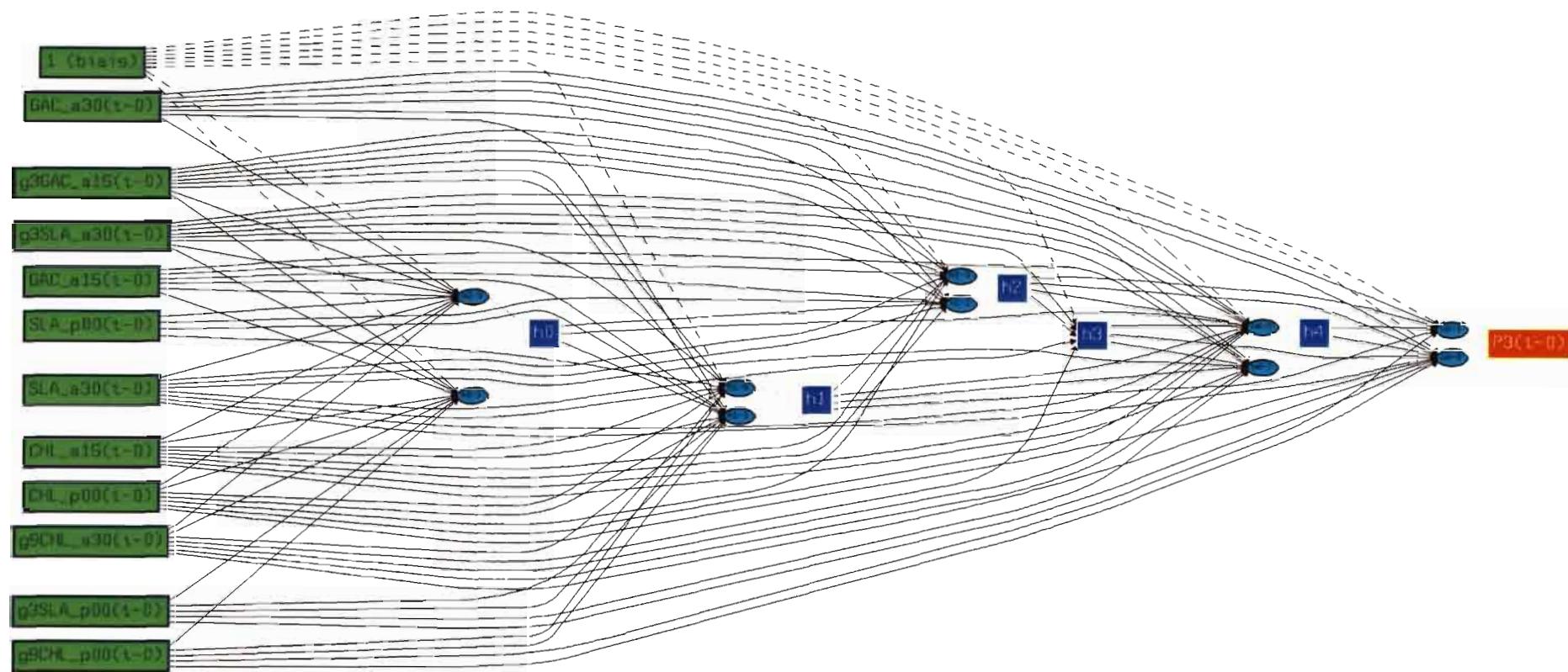


Figure 57: Exemple de réseau de neurones expérimenté sur les données de pêches

Pour tous les neurones du réseau (incluant le neurone de sortie), nous avons utilisé une fonction d'activation tangente hyperbolique. Cette fonction produit des valeurs dans l'intervalle [-1,1], soit le même intervalle que celui que visaient les modèles linéaires du chapitre 4.4. La sortie ainsi obtenue a un caractère binaire, qu'on pourrait décrire par une probabilité de présence (+1) ou d'absence (-1) de thons.

Ce réseau de neurones obtient un taux de succès d'environ 80%. À titre de comparaison, notre meilleur modèle linéaire (section 4.4.3.9) obtenait un taux de succès d'environ 72%. La distribution des valeurs de sorties du réseau de neurones (figure 58) présente un aspect différent de celui des modèles linéaires (par exemple la figure 52). Les valeurs près de 0 sont moins nombreuses, tandis que les valeurs près de ± 1 sont majoritaires. Cette dernière observation s'explique par la forme sigmoïdale (figure 55) de la fonction de transfert du dernier neurone. Toutes les valeurs positives, même les plus élevées, sont ramenées à une valeur qui ne dépasse jamais 1. Ce comportement contraste avec celui des modèles linéaires, qui n'imposent aucune limite sur les valeurs de sorties même si les modèles essaient de reproduire des sorties entre -1 et +1. Le comportement du réseau de neurones revient à accumuler dans la dernière barre de l'histogramme toutes les valeurs supérieures à 1 des modèles linéaires. Le même raisonnement s'applique pour les valeurs négatives.

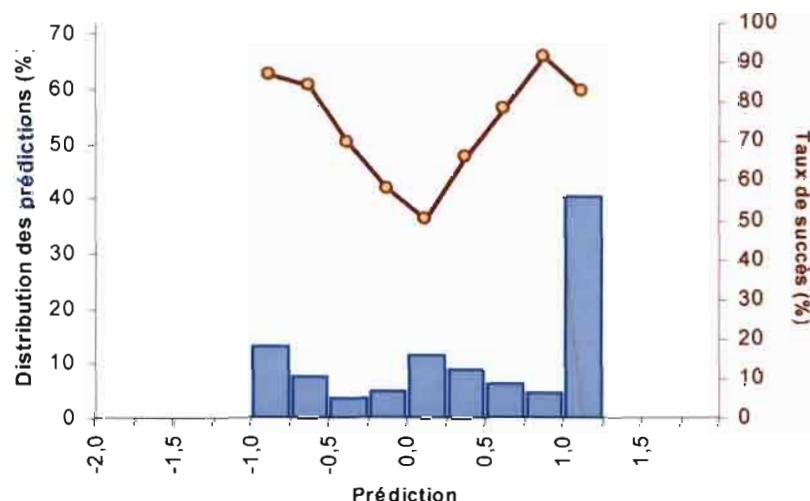


Figure 58: Distribution des prédictions du modèle basé sur un réseau de neurone

Le réseau de neurones a été testé en utilisant 70% des données pour l'apprentissage, et en comparant les prédictions sur les 30% de données restant. Les données d'apprentissage et de tests ont été tirées aléatoirement. Les résultats étant sensiblement équivalents (environ 80% de succès) sur les deux bases, le modèle semble avoir de bonnes capacités de généralisation. Le gain qu'apporte le réseau de neurones par rapport aux modèles linéaires indique qu'il existe des corrélations non-linéaires (autres que celles qui sont ajoutées dans les modèles) entre les variables.

Nous n'avons pas utilisé la latitude dans ce réseau afin de rester cohérent avec les modèles linéaires des sections précédentes. Toutefois, un réseau de neurones saurait utiliser la latitude (et peut-être la longitude) mieux que les modèles linéaires pour prendre en compte les variations de la dynamique selon les régions. Cela revient à faire plusieurs modèles pour différentes régions géographiques (Mangeas, 1996).

Chapitre 5

Résumé et conclusion

5.1 Résumé des résultats statistiques

Nous avons tenté d'estimer une probabilité de présence ou d'absence de thons en fonction du paysage océanique. Cette étude ne distingue pas les différentes espèces de thon, dès lors que les individus ont plus de 10 kg (section 4.1.1 à la page 90) et ont été observés en bancs libres. Trois autres catégories de données de pêches ont été explorées, mais de manière plus superficielle. Les quatre types de données de pêches utilisées sont :

- Moins de 10 kg en bancs libres } catégorie C1
- Moins de 10 kg en bancs associés }
- Plus de 10 kg en bancs libres } catégorie C3
- Plus de 10 kg en bancs associés }

Les descripteurs du paysage océanique ont été dérivés à partir de différentes combinaisons d'un paramètre environnemental, d'une opération et d'une position relative (section 4.1.2 à la page 91). Les principaux paramètres environnementaux utilisés sont :

- Concentration en chlorophylle-*a* (CHL)
- Température de surface GAC (SST)
- Anomalie de la hauteur de l'eau (SLA)

Le pompage d'Ekman et les courants géostrophiques ont été omis dans les modèles linéaires (discutés plus bas) à cause du faible nombre de données disponibles (section 4.4.1 à la page 105). Les corrélations de Spearman ont été calculées entre ces paramètres environnementaux et les quantités de poissons pêchés (chapitre à la page). Elles ont permis de dégager deux remarques :

- Les bancs associés ont des corrélations de même signe que les bancs libres, mais plus faibles. On sait que la présence d'objets flottants et autres dispositifs d'agrégations est un facteur déterminant dans le comportement du thon (section 1.4.4 à la page 26), probablement au détriment des paramètres environnementaux testés (CHL, SST et SLA).
- Le comportement des petits poissons semble très différent de celui des gros individus, si l'on se réfère aux paramètres environnementaux testés. Le tableau 29 synthétise le signe de quelques-unes des corrélations de Spearman les plus significatives de la figure 35 (le symbole ↗ indique que les quantités pêchées augmentent avec la valeur du paramètre) :

Tableau 29 : Sommaire des corrélations de Spearman de la figure 35

Catégorie	✓CHL	SST	✓SST	SLA	✓SLA
C1 (< 10 kg)	↗	↗	↘	↗	↗
C3 (> 10 kg)	↘	↘	↗	↘	↘

Ce tableau suggère que les gros poissons préfèrent les eaux plus froides (souvent associées à une anomalie de la hauteur de l'eau négative⁹¹), près de fronts thermiques, dans des régions pourvues d'une concentration en chlorophylle-*a* et/ou⁹² d'une hauteur de l'eau relativement uniforme. Les petits poissons se retrouvent eux dans des conditions environnementales opposées : une eau chaude (souvent associée à une anomalie de hauteur de l'eau positive), de température uniforme, près de gradients de la concentration en chlorophylle-*a* et/ou de la hauteur de l'eau. Une partie de ces relations pourrait s'expliquer par la physiologie des gros poissons mieux adaptée à l'exploration des eaux froides que celle des petits poissons (section 1.4.1).

⁹¹ Rappelons qu'une anomalie de la hauteur de l'eau négative est un indicateur de tourbillons cycloniques, donc de divergences pouvant favoriser une remontée de la thermocline.

⁹² L'expression « et/ou » vient du fait que les corrélations calculées dans cette thèse (figure 35) n'indiquent pas comment se comporte le poisson lorsque plusieurs conditions *a priori* favorables sont remplies en même temps. Même si on peut raisonnablement supposer que plusieurs conditions favorables lorsque analysées séparément se renforcent lorsqu'elles se réalisent en même temps, nous n'avons pas effectué de test de corrélation de Spearman pour le vérifier.

Les corrélations de Spearman ne comparent pas l'importance relative des paramètres CHL, SST et SLA. Cette importance relative a été estimée à l'aide de modèles linéaires. Plusieurs de ces modèles ont été construits afin de prédire la présence ou l'absence de bancs libres de thons de plus de 10 kg en fonction des descripteurs retenus (section 4.4.3 à la page 107). Ces modèles sont décrits par les tableaux 19 à 28, et contiennent tous des termes croisant des descripteurs dérivés d'un même paramètre (par exemple CHL × VCHL). En outre, certains de ces modèles (les tableaux 26 et 28) contiennent des termes croisant des descripteurs de paramètres distincts, par exemple CHL × VSLA. Ces derniers modèles ne permettent pas de séparer les contributions respectives des paramètres CHL, SST et SLA.

Un modèle combinant les trois paramètres, et qui contient des croisements de paramètres différents, a été expérimenté (tableau 28). Il prédit correctement l'absence ou la présence de thon dans $(71,7 \pm 1,7)\%$ des cas, où le $\pm 1,7\%$ représente un intervalle de confiance de 95%. Ce taux de succès est maintenu même lorsque l'on ne calcule la régression linéaire que sur la moitié des données et testons le taux de succès sur l'autre moitié. Si les présences ou absences de thons aux positions prospectées étaient prédites de manière totalement aléatoire, le taux de succès aurait été d'environ 50% (section 4.4.2 à la page 107). Le succès de ce modèle linéaire apparaît donc significatif. Rappelons toutefois que la probabilité de trouver du poisson aux positions prospectées n'est pas représentative de la probabilité d'en trouver à des positions aléatoires dans l'océan, car les positions prospectées par les pêcheurs ne sont pas aléatoires (section 1.3.2 à la page 22).

En plus du modèle décrit dans le tableau 28, nous avons aussi testé un modèle dont l'ensemble des termes est l'union des termes utilisés par les modèles utilisant les paramètres CHL, SST et SLA séparément. Ce modèle est décrit par le tableau 27 (page 129) et résumé dans la première ligne du tableau 30. Ce tableau 30 présente un sommaire des taux de succès des principaux modèles testés dans la section 4.4.3 de cette thèse. Dans ce tableau, l'abréviation « SST » désigne la température de surface GAC (*Global Area Coverage*), déterminée plus significative que la température de surface LAC (*Local Area Coverage*) dans le cas des poissons de plus de 10 kg (section 4.4.3.4 à la page 121). On remarque que les modèles ayant les meilleurs taux de succès sont aussi ceux dont les valeurs prédites sont les mieux distribuées (comparer par exemple la figure 47 à la figure 52, les deux extrêmes).

Tableau 30: Sommaire des taux de succès des modèles

CHL	SST	SLA	Effectif	Succès (%)	Référence
✓	✓	✓	1547	69,2 ± 2,4	Section 4.4.3.9 (page 129)
	✓	✓	10945	63,3 ± 0,9	Section 4.4.3.8 (page 127)
✓		✓	1591	69,0 ± 2,3	Section 4.4.3.6 (page 125)
✓	✓		1547	65,2 ± 2,4	Section 4.4.3.7 (page 126)
✓			1591	62,2 ± 2,4	Section 4.4.3.2 (page 114)
	✓		12633	59,2 ± 0,9	Section 4.4.3.3 (page 117)
		✓	12558	62,6 ± 0,9	Section 4.4.3.5 (page 123)

Des trois modèles utilisant un seul paramètre (CHL, SST ou SLA), celui utilisant la température de surface semble le moins efficace. On note aussi qu'en partant du modèle combinant les trois paramètres et en retirant la SST, aucun impact significatif n'est observé sur le taux de succès. Le retrait de la CHL a le plus grand impact, suivi par le retrait de la SLA. Ce caractère non-significatif de la température dans notre étude semble aller à l'encontre de ce qui est communément admis dans la littérature et résumé dans les fiches de la FAO (Pesce et Gentile, 2003). Cette observation est toutefois à prendre avec précaution :

- Les descripteurs utilisés dans les modèles linéaires ont été choisis de manière semi-empirique (section 4.4.1 à la page 105) et ne représentent probablement pas les modèles optimaux. D'autres approches plus rigoureuses devront être explorées plus en détails (par exemple la section 4.5.3).
- Les descripteurs dérivés des coordonnées spatio-temporelles (notamment la latitude et la saison) ont été délibérément exclus de ces modèles. Or, nous savons qu'ils améliorent significativement la pertinence de la température (page 119 dans la section 4.4.3.3).
- Les gradients de température (comme tous les autres gradients dans cette thèse) n'ont été estimés qu'à la position de la prospection. Or, les gradients peuvent être des régions turbides. Le thon se situera typiquement à proximité du gradient plutôt que directement dedans. L'estimation des descripteurs avec un décalage spatial, déjà prévue dans la partie informatique de cette thèse (page 80 dans la

section 3.5.3), devra être poursuivie. La pertinence du décalage spatial dépendra toutefois de la précision des coordonnées de prospection (figure 27 à la page 91).

- Les analyses ont été faites sur des données de pêches. Plusieurs flottilles disposaient de cartes de température lors de leurs prospections. Si l'expérience des marins leur a permis d'utiliser ces cartes au mieux et de ne prospecter que les régions dans lesquelles les conditions de température étaient favorables, alors les modèles linéaires auront plus de difficultés à utiliser ce paramètre puisque des prospections sans pêche ne seront pas corrélées à des températures défavorables.

Notons toutefois que, dans le cas de l'albacore, (Cayré, 1991) avait déjà soulevé un doute sur la pertinence des cartes de température de surface en l'absence de gradients thermiques prononcés. En effet, l'albacore passe relativement peu de temps dans la couche de surface (0 – 20 mètres). Or, cette espèce représente 41% des tonnages dans nos données (figure 26).

Les résultats de ces modèles linéaires résumés dans le tableau 30 suggèrent que l'anomalie de la hauteur de l'eau est un paramètre pertinent en halieutique appliquée à la pêche thonière. Les arguments en faveur de la SLA sont :

- Le taux de succès du modèle utilisant la SLA est aussi élevé que celui utilisant la CHL.
- Les données de SLA sont au moins 6 fois plus nombreuses que les données de CHL pour la plage de temps étudiée (années 1991 à 2001). Cette différence s'explique par la couverture temporelle plus grande (celle des données de CHL ne commencent qu'en 1997) et par la capacité du capteur de SLA à effectuer ses coups de sonde malgré la couverture nuageuse.
- Parmi toutes les paires possibles des paramètres CHL, SST et SLA, la paire CHL – SLA est celle qui marche le mieux. Elle produit un taux de succès proche du maximum que nous ayons obtenu dans cette thèse. L'ajout de la SST à ce modèle ne semble pas avoir d'impact significatif.

À partir des modèles de la section 4.4.3, on a généré plus de 10 000 cartes de potentiel de pêches. Les figures 42, 45, 48, 50, 51 et 53 en sont des échantillons. Ces cartes couvrent une région allant de 30°S à 30°N et 30°E à 90°E. Comme mentionné dans la section 4.4.3.2, seulement 34% de la surface océanique des cartes de potentiel ont fait l'objet d'au moins une prospection (figure 33 à la page 96). Les potentiels de pêches affichés à l'est de Madagascar sont des extrapolations, car nous avons peu de données dans cette région (figure 39 à la page 111). Ces extrapolations sont d'autant plus hasardeuses que les modèles linéaires utilisent peut-être des paramètres environnementaux afin de discriminer des régions géographiques (par exemple l'équateur et le canal du Mozambique) pour des raisons que nous ne mesurons pas (page 112 dans la section 4.4.3.1).

Ces cartes de potentiels permettraient de prédire quelles sont, parmi les positions prospectées par un senneur, celles où l'on a le plus de chance de trouver des thons. Les valeurs représentées par ces cartes ne sont toutefois pas des probabilités de pêche. Ce sont les valeurs de sorties des modèles, dans un intervalle de l'ordre de -1 à +1 (les valeurs exactes de l'intervalle dépendent des extrapolations effectuées par le modèle utilisé). Ces valeurs de sorties pourraient être transformées en une probabilité de présence en utilisant par exemple la figure 52 (page 131) comme une table des correspondances. Mais ces probabilités de présence, établies pour les positions prospectées, ne sont pas nécessairement généralisables à l'ensemble d'une carte puisque les pêcheurs ne prospectent pas au hasard. Si en plus ces cartes devenaient disponibles aux pêcheurs, elles pourraient modifier leurs stratégies de prospection et par le fait même l'échantillonnage qui sert à établir les modèles. Cette rétroaction entre les résultats statistiques et l'échantillonnage pourrait donc accentuer le biais dans le calcul des probabilités de présence.

La mise à disposition de cartes de potentiels pose le problème de la surexploitation de la ressource. S'il existait une organisation internationale capable d'imposer le respect de quotas, les cartes de potentiels permettraient aux pêcheurs de consommer moins de fuel en prospection pour l'atteinte de leurs quotas. Cela aurait le double avantage d'augmenter leurs marges tout en diminuant l'émission des gaz issue des combustibles fossiles. En l'absence de quota toutefois, il est probable que plusieurs chercheront à pêcher plus. Pour cette raison, ces images s'adressent aux gestionnaires. Si elles se révèlent efficaces dans la pratique, elles pourraient servir à :

- Diminuer les efforts de pêches sur objets flottants en les redirigeant vers les bancs libres
L'efficacité des dispositifs concentrateurs de poissons a incité un nombre croissant de pêcheurs à y concentrer leurs efforts. Or, nous avons mentionné à la page 12 que les pêches sur objets flottants

semblent représenter un danger pour le recrutement du thon obèse et de l'albacore. Si les cartes de potentiel de pêche s'avèrent suffisamment efficaces pour « concurrencer » les dispositifs concentrateurs de poissons, la valeur économique des plus gros poissons incitera peut-être certains pêcheurs à diminuer la pression sur les objets flottants.

- **Diminuer les efforts de pêches sur une catégorie en les redirigeant vers une autre**

Différentes cartes pour différentes catégories (notamment C1 et C3) pourraient être diffusées de manière contrôlée afin de diriger les efforts de pêches vers une catégorie pour en soulager une autre. Par exemple, la catégorie C3 est constituée majoritairement d'albacores (figure 26 à la page 90), une espèce considérée surexploitée par la commission des thons de l'océan Indien. La catégorie C1 en revanche est constituée majoritairement de listaos (figure 26), une espèce considérée plus résistante à la surexploitation (voir page 12). Toutefois, les bancs de listao sont souvent associés aux albacores juvéniles, ce qui risque de rendre difficile l'application d'une pêcherie sélective.

5.2 Travaux futurs

Parmi les nombreuses améliorations possibles à cette étude, on peut citer deux catégories : 1) expérimenter d'autres descripteurs du paysage océanique et 2) expérimenter d'autres types de modèles. En particulier, les réseaux bayésiens pourraient être bien adaptés à la problématique de cette thèse.

5.2.1 Nouveaux descripteurs du paysage océanique à considérer

- Les corrélations de Spearman effectuées sur l'amplitude du courant géostrophique (figure 35 à la page 102) suggèrent que ce paramètre est à explorer d'avantage. Il est le seul paramètre testé à présenter une variation remarquable en fonction du décalage temporel.
- Il est possible que les anomalies de température soient plus pertinentes que les valeurs directes de température. Les anomalies permettraient d'atténuer les biais dus aux différences de latitude. Par exemple une température de 30°C n'a pas la même signification à l'équateur qu'à 40°S ; une anomalie de 1°C peut peut-être être utilisée plus facilement par les modèles sur l'ensemble de la zone d'étude.
- La section 4.4.3.4 suggère que des paramètres dérivés des coordonnées spatio-temporelles seraient pertinents. Les modèles peuvent l'utiliser par exemple pour pondérer les températures directes (par opposition aux anomalies, voir le point précédent) en fonction de la latitude et de la saison. En outre, un réseau de neurones artificiels pourrait les utiliser pour apprendre à reproduire des dynamiques différentes selon les régions, notamment le canal du Mozambique (surtout la partie nord), la région entre la Somalie et les Seychelles, la bande équatoriale et le golfe d'Oman.
- Dans cette étude, les positions relatives (telles que utilisées dans l'équation 2 à la page 91) ne comprenaient qu'une composante temporelle. Le concept devrait être étendu à l'exploration spatiale des valeurs autour de la position prospectée. La composante spatiale intervient par exemple si les thons prospectent à proximité des fronts thermiques plutôt que dedans.
- Le mode de calcul des gradients a besoin d'être affiné. D'autres techniques qu'un opérateur isotrope ou de Sobel peuvent être testées. Par exemple on pourrait calculer les gradients entre tous les pixels dans une certaine zone, et retenir la valeur médiane.
- Cette étude n'a pas du tout pris en compte le nombre d'heures qu'a duré la pêche ou la prospection (chapitre 2.1 à la page 33), ni le trajet accompli depuis la dernière position relevée. Ces informations pourraient être utilisées pour affiner les conditions environnementales des régions prospectées sans pêche.
- Outre les courants géostrophiques et le pompage d'Ekman que nous avons écartés, d'autres paramètres seraient à prendre en compte comme entrées des modèles de prédiction du potentiel de pêches :
 - les sorties des modèles de production primaire ;
 - la bathymétrie et/ou la distance à la côte la plus proche ;
 - le champ magnétique terrestre (on croit que les thons l'utilisent pour s'orienter).

5.2.2 Réseaux bayésiens

Les modèles linéaires et le réseau de neurones présentés dans cette thèse sont parvenus à des taux de succès significatifs grâce à une exceptionnelle abondance de données dans le sud-ouest de l'océan Indien. Ce succès peut être difficile à reproduire dans des régions où les données sont moins abondantes, en raison de deux limitations importantes. La seconde limitation est particulièrement sévère, et découle en fait de la première :

- Nos modèles linéaires ne prennent pas en compte l'incertitude sur les données. Or, les données utilisées pour l'apprentissage ne sont pas toutes de qualité égale. Par exemple la présence de nuages font paraître certaines valeurs de température légèrement plus basses que ce qu'elles devraient être ; des valeurs de synthèses hebdomadaires sont remplacées par des synthèses mensuelles lorsque les premières sont absentes à une position donnée ; les anomalies de la hauteur de l'eau ne sont pas toujours dérivées du même ensemble de satellites ; la valeur d'un pixel peut être la moyenne des pixels avoisinants lorsqu'une donnée manquante est entourée de données valides ; les positions près

du centre des voxels (la troisième dimension du voxel étant le temps) seront moins sujettes aux interpolations que les valeurs près des bords, *etc.*

- Les modèles testés dans cette thèse exigent que tous les descripteurs soient présents. S'il manque ne serait-ce qu'un seul descripteur, toute la ligne de données est écartée. C'est ainsi que l'on passe de 12 633 données pour le modèle qui cause le moins de perte, à 1 547 données pour le modèle le plus complet. Même le modèle le plus permissif exploite bien peu de données en comparaison des 150 273 que contient la base de données initiale. De très nombreuses données avaient dû être écartées par manque d'information sur le caractère libre ou associé des bancs, les espèces visées lors d'un coup nul, *etc.* Disposer d'un modèle capable de fonctionner avec des informations incomplètes est donc hautement souhaitable.

Ces deux problèmes sont liés, car en l'absence d'une donnée on peut spécifier une valeur probable avec une incertitude plus grande. Cette valeur probable peut être spécifiée avec d'autant plus de précision (c'est-à-dire une incertitude plus étroite) qu'elle est corrélée avec les valeurs présentes. C'est le cas de nos paramètres océanographiques ; par exemple la probabilité que la température soit froide est plus élevée si l'anomalie de la hauteur de l'eau est négative. Dans cette situation, l'approche des réseaux bayésiens paraît séduisante.

Un réseau bayésien attend pour chaque paramètre en entrée non pas un scalaire (la valeur exacte), mais plutôt une distribution. Si une valeur est connue avec exactitude, sa distribution est une fonction de Dirac. Dans le cas plus réaliste où elle est issue d'un instrument de mesure doté d'une précision limitée, la distribution peut être gaussienne. L'écart-type de cette distribution peut être plus large pour une valeur extraite d'une synthèse mensuelle que pour une valeur extraite d'une synthèse hebdomadaire par exemple.

Indépendamment des réseaux bayésiens, on peut préparer une base de données contenant les distributions statistiques d'un paramètre en fonction de quelques autres. Par exemple on pourrait exprimer la température (T) en fonction de l'anomalie de la hauteur de l'eau (h), de la latitude (φ) et de la saison (t). Les paramètres d'entrées (h, φ, t) peuvent être des scalaires, tandis que la sortie T est une distribution statistique. Cette distribution peut être déterminée à partir de la totalité des données satellitaires, et non pas seulement les valeurs aux positions de pêche. On exploite ainsi des millions de données totalement ignorées dans cette thèse. Cette distribution peut être fournie en entrée au réseau bayésien lorsque la température est absente tandis que l'anomalie de la hauteur de l'eau est présente. Cette situation est courante puisque la hauteur de l'eau, mesurée par radar, est moins sensible à la présence de nuages.

L'utilisation de distributions plutôt que des scalaires présente donc des avantages séduisants dans le cas des paramètres environnementaux. Ces avantages s'appliquent aussi aux données de pêches. Par exemple nous avons écarté plusieurs données parce que nous ne savions pas si un coup nul visait des petits ou des gros poissons. Or, nous avons vu dans cette thèse que les petits et gros thons semblent préférer des conditions environnementales différentes. Il semble donc plausible que pour les positions où il y a certitude qu'un banc existait mais dont on ignore la nature, un modèle puisse prédire en fonction des paramètres environnementaux si ce banc était constitué plus probablement de petits poissons, de gros poissons ou des deux. Dans ce cas, l'information « présence ou absence de gros poissons » (par exemple) contiendrait une part d'incertitude, comme pour les paramètres environnementaux. Une approche similaire pourrait être tentée pour le caractère libre et associé des bancs.

Annexe

Requêtes SQL utilisées

Les requêtes exposées ici ont été utilisées pour extraire les données de la base halieutique (figure 22 à la page 78). Ces données ont ensuite été introduites dans S+ pour les analyses statistiques.

A.1 Ré-essais

Donne la liste des journées au cours desquelles une marée a fait plus d'un coup de sennes. Le nombre total d'heures passées en mer est calculé (*hMer*). Ce nombre ne devrait jamais dépasser 24 heures. Les enregistrements qui ne respectent pas cette condition seront marqués comme « incohérent » à l'étape A.3.

```
SELECT marée,
       date,
       COUNT(ID) AS compte,
       SUM(Captures.nsennes) AS nsennes,
       SUM(Captures.hMer) AS hMer
  FROM Captures
 GROUP BY marée, date HAVING (compte > 1)
 ORDER BY marée, date;
```

A.2 Tonnages non-répartis

Donne la liste des marées pour lesquelles la somme des captures par catégories ne correspond pas au total attendu. La somme des valeurs apparaissant dans les colonnes *YFTc1* et *YFTc3* devrait être égale à la valeur apparaissant dans la colonne *YFT*, et de même pour *BET*. Ce n'est pas toujours le cas avec les données qui nous ont été fournies. Dans la requête ci-dessous, la somme des erreurs absolues est calculée dans la colonne « *total* ». Les enregistrements qui présentent une trop grande erreur seront marqués « incohérent » à l'étape A.3.

```
SELECT ID,
       marée,
       date,
       Captures.YFT - (YFTc1 + YFTc3) AS YFT,
       Captures.BET - (BETc1 + BETc3) AS BET,
       ABS(YFT) + ABS(BET) AS total
  FROM Captures WHERE (YFT <> 0) OR (BET <> 0)
 ORDER BY ABS(total) DESC;
```

A.3 Incohérences

Cette requête utilise les résultats des étapes précédentes pour produire une liste d'enregistrements à déclarer incohérents. Ces enregistrements seront exclus des requêtes aux étapes A.1 et suivantes. Plus spécifiquement, nous identifions ici tous les enregistrements qui remplissent au moins une des conditions suivantes :

- Des coups de sennes ont été effectués alors que le nombre d'heures en pêche est proche de 0. Aucun enregistrement ne présente cette erreur dans notre jeu de données.
- Le nombre d'heures en pêche est supérieur au nombre d'heures en mer. Cette erreur concerne 146 enregistrements. Il s'agit toutefois du seul type d'erreur dont nous prenons note sans rejeter les enregistrements incriminés. En effet, deux observations suggèrent que l'erreur s'est faite à la saisie des heures : soit le nombre d'heure en mer est 0, ou soit le nombre d'heures en pêche reste inférieur au nombre total d'heures en mer tel que calculé à l'étape A.1. Puisque nous ne calculons pas de prises par unité d'effort, l'exactitude de ces champs n'est pas un pré-requis pour nos travaux.
- La somme du nombre d'heures en mer dans une journée (telle que calculée à l'étape A.1) est supérieure à 24½ heures. Cette erreur concerne 282 enregistrements. Nous tolérons un dépassement d'une demi-heure car il nous paraît suffisamment faible pour ne pas être symptomatique d'un problème grave, alors que 230 enregistrements sont dans cet intervalle.
- La somme des tonnages par catégories diffère de plus de 0,2 tonne de la somme attendue. Cette erreur concerne 5 enregistrements, avec des erreurs variant de 6,8 à 65,2 tonnes.

- Le nombre de calées est supérieur au nombre de coups de sennes. Aucun enregistrement ne présente cette erreur dans notre jeu de données.

La clause WHERE de la requête ci-dessous exprime en langage SQL tous les points que nous venons d'énumérer, dans le même ordre. Le nombre d'enregistrements sélectionnés s'élève à 437.

```

SELECT Captures.ID AS ID,
       Captures.date AS date,
       Captures.nSennes AS nSennes,
       Captures.hPêche AS hPêche,
       Réessais.hMer AS hMer,
       "Tonnages non-répartis".YFT AS YFT,
       "Tonnages non-répartis".BET AS BET,
       nCalées > nSennes AS surnombre,
       incohérence AS incohérence
  FROM (Captures LEFT JOIN Réessais ON (Captures.marée = Réessais.marée) AND
        (Captures.date = Réessais.date))
 LEFT JOIN "Tonnages non-répartis" ON Captures.ID = "Tonnages non-répartis".ID
 WHERE (nSennes > 0 AND hPêche <= 0.2) OR
       (Captures.hMer < hPêche) OR
       (Réessais.hMer >= 24.5) OR
       (YFT > 0.2) OR
       (BET > 0.2) OR
       (surnombre = TRUE) OR
       (incohérence = TRUE)
 ORDER BY date, hMer;

```

A.4 Présences par espèces

Cette requête effectue une première sélection sur les données de pêches qui seront utilisées dans le modèle. Elle exclut les enregistrements marqués incohérents (étape A.3), ceux dont l'IFREMER considère la qualité douteuse (qualité >= 3), ainsi que ceux pour lesquels le bateau se déplaçait sans être en pêche (hPeche = 0). Afin de distinguer les bancs libres des bancs non-libres, nous excluons les données pour lesquelles le code d'association est inconnu ou non-spécifié. Nous excluons aussi les coups de sennes sans calées positives, car ils ne nous permettent pas de savoir quelles espèces le pêcheur a vu (chapitre 3.5.3). Enfin, nous incluons un ensemble de données prises à des positions aléatoires (fournisseur = 3), qui seront identifiées comme des positions non-visitées.

```

SELECT Captures.ID, date, x, y,
       ABS(fournisseur<>3) AS visité,
       IIf(Associations.libre,1,0) AS libre,
       IIf(YFT>0 Or YFTc1>0 Or YFTc3>0 Or SKJ>0 Or
           BET>0 Or BETc1>0 Or BETc3>0 Or ALB>0 Or OTH>0, 1, 0)) AS présence,
       ABS(Captures.YFT > 0) AS YFT,
       ABS(Captures.YFTc1 > 0) AS YFTc1,
       ABS(Captures.YFTc3 > 0) AS YFTc3,
       ABS(Captures.SKJ > 0) AS SKJ,
       ABS(Captures.BET > 0) AS BET,
       ABS(Captures.BETc1 > 0) AS BETc1,
       ABS(Captures.BETc3 > 0) AS BETc3,
       ABS(Captures.ALB > 0) AS ALB
  FROM Captures INNER JOIN Associations ON Associations.ID=Captures.association
 WHERE (fournisseur=3) OR
       (fournisseur=1 AND incohérence=FALSE AND qualité<3 AND hPêche>0 AND
        (nSennes=0 OR (nCalées>0 AND Associations.inconnu=FALSE)))
 ORDER BY date;

```

A.5 Présences par catégories

Cette requête se base sur les présences par espèces (A.4), mais en les synthétisant par les catégories C1 et C3.

```

SELECT ID, date, x, y, visite, libre, présence,
       IIf(YFTc1 Or BETc1 Or SKJ, 1, 0) AS C1,
       IIf(YFTc3 Or BETc3, 1, 0) AS C3,
       ALB
  FROM "Présences par espèces";

```

Bibliographie

- A.A. Anganuzzi, K.A. Stobberup et N.J. Webb (*éditeurs*), 1996. Proceedings of the sixth expert consultation on indian ocean tunas. IPTP collective volume . vi. 9 p.
- Barbieri, M.A., E. Yanez et M. Ferias, 1991. La télédétection et la pêche artisanale du germon et de l'espodon au Chili : un cas de transfert de technologie in Durand, Lemoalle et Weber (*éditeurs*), La recherche face à la pêche artisanale. Symp. ORSTOM. - pp. 817-824
- Bard, F.X., J.-M. Stretta et M. Slepoukha, 1985. Les épaves artificielles comme auxiliaires de la pêche thonière en océan atlantique, quel avenir? La pêche maritime. - pp. 655-659
- Barron, A., 1993. Universal approximation bounds for superpositions of a sigmoidal function. IEEE transactions on information theory. 3 - 39 p.
- Block, B.A., H. Dewar, C.J. Farwell et E.D. Prince, 1998. A new satellite technology for tracking the movements of Atlantic bluefin tuna. Proc. National academy of sciences. 95 - pp. 9384-9389
- Block, B.A., H. Dewar, S.B. Blackwell, T.D. Williams, E.D. Prince, C.J. Farwell, A. Boustany, S.L.H. Teo et A. Seitz, 2001. Migratory movements, depth preferences, and thermal biology of Atlantic bluefin tuna. Science. 293 - pp. 1310-1314
- Boehlert, G.W. et A. Genin, 1987. A review of the effects of seamounts on biological processes in B.H. Keating, P. Fryer, R. Batiza et G.W. Boehlert (*éditeurs*), Seamounts, islands, and atolls. Geophy. union. - pp. 319-334
- Borstad, G.A., D.A. Hill, R.C. Kerr et B.S. Nakashima, 1982. Direct digital remote sensing of herring schools. International journal of remote sensing. 13 - pp. 2191-2198
- Brill, R.W., 1994. A review of temperature and oxygen tolerance studies of tunas pertinent to fisheries oceanography, movement models and stock assessments.. Fisheries oceanography. 3 - pp. 204-216
- Brill, R.W., 1996. Selective advantage conferred by the high performance physiology of tunas, billfished and dolphin fish. Comp. Biochem. Physiol.. 113A:1 - pp. 3-15
- Brill, R.W., B.A. Block, C.H. Boggs, K.A. Bigelow, E.V. Freund et D.J. Marcinek, 1999. Horizontal movements and depth distribution of large adult yellowfin tuna near the Hawaiian islands, recorded using ultrasonic telemetry: implication for the physiological ecology of pelagic fishes. Marine biology. 133 - pp. 395-408
- Brill, R.W., D.B. Holts, R.K.C. Chang, S. Sullivan, H. Dewar et F.G. Carey, 1993. Vertical and horizontal movements of striped marlin (Tetrapturus audax) near the Hawaiian Islands, determined by ultrasonic telemetry, with simultaneous measurement of oceanic currents. Marine biology. 117 - pp. 567-574
- Cayré, P., 1991. Behaviour of yellowfin tuna and skipjack tuna around fish aggregating devices in the Comoros islands as determined by ultrasonic tagging. Aquatic living resources. 4 - pp. 1-12
- Coakley, J.A. et F.P. Bretherton, 1982. Cloud cover from high-resolution scanner data: detecting and allowing for partially filled fields of view. Journal of geophysical research. 87 - pp. 4917-4932
- Collignon, J., 1991. Écologie et biologie marines: introduction à l'halieutique. Masson. 298 p.
- Cram, D.L. et I. Hampton, 1976. A proposed arial/acoustic strategy for pelagic fish stock assessemment.. J. cons. int. explor.. 37(1) - pp. 91-97
- CTOI, 2003. Rapport de la sixième session de la du comité scientifique de la commission des thons de l'océan Indien. Commission des thons de l'océan Indien. - 76 p.
- Cybenko, G., 1989. Approximation by superposition of sigmoidal functions. Mathematics of control, signals, and systems. 2 - pp. 303-314
- Dagorn, L., E. Josse et P. Bach, 2000. Individual differences in horizontal movements of yellowfin tuna in nearshore areas in French Polynesia, determined using ultrasonic telemetry. Aquatic living resources. 13 - pp. 1-10
- Dagorn, L., E. Josse, P. Bach et A. Bertrand, 2000. Modelling tuna behaviour near floating objects: from individuals to aggregations. Aquatic living resources. 13 - pp. 1-10
- Dagorn, L., F. Menczer, P. Bach et R.J. Vazquez-Cuervo, 2000. Co-evolution of movement behaviours by tropical pelagic predatory fishes in response to prey environment: a simulation model. Ecological

- modelling. 134 - pp. 325-341
- Dagorn, L., P. Bach et E. Josse, 2000. Movement patterns of large bigeye tuna in the open ocean, determined using ultrasonic telemetry. Marine biology. 136 - pp. 361-371
- Desruisseaux, M., M. Petit, L. Gardel, A.G. Ramos, M. Slepoukha, E. Rivière, J. Coca, P. Malard et Y. Basson, . Des phénomènes physiques aux paramètres océanographiques in M. Petit et F. Huyn (éditeurs), Halieutique et environnement océanique : le cas de la pêche palangrière à l'espadon depuis l'Île de La Réunion. in press. -
- Dewar, H. et J. Graham, 1994. Studies of tropical tuna swimming performance in a large water tunnel - kinematics. Journal of experimental biology. 192 - pp. 45-59
- Dizon, A.E., 1977. Effect of dissolved oxygen concentration and salinity on the swimming speed of two species of tuna. Fishery bulletin. 75 - pp. 649-653
- Farrugio, H., 1981. Exploitation et dynamique des populations de thon rouge Atlantico-Méditerranéenne. Université des sciences et techniques du Languedoc. 266 p.
- Fotsing, J.-M., 1998. Paysages de l'Ouest-Cameroun : approche géographique des dynamiques de l'espace rural par télédétection. Université de Paris IV - Sorbonne. 751 p.
- Fréon, P. et L. Dagorn, 2000. Review of fish associative behaviour: toward a generalisation of the meeting point hypothesis. Reviews in fish biology and fisheries. 10 - pp. 183-207
- Fréon, P. et O.A. Misund, 1999. Dynamics of pelagic fish distribution and behaviour: effects on fisheries and stock assessment. Blackwell. 348 p.
- G.H. Tomczak (éditeurs), 1977. Environmental analyses in marine fisheries research - fisheries environmental services. FAO fisheries technical paper . 170. 141 p.
- Garcia, M. et C. Newton, 1997. Current situation, trends and prospects in world capture fisheries in E.K. Pikitch, E.D. Huppert et M.P. Sissenwine (éditeurs), Global trends: fisheries management. AFSS. -
- Gréboval, D. et G. Munro, 1999. Overcapitalization and excess capacity in world fisheries: underlying economics and methods of control, in managing fishing capacity. FAO fisheries technical paper. 386 -
- Hall, M., C. Lennert-Cody, M. Garcia et P. Arenas, 1999. Characteristics of floating objects and their attractiveness for tuna in M.D. Scott, W.H. Bayliff, C.E. Lennert-Cody et K.M. Chaefer (éditeurs), Proceedings of the international workshop on the ecology and fisheries for tunas associated with floating objects. IATTC report. - pp. 396-446
- Herron, R.C. et T.D. Leming, 1989. Satellite-detected fronts and butterfish ageastern Gulf of Mexico. Continental shelf research. 9(6) - pp. 569-588
- Hertz, J., A. Krogh et R.G. Palmer, 1991. Introduction to the theory of neural computation. Addison-Wesley publishing compagny inc.. 327 p.
- Holland, K.N., R.W. Brill et R.K.C. Chang, 1990. Horizontal and vertical movements of yellowfin and bigeye tuna associated with fish aggregating devices. Fishery bulletin. 88 - pp. 493-507
- Holland, K.N., R.W. Brill, R.K.C. Chang et D.A. Fournier, 1992. Physiological and behavioural thermoregulation in bigeye tuna. Nature. 358 - pp. 410-412
- Hornik, K., M. Stinchcombe et H. White, 1989. Multilayer feedforward networks are universal approximators. Neural networks. 2 - pp. 359-366
- J.R. Hunter et J.H. Churnside (éditeurs), 1995. Airborne fishery assessemment technology - a NOAA worshop report. SWFSC admin. rep. . LJ-95-02. 71 p.
- Joseph, J., 2003. Managing fishing capacity of the world tuna fleet. FAO fisheries cicular. 982 - 67 p.
- Josse, E., P. Bach et L. Dagorn, 1998. Simultaneous observations of tuna movements and their prey by sonic tracking and acoustic surveys. Hydrobiologia. 371 - pp. 61-69
- Klimley, A.P. et C.F. Holloway, 1999. School fidelity and homing synchronicity of yellowfin tuna, Thunnus albacares. Marine biology. 133 - pp. 307-317
- Korsmeyer, K.E. et H. Dewar, 2001. Tuna metabolism and energetics in B.A. Block et E.D. Stevens (éditeurs), Tuna : physiology, ecology and evolution. Academic Press. - p. 35-78
- Laurs, R.M., 1989. Application de la télédétection satellitaire au germon du Pacifique Nord, Thunnus alalunga (Bonnerre). FAO fisheries technical paper. 302 - pp. 87-97

- Laurs, R.M., 1989. Review of satellite applications to fisheries in (éditeurs), Proceeding of IGARSS'89, 12th canadian symposium on remote sensing.. - pp 2037-2040
- Le Gall, J.Y., 1989. Télédétection satellitaire et pêches thonières océaniques. Document technique sur les pêches (FAO). **302** - 148 p.
- Lebart, L., A. Morineau et M. Piron, 1995. Statistique exploratoire multidimensionnelle. Dunod. 439 p.
- Legendre, L. et P. Legendre, 1979. Écologie numérique. Tome 1: Le traitement multiple des données écologiques; Tome 2: La structure des données écologiques. Masson.
- Leroy, B., 2003. 7 mois dans la vie d'un thon obèse du Pacifique. Secrétariat général de la communauté du Pacifique. - 21 p.
- Leshno, M., V. Y. Lin, A. Pinkus et S. Schoken, 1993. Multilayer feedforward networks with a nonpolynomial activation function can approximate any function. Neural networks. **6** - pp. 861-867
- Lohofner, R., C. Roden, C. Roger, W. Hoggard et K. Mullin, 1988. Distribution and relative abundance of "baitfish" schools in the Gulf of Mexico : spring and fall 1987. Clupeid purse seine fisheries workshop (septembre 1985). - 15 p.
- Longhurst, A.R., 1976. Vertical migration in D.H. Cushing et J.J. Wash (éditeurs), The ecology of the seas. Blackwell. - pp. 117-137
- Lowe, T.E., R.W. Brill et K.L. Cousins, 2000. Blood oxygen-binding characteristics of bigeye tuna (*Thunnus obesus*), a high-energy-demand teleost that is tolerant of low ambient oxygen. Marine biology. **136** - pp. 1087-1098
- M. Petit et F. Huyn (éditeurs), . Halieutique et environnement océanique : le cas de la pêche palangrière à l'espadon depuis l'Île de La Réunion. (sous presse) ..
- M.J.A. Butler, M.-C. Mouchot, V. Barale et C. Leblanc (éditeurs), 1988. The application of remote sensing technology to marine fisheries : an introductory manual. FAO fisheries technical paper . **295**. 165 p.
- Mangeas, M., 1996. Propriétés statistiques des modèles paramétriques de prévision de séries temporelles - application aux réseaux de neurones à propagation directe. Université de Paris I.
- Marsac, F., M. Petit et J.-M. Stretta, 1987. Radiométrie aérienne et prospection thonière à l'ORSTOM. Méthodologie, bilan et perspective in (éditeurs), Initiations et documents techniques n° 68, (Télédétection 12). ORSTOM. - 33 p.
- Minsky, M.L. et S.A. Papert, 1969. Perceptrons. Cambridge, presse MIT.
- Nakashima, B.S., 1990. Capelin scholl surface area index for NAFO div. 3L during the 1989 spawning season. Northwest Atlantic fisheries organization SCR. **90/59** - N1780
- Nakashima, B.S. et G.A. Borstad, 1993. Detecting and measuring pelagic fish schools using remote sensing techniques. ICES Report C.M. 1993/B:7 Session T. Fish capture committee. - 19 p.
- Oliver, C.W., W.A. Armstrong et J.A. Young, 1994. Developments of an airborne LIDAR system to detect tunas in the eastern tropical Pacific purse-seine fishery. NOAA-TM-NMFS-SWFSC-204.. Southwest fisheries science center. - 67 p.
- Pauly, D., V. Christensen, J. Dalsgaard, R. Froese et F.jr. Torres, 1998. Fishing down marine food webs. Science. **279** - pp. 860-863
- Pesce, M. et A. Gentile, 2003. Biological characteristics of tunas and tuna-like species. [#http://www.fao.org/figis/servlet/static?dom=root&xml=speciesgroup/tunalike.xml#](http://www.fao.org/figis/servlet/static?dom=root&xml=speciesgroup/tunalike.xml)
- Petigas, P., 1996. Geostatistics and their application to fisheries survey data in (éditeurs), Computers in fisheries research. Chapman-Hall. -
- Petit, M., 1984. La pêche des thoniers senneurs dans le Pacifique tropical ouest. La pêche maritime. -
- Petit, M., 1991. Contribution de la télédétection aérospatiale à l'élaboration des bases de l'halieutique opérationnelle: l'exemple des pêcheries thonières tropicales de surface (aspect évaluatif). Université Paris VI. 128 p.
- Petit, M., 1995. Operational halieutics and satellite observation : a research action driven by fishermen managers and scientists in SW Indian Ocean in Gerard et Biarritz (éditeurs), Proceedings of the international symposium on operational oceanography and satellite observations.. -
- Petit, M. et J.-M. Stretta, 1989. Sur le comportement des bancs de thons observés par avion. Documents

- scientifiques ICCAT. **30(2)** - pp. 488-490
- Petit, M. et P. Hazane, 1983. Radiométrie aérienne et prospection thonière dans la ZEE de Nouvelle-Calédonie. Rapport de convention mars 1983. ORSTOM Nouméa. -
- Petit, M., J.-M. Stretta, M. Simier et A. Wadsworth, 1989. Anomalies de surface et pêche thonière. Un potentiel inattendu de SPOT pour la détection des zones de pêche par l'inventaire des hauts-fonds. Mappemonde. **89(3)** - pp. 13-16
- Petit, M., L. Dagorn, P. Lena, M. Slepoukha, A.G. Ramos et J.-M. Stretta, 1994. Oceanic landscape concept and operational fisheries oceanography. Mémoires de l'Institut océanographique. **18** - pp. 85-97
- Power, J.H. et L.N. May jr., 1991. Satellite observed sea-surface temperatures and yellowfin tuna catch and effort in the gulf of Mexico. Fishery bulletin. **89** - pp. 429-439
- R.B. Deriso et W.H. Bayliff (éditeurs), 1991. World meeting on stock assessment of bluefin tunas: strengths and weaknesses. Inter-American tropical tuna commission special report . 7.
- Roger, C., 1994. Relationships among yellowfin and skipjack tuna, their prey-fish and plankton in the tropical western Indian Ocean. Fisheries oceanography. **3:2** - pp. 133-141
- Roithmayr, C.M., 1970. Airborne low-light sensor detects luminescing fish schools at night. Commer. fish. rev.. **32(12)** - pp. 42-51
- Roithmayr, C.M. et F.P. Wittmann, 1973. Low light level sensor developments for marine resource assessment. Mar. tech. soc. 8th annual conf.. - pp. 277-288
- Rose, G.A. et W.C. Leggett, 1993. Use of oceanographic forecasts and echosounders to guide and enhance an inshore gillnet fishery for Atlantic cod. Canadian journal of fisheries and aquatic science. **50** - pp. 2129-2136
- Rosenblatt, F., 1962. Principles of neurodynamics. Spartan.
- Santiago, J., A.G. Ramos et M. Canton, 1993. Teledetección y pesca de atún blanco en NE Atlántico. Análisis de las campañas de pesca de 1990-1992. Gobierno Vasco, Departamento de agricultura y pesca. **51** - 142 p.
- Scherrer, B., 1984. Biostatistique. . 850 p.
- Schott, F.A., M. Dengler et R. Schoenfeldt, 2002. The shallow overturning circulation of the Indian Ocean. Progress in oceanography. **53** - pp. 57-103
- Seki, M.P., R. Lumpkin et P. Flament, 2002. Hawaii cyclonic eddies and blue marlin catches: the case study of the 1995 Hawaiian international billfish tournament. Journal of oceanography. **58** - pp. 739-745
- Shankar, D., P.N. Vinayachandran et A.S. Unnikrishnan, 2002. The monsoon currents in the north Indian Ocean. Progress in oceanography. **52** - pp. 63-120
- Sharp, G.D., 1981. What is a tuna school?. ICCAT symposium. **WP2620** -
- Sharp, G.D., 2001. Tuna oceanography: an applied science in B.A. Block et E.D. Stevens (éditeurs), Tuna: physiology, ecology and evolution. Academic Press. - pp. 345-389
- Shenoi, S., P. Saji et A. Almeida, 1999. Near-surface circulation and kinetic energy in the tropical Indian ocean derived from lagrangian drifters. Journal of marine research. **57** - pp. 885-907
- Shin, Y. et J. Ghosh, 1991. The pi-sigma network: an efficient higher-order neural network for pattern classification and function approximation. Proceedings IJCNN. - pp. 13-18
- Shinomi, H. et H. Tameishi, 1988. Discriminant prediction of formation of saury fishing grounds by satellite infrared imageries. Nippon Suisan Gakkaishi. **7** - pp. 1093-1099
- Simpson, J.J., 1990. On the accurate detection and enhancement of oceanic features observed in satellite data. Remote sensing environment. **33** - pp. 17-33
- Squire, J.L. et H. Krumboltz, 1981. Profiling pelagic fish schools using airborne optical lasers and other remote sensing techniques. Mar. tec. soc. j.. **15(4)** - pp. 27-31
- Stretta, J.-M., 1977. Température de surface et pêche thonière dans la zone frontale du cap Lopez (Atlantique tropical oriental) en juin et juillet 1972, 1974 et 1975. Cahiers ORSTOM série océanographie. **XV(2)** - pp. 163-180
- Stretta, J.-M., 1991. Contribution de la télédétection aérospatiale à l'élaboration des bases de l'halieutique opérationnelle: l'exemple des pêches thonières tropicales de surface (aspect prédictif). Université Paris

VI. 126 p.

- Sund, P.N., M. Blackburn et F. Williams, 1981. Tunas and their environment in the Pacific Ocean: a review. Oceanographical marine biology annual review. **19** - pp. 443-512
- Swain, D.P. et A.F. Sinclair, 1994. Fish distribution and catchability: what is the appropriate measure of distribution?. Canadian journal of fisheries and aquatic science. **51** - pp. 1046-1054
- Thorpe, S.A., 1995. Dynamical process of transfer at the sea surface. Progress in oceanography. **35** - pp. 315-352
- Widrow, B. et S. Stearns, 1985. Adaptive signal processing. Prentice-Hall.
- Wolberg, G., 1990. Digital image warping. IEEE computer society press. pp. 129-131
- Yamanaka, I., S. Ito, K. Niwa, R. Tanabe, Y. Yabuta et S. Chikuni, 1988. The fisheries forecasting system in Japan for coastal pelagic fish. FAO fisheries technical paper. **301, VIII** - 72 p.