THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par
l'Université Toulouse III - Paul Sabatier
Discipline ou spécialité : *Océanographie*

Présentée et soutenue par *Olga Hernández*
Le *21/06/2012*

Titre :

Modélisation eulérienne de l'habitat de ponte et de la dynamique des larves des anchois et sardines dans le système d'upwelling du Pérou

---

JURY
Isabelle Dadou (Présidente)
Pierre Brasseur (Rapporteur)
Christian Mullon (Rapporteur)
Martin Huret (Examinateur)
Vincent Echevin (Examinateur)
Philippe Gaspar (Directeur de thèse)
Patrick Lehodey (Co-directeur de thèse)
Arnaud Bertrand (Co-directeur de thèse)

École doctorale :
Sciences de l'Univers, de l'Environnement et de l'Espace (SDU2E)
Unité de recherche :

Collecte Localisation Satellites (CLS)- Direction Océanographie Spatiale (DOS)
Directeur(s) de Thèse:
Philippe Gaspar, Coordinateur Scientifique (DOS), Toulouse
Patrick Lehodey, Chef du Département Ecosystèmes Marins (DOS), Toulouse
Arnaud Bertrand, Directeur de Recherche, IRD, Sète

Résumé

Le Système du Courant de Humboldt (SCH) au large de la côte péruvienne est le système le plus productif au monde en terme de biomasse de poisson. Les petits poissons pélagiques tels que l'anchois du Pérou (*Engraulis ringens*) et la sardine (*Sardinops sagax*) occupent une place centrale dans son écosystème et représentent une partie importante de la biomasse totale. Historiquement, le système a connu d'importantes variations des populations d'anchois et de sardine, une particularité qui continue d'interpeller les océanologues. Actuellement, l'anchois est l'espèce principale, et soutient la pêcherie monospécifique la plus importante au monde, avec des débarquements annuels représentant environ 10 % des captures mondiales de pêches. La sardine en revanche est absente de l'écosystème du SCH depuis la fin des années 90. Il est ainsi crucial de comprendre les facteurs qui contrôlent les stocks de petits pélagiques le long de leur côte, afin de mieux gérer cette importante ressource.

L'objectif de ce travail de thèse est d'implémenter pour la première fois une approche eulérienne visant à étudier les premiers stades de vie des anchois et sardines dans le SCH, et à comprendre l'influence des paramètres environnementaux et de la variabilité climatique sur la dynamique des populations. Le modèle est adapté d'un modèle eulérien de dynamique spatiale des écosystèmes (SEAPODYM), initialement développé pour les grands poissons pélagiques (comme le thon) à l'échelle du bassin océanique, et utilise le modèle couplé physico-biogéochimique ROMS-PISCES en entrée.

Afin de contraindre les paramètres du modèle, nous avons développé une approche par assimilation de données basée sur un jeu unique de données de densités d'oeufs, larves et adultes, acquis par l'Institut de la Mer du Pérou (IMARPE) dans les eaux du SCH au cours des 40 dernières années. Ces données font l'objet d'une analyse détaillée.

Le modèle permet de tester différents mécanismes susceptibles de contrôler l'habitat de ponte et le recrutement larvaire : température optimale, abondance des proies, présence/absence de prédateurs, et influence des courants sur la rétention et la dispersion. Nous montrons que la combinaison de quelques concepts simples permet d'expliquer raisonnablement la répartition spatiale des oeufs et larves ainsi que la variabilité interannuelle, mais la description actuelle ne permet pas d'expliquer la saisonnalité observée dans les zones côtières.

A l'échelle saisonnière, les résultats suggèrent que l'abondance spatiale et temporelle des proies, ainsi que la présence ou l'absence de prédateurs, contrôlent majoritairement l'habitat de ponte et la dynamique larvaire, tandis que la température, souvent présentée comme un facteur crucial dans la dynamique des premiers stades de vie, semble être de moindre importance. A l'échelle interannuelle, une paramétrisation différente est

obtenue, dans laquelle la température semble jouer un rôle plus important, alors que l'impact des prédateurs devient nul pour l'anchois. Notre modèle prédit une diminution de l'abondance des oeufs et larves pendant le phénomène "El Niño" de 1997-1998, conformément aux observations.

Une étude de sensibilité des principales sources de variabilité et d'incertitudes du modèle suggère quelques pistes d'amélioration de la qualité de nos résultats.

Enfin, certains mécanismes absents du modèle à ce stade de développement, tels que l'influence de l'oxygène, la prise en compte des adultes, et la mortalité par pêche, sont susceptibles d'améliorer les résultats du modèle et devraient être pris en compte à l'avenir.

Mots clés: Anchois, Sardine, *Engraulis ringens*, *Sardinops sagax*, SEAPODYM, modèle de dynamique spatiale des populations, assimilation de données, Estimation du Maximum de Vraisemblance, méthode adjointe, upwelling, Humboldt Current system, Pérou, habitat de ponte

Abstract

The Humboldt Current System (HCS) off the Peruvian coast is the most productive oceanic system in terms of fish biomass, dominated by small pelagic species such as the Peruvian anchovy (*Engraulis ringens*) and sardine (*Sardinops sagax*). Historically, the system has known important variations in the anchovy and sardine populations, a particularity which has questioned oceanographers for many years. Anchovy currently dominates the HCS ecosystem, and sustains the world's most important monospecific fisheries, with annual landings representing about 10% of the world fish catches. The sardine however is absent in the HCS ecosystem since the end of the 1990s. It is therefore crucial to understand the factors which control small pelagics stocks to help in the sustainable management of this important resource.

The goal of the present work is to implement an Eulerian approach to study the early life stages of anchovy and sardine in the HCS, and understand the influence of environmental parameters and climate variability. The model is adapted from the Spatial Eulerian Ecosystem and Population Dynamic Model (SEAPODYM), initially developed for large pelagic fishes (e.g., tuna) at ocean basin scale. It uses the outputs of a regional ROMS-PISCES coupled physical-biogeochemical model as environmental forcing.

In order to constrain model parameters, we developed a data assimilation framework using a unique high record dataset of eggs, larvae and adult abundances, collected by the Instituto del Mar del Peru (IMARPE) in the HCS waters over the last 40 years, which is the object of an in-depth analysis.

This framework allows to test different mechanisms proposed to control fish spawning habitat and larval recruitment : optimal temperature, prey abundance, trade-off between preys and predators, and influence of currents on retention and dispersion. We show that the combination of a few simple concepts can reasonably explain the overall spatial distribution of eggs and larvae and the interannual variability, but the current description is still not sufficient to explain the abundance seasonality observed in coastal areas.

At seasonal time scale, the results suggest that spatial and temporal prey availability, and the presence or absence of predators, mainly control spawning habitat and larvae dynamics, whereas temperature, a controlling factor often used in early life stages studies, seem to be of lesser importance. At interannual scale, a different parameterization is achieved, for which the temperature appears to play a more significant role, whereas the impact of predators become null for anchovy. The model solution predicted a decrease of eggs and larvae abundances during the El Niño event 1997-1998 in accordance with

observations.

A sensitivity study of the main sources of variability and uncertainty of our model and data suggests a few possibilities to improve the accuracy of the model results.

Finally, some mechanisms that are missing from the current model description, such as the role of the oxygen, the dynamics of adults and the fishing mortality, could likely improve the results of our model, and should be taken into account in the future.

Keywords: Anchovy, Sardine, *Engraulis ringens*, *Sardinops sagax*, SEAPODYM, spatial dynamics population, data assimilation, Maximum Likelihood Estimation, adjoint method, upwelling, Humboldt Current system, Peru, Spawning habitat

Resumen

El sistema de la corriente de Humboldt (SCH) frente a la costa peruana es el sistema más productivo del mundo, en términos de biomasa de peces. Los pequeños pelágicos como la anchoveta peruana (*Engraulis ringens*) y la sardina (*Sardinops sagax*) ocupan un posición central en este ecosistema y representan una parte importante de la biomasa total. Históricamente, el sistema ha conocido importantes variaciones de las poblaciones de anchoveta y sardina, una particularidad que continua interrrogando a los oceanógrafos. Hoy en día, la anchoveta es la especie principal, y sostiene la pesquería monoespecífica más importante del mundo, con desembarques anuales que representan alrededor del 10% de la captura mundial de peces. La sardina en cambio está ausente del ecosistema del SCH desde el final de los años noventa. Por ello, es crucial entender cuales son los factores que controlan los stocks de pequeños pelágicos, para poder gestionar mejor este importante recurso.

El objetivo de este trabajo de tesis es de implementar por primera ver un enfoque euleriano con el fin de estudiar los primeros estadios iniciales (huevos y larvas) de vida de la anchoveta y la sardina en el SCH, así como para entender la influencia de los parámetros medio-ambientales y la variabilidad climática en la dinámica de estas poblaciones. El modelo ha sido adaptado de un modelo euleriano de dinámica espacial de los ecosistemas (SEAPODYM), inicialmente desarrollado par los grandes peces pelágicos (como el atún) a escala de la cuenca oceánica. El modelo utiliza un modelo acoplado físico-bioquímico, ROMS-PISCES, en entrada.

Con el fin de definir lo mejor posible los parámetros del modelo, hemos desarrollado un método con asimilación de datos basados en una base única de datos de huevos, larvas y adultos, adquiridos por el Instituto del Mar del Perú (IMARPE) en las aguas del SCH a lo largo de estos 40 últimos años. Estos datos son objeto de un detallado análisis.

El modelo permite ensayar diferentes mecanismos susceptibles de controlar el hábitat de la puesta y el reclutamiento larvario: temperatura óptima, abundancia de presas, presencia/ausencia de predadores e influencia de las corrientes en la retención y la dispersión. Mostramos que la combinación de algunos conceptos simples permite explicar razonablemente la repartición espacial de los huevos y las larvas, así como la variabilidad interanual; no obstante, la descripción actual no permite explicar la variabilidad estacional observada en las areas costeras.

A escala estacional, los resultados sugieren que la abundancia espacio-temporal de las presas, así como la presencia o ausencia de predadores controlan mayoritariamente el hábitat de puesta y la dinámica larvaria, mientras que la temperatura, factor a menudo

puesto en evidencia en el estudio de los primeros estadios de vida, parece ser de menor importancia. A escala interanual, obtenemos una parametrización diferente en la cual la temperatura aparece jugar un papel más significativo, mientras que el impacto de los predadores se vuelve nulo para la anchoveta. La solución del modelo predice una disminución de la abundancia de huevos y larvas durante el fenómeno El Niño 1997-1998, de acuerdo con las observaciones.

Un estudio de sensibilidad de las principales fuentes de variabilidad e incertidumbres del modelo sugieren algunas pistas de mejora de los resultados.

Finalemente, algunos mecanismos ausentes del modelo a este nivel de desarollo, como lo son la influencia del oxígeno, la dinámica de los adultos y la mortalidad por pesca, son susceptibles de mejorar los resultados del modelo y deberían de tomarse en cuenta en el futuro.

Palabras claves: Anchoveta, Sardina, *Engraulis ringens*, *Sardinops sagax*, SEAPODYM, modelo de dinámica espacial de poblaciones, asimilación de datos, Estimación de Máxima Verosimilitud, modelo adjunto, upwelling, Humboldt Current system, Perú, hábitat de puesta

Remerciements

Trois ans et demi : une longue étape s'achève, et ma thèse n'aurait sans doute pas vu le jour sans l'aide de nombreuses personnes.

Je souhaite tout d'abord remercier mes directeurs de thèse, Philippe, Patrick et Arnaud pour leur aide au cours de ma thèse. Philippe, merci d'avoir accepté d'être mon directeur de thèse et de m'avoir permis de revenir au monde de l'océanographie. Patrick, merci de m'avoir initié aux joies (et frustations!) de la modélisation. Merci pour ton aide au cours de ces trois années, et notamment pendant la rédaction du manuscrit, malgré ton planning toujours bien chargé! Arnaud, merci pour ton aide et ton soutien permanent malgré la distance, et merci beaucoup pour ton enthousiasme et ton accueil toujours chaleureux à Sète et à Lima!

Merci beaucoup à Pierre Brasseur et Christian Mullon pour avoir accepté d'être les rapporteurs de ce travail de thèse, et à Isabelle Dadou pour avoir accepté de présider le jury. Merci pour vos commentaires constructifs qui m'ont permis d'améliorer ce manuscrit.

Merci à Vincent Echevin, qui a coordonné le projet PEPS. Merci pour ce beau projet, pour toutes ces discussions scientifiques à Lima et à Paris, et pour ton soutien jusqu'au bout! J'espère que vous continuerez longtemps à travailler sur cet ecosystème encore mal compris. Ce projet a été l'occasion de rencontrer de nombreuses personnes au LOCEAN à Paris, au LEGOS à Toulouse, à l'IRD à Sète ou à l'IMARPE au Pérou. Merci donc à Ainhoa, Thimothée Brochier, Aurélie Albert, Patricia Ayon, Alexis Chaigneau, Boris Dewitte, Sophie Bertrand, Betsy Buitron, Ramiro Castillo, Jorge Tam, ... et à ceux que j'oublie. Gracias!

Merci à CLS et en particulier à l'équipe MEMMS. Merci à Inna pour son aide sur le code de SEAPODYM et l'optimisation, merci à Beatriz pour son écoute. Merci à Eric Greneir de Mercator pour nos échanges sur l'optimisation, à Marie-Helène Rio et Sandrine pour les explications de Surcouf! Aux informaticiens de CLS qui m'ont aidée pour mes nombreuses galères d'ordinateurs et à Glawdys et Nelly pour leur disponibilité et leur aide pour tous les aspects administratifs. A mes ex-co bureaux, Julien et Anna pour les bons moments passés à CLS mais surtout pour leur amitié! Aux cls-siens pour tous les bons moments passés ensemble, en particulier à Marine, Sandrine, Seb, Olivier, Jeff, Guilllaume, Marion, François, Bruno, Laurent, Anabelle, Rémi, etc.

Merci à Martin Huret, pour avoir accepté de faire part de mon jury, mais également pour m'avoir donné la chance de participer à la campagne PELGAS 2011 et d'en apprendre un peu plus sur l'anchois et la sardine du golfe de Gascogne. Et surtout merci pour tous ces bons moments passés à bord de la Thalassa! Je n'oublierai pas! J'en profite pour remercier aussi Momo, Corinne, Pieter et tous les gens de la Thalassa!

Thanks to Rob Dunn for this first internship in Hawaii and this beautiful subject of understanding the vocalization of blue whales. Thanks for proposing me to participate to the Fiji's cruise at the begginning of my PhD.

Merci aux amis de Toulouse, Paris, Barcelone et Hawaii! A Balma et Seb pour ces week-end de rando et tous les autres bons moments passés ensemble! Aux apnéistes toulousains, à ces jeudi sous l'eau qui vont me manquer! A Olivier, pour son amitié en plus de son écoute toujours présente à CLS! A Mélanie, avec qui l'aventure commença il y a maintenant 6 ans dans une rue à Hawaï. Merci pour tous ces moments de partages et de voyages entre Toulouse, Paris, Hawaii et le Japon. On n'en aura peut être plus dans le travail, mais j'espère qu'ils seront nombreux à côté!

Merci aux amis de l'ENS et particulièrement à Philou pour avoir toujours été présent malgré la distance! Merci à Eric le ricain, à Cécile la désormais Madrilène pour tes encouragements constants! Aux amis de Barcelone: Marta y Marina, ¡espero que nos veamos ahora un poco más! A Lourdes y Jose por iniciarme desde pequeña al buceo, ¡por su amistad y amor al mar! Un grand merci à Marine et Sandrine pour leur amitié, leur bonne humeur et tous ces bons moments passés ensemble à CLS mais surtout en dehors du boulot! Merci d'avoir été toujours là pour moi ces deux dernières années à Toulouse!

Muchas gracias a toda mi familia, por su amor y cariño. A mis padres, por apoyarme en todo momento; por haberme dejado soñar con el mar desde pequeña, ¡aunque a veces no les gustase que pasara tanto tiempo bajo el agua buceando! ¡Gracias por haberme dado esa curiosidad y amor por la naturaleza! A mis hermanas por haber estado siempre a mi lado: Clara, ¡espero que ahora podamos vernos un poco más! y a Elena, muchas gracias por estos tres años en Toulouse, ¡por todos esos detalles que siempre has tenido conmigo en los momentos más difficiles! Gracias a Gregoire, a Abel y gracias también a los tres peques: Laura, Emma y Raphael,¡¡gracias por vuestra alegría!! A Papi, qui est parti avant que je soutiennes cette thèse... et qui aurait voulu être là. Merci pour ta philosophie de la vie, merci pour ces conversations philosophiques et ces bons moments passés lors des petites escapades aux Sables entre conférences et vacances. J'aurais aimé quelles soient plus nombreuses! A mes oncles et tantes et en particulier à Odile qui a suivi cette thèse de près!, à mes cousins, à mes grands-parents, à Claire et Patrick, a mis primos, a mis tíos y tías (Montse; ¡gracias por las tortillas de patata que le gustaron mucho al jurado!), a mis abuelos... gracias pour vuestro apoyo, merci de vos encouragements!

Y por último muchas gracias a tí, Quentin, no solo por todas las conversaciones científicas y tu interés constante por esta tesis, sino por tu amor, tu infatigable apoyo, tu sensibilidad, tu paciencia, por creer en mí, por hacerme soñar. ¡Esta tesis tambien es tuya!

No hay palabras para daros las gracias, así que esta tesis es para vosotros: para mi familia, para Quentin. Os la dedico.

Bancs de sardines. ©: Henry Jager

" Little Fish : Big Impact ! " *

Arrivée des anchois à l'usine de farine de poisson. Pérou. ©Lenfest Ocean Program

Flottille de pêche et usines de farine de poisson sur le littoral péruvien. ©IRD - Film: "L'anchois du Pérou, un El Dorado ?"

*Rapport du Lenfest Ocean Program (Pikitch et al., 2012).

Table des Matières

1 Introduction **1**

1.1 Introduction (French Version) . . . 2
1.1.1 Motivations . . . 2
1.1.2 Le système d'upwelling de Humboldt: Océanographie . . . 7
1.1.3 Populations d'anchois et sardines . . . 11
1.1.4 Approche par modélisation numérique . . . 14
1.1.5 Objectifs de la thèse . . . 17
1.1.6 Structure de la thèse . . . 19
1.2 Introduction (English version) . . . 22
1.2.1 Motivations . . . 22
1.2.2 The Humboldt Current system: Oceanography . . . 26
1.2.3 Anchovy and Sardines populations . . . 31
1.2.4 Approach by numerical modelling . . . 33
1.2.5 Thesis aims and objectives . . . 36
1.2.6 Structure of the thesis . . . 38
Bibliography . . . 40

2 Fundamentals of the small pelagic fish SEAPODYM model version **45**

2.1 Introduction . . . 46
2.2 Main processes controlling spawning habitat and eggs and larvae dynamics 46
2.3 Definition of spawning habitat and larvae recruitement in the SEAPODYM standard version . . . 47
2.4 Revision from SEAPODYM standard version for small pelagics . . . 47
2.4.1 Spawning habitat definition . . . 48
2.4.1.1 Impact of temperature . . . 49
2.4.1.2 Effect of a relationship to prey abundance . . . 49
2.4.1.3 Effect of a relationship to predator abundance . . . 50
2.4.1.4 Accessibility of adults to spawning areas . . . 50
2.4.1.5 Illustration of spawning habitat definition . . . 50
2.4.2 Predicted eggs density . . . 51
2.4.3 Predicted larvae density . . . 52

2.5 Conclusion . . . 53
Bibliography . . . 54

3 Observed variability on anchovy and sardine dynamics **57**
3.1 Introduction . . . 58
3.2 Available Data . . . 58
3.2.1 Eggs, larvae data . . . 58
3.2.1.1 Spatial and temporal heterogeneity of datasets . . . 58
3.2.1.2 Statistical characteristics . . . 61
3.2.2 Adult data . . . 67
3.2.2.1 Spatial and temporal heterogeneity of datasets . . . 68
3.2.2.2 Statistical characteristics . . . 70
3.2.3 Gonadosomatic Index . . . 71
3.3 Spatial Distribution . . . 72
3.3.1 All datasets . . . 72
3.3.2 Climatology maps . . . 76
3.4 Seasonal variability . . . 77
3.4.1 Anchovy . . . 77
3.4.2 Sardine . . . 80
3.5 Interannual Variability: El Niño events . . . 82
3.6 Decadal variability . . . 85
3.7 Physical observations for eggs, larvae and adults . . . 89
3.7.1 Effect of temperature . . . 89
3.7.2 Vertical distribution of eggs and larvae . . . 92
3.7.3 Spatial co-occurrence of eggs and adults . . . 94
3.8 Predators of anchovy and sardines . . . 96
3.9 Conclusion . . . 98
Bibliography . . . 100

4 Environmental forcing of Seapodym Model **103**
4.1 Introduction . . . 104
4.2 Satellite and *in situ* observations . . . 105
4.2.1 Sea Surface Temperature (SST) . . . 105
4.2.2 Chlorophyll and Primary Productivity . . . 105
4.2.3 Oxygen . . . 107
4.2.4 Currents . . . 109
4.3 ROMS-PISCES model outputs . . . 111
4.3.1 The model . . . 111
4.3.2 Vertical habitat definition . . . 112
4.3.3 Climatological Run . . . 113
4.3.3.1 Configuration . . . 113

4.3.3.2 Euphotic depth and mixed layer depth . . . 113
4.3.3.3 Oxygen . . . 114
4.3.3.4 Currents . . . 117
4.3.3.5 Temperature, Primary Productivity . . . 119
4.3.4 Interannual Run 1992-2000 . . . 123
4.3.5 Interannual Run 2000-2006 . . . 126
4.4 Eggs and larvae predators: the micronekton model . . . 129
4.5 Conclusion . . . 133
Bibliography . . . 134

5 Modeling early life history **137**
5.1 Introduction . . . 139
5.2 Materials and Methods . . . 141
5.2.1 Anchovy data . . . 141
5.2.2 Bio-physical environment . . . 142
5.2.3 SEAPODYM spawning habitat and larval recruitment . . . 143
5.2.4 Simulations . . . 145
5.3 Results . . . 146
5.3.1 Eggs and larvae distributions . . . 146
5.3.2 Seasonal variability of predicted environment . . . 148
5.3.3 Observed seasonal variability in anchovy reproduction . . . 150
5.3.4 Simulated spatial distributions . . . 153
5.3.5 Simulated seasonality . . . 154
5.4 Discussion . . . 158
A Appendices . . . 161
A.1 Continuous model of larvae spatial dynamics. . . . 161
A.2 Spawning . . . 161
A.3 Mortality . . . 162
Bibliography . . . 163

6 Parameter estimation in data assimilation framework **169**
6.1 Introduction to data assimilation concepts . . . 170
6.2 Bayesian Approach . . . 171
6.3 Global optimization approach: Grid search method . . . 174
6.4 Local optimization approach: SEAPODYM framework . . . 175
6.4.1 Adjoint methodology . . . 177
6.4.2 Parameter scaling and bounds . . . 179
6.4.3 Analysis of the quality of the estimate: Hessian Information . . . 181
6.5 Implementation of a new likelihood function on SEAPODYM . . . 182
6.5.1 Observations and predictions . . . 182
6.5.2 Cost function . . . 183

6.5.3 Initial conditions . . . 184
6.5.4 Sensitivity analysis . . . 185
6.5.4.1 Sensitivity of model to parameters . . . 185
6.5.4.2 Sensitivity of objective function to parameters . . . 185
6.6 Selecting the distribution of the likelihood function . . . 186
Bibliography . . . 192

7 Optimization of early life history parameters 195
7.1 Introduction . . . 196
7.2 Optimization with climatology time series . . . 197
7.2.1 Sensitivity Analysis . . . 197
7.2.2 Correlated parameters . . . 198
7.2.3 Parameters estimation . . . 199
7.2.4 Exploring the effect of oxygen concentration . . . 210
7.2.5 Exploring the impact of larvae retention for older larvae cohorts . 210
7.3 Analysis of the sources of uncertainties . . . 212
7.3.1 Optimization at coarser resolution . . . 212
7.3.2 Sensitivity to data uncertainties . . . 214
7.3.2.1 Impact of extreme high data values . . . 214
7.3.2.2 Impact of the cost function . . . 215
7.3.3 Impact of the vertical layer definition . . . 216
7.3.4 Impact of mesoscale variability . . . 218
7.3.5 Impact of physical inputs . . . 219
7.3.6 Sensitivity to simulated predator inputs. . . . 221
7.3.7 Conclusion . . . 222
7.4 First optimization experiments at interannual scale . . . 223
7.4.1 Optimization with the 1992-2000 run at 1/6° of resolution . . . 223
7.4.2 Optimization with the run 2000-2006 at 1/9° of resolution . . . 234
Bibliography . . . 236

8 Conclusion 237
8.1 Conclusion (English version) . . . 238
8.1.1 Synthesis of the study . . . 238
8.1.2 Main results . . . 239
8.1.3 Limitations of the study . . . 240
8.1.4 Perspectives of the study . . . 241
8.1.4.1 Improvements of the optimization method . . . 242
8.1.4.2 Taking into account adults and catch . . . 242
8.1.4.3 Long term perspectives . . . 243
8.2 Conclusion Générale (French version) . . . 245
8.2.1 Synthèse de l'étude . . . 245

8.2.2 Résultats majeurs 246
8.2.3 Limitations de l'étude 247
8.2.4 Perspectives de l'étude 249
8.2.4.1 Perspectives d'amélioration de la méthode d'optimisation 249
8.2.4.2 Prise en compte des adultes et de la pêche 250
8.2.4.3 Perspectives à long terme 251
Bibliography 252

A Spatial Map for Anchovy and Sardines **253**

B Practical coding of the adjoint **267**
Bibliography 268

C Tracking blue whales in the eastern tropical Pacific with an ocean-bottom seismometer and hydrophone array **269**

D Correlation coefficient and standard deviation uncertainties **281**

Chapter 1

Introduction

Sommaire

1.1 Introduction (French Version) 2

1.1.1 Motivations 2

1.1.2 Le système d'upwelling de Humboldt: Océanographie 7

1.1.3 Populations d'anchois et sardines 11

1.1.4 Approche par modélisation numérique 14

1.1.5 Objectifs de la thèse 17

1.1.6 Structure de la thèse 19

1.2 Introduction (English version) 22

1.2.1 Motivations 22

1.2.2 The Humboldt Current system: Oceanography 26

1.2.3 Anchovy and Sardines populations 31

1.2.4 Approach by numerical modelling 33

1.2.5 Thesis aims and objectives 36

1.2.6 Structure of the thesis 38

Bibliography 40

1.1 Introduction (French Version)

1.1.1 Motivations

Au cours des 60 dernières années, la population mondiale s'est accrue de 4,25 milliards d'habitants pour atteindre aujourd'hui 7 milliards de personnes. Dans plusieurs parties du monde, l'évolution démographique s'est accompagnée d'un développement économique associé à une diversification des sources protéiques dans l'alimentation, la consommation de produits d'origines animale (viande et poisson) augmentant. En particulier, la consomation moyenne annuelle de consommation de poisson est passée de 9,9 kg par habitant dans les années 60 à 17 kg en 2007 (FAO, 2010). Cette demande accrue a provoqué une croissance explosive de la production mondiale de poisson, qui est passée de 19 millions de tonnes en 1950, à 90 millions de tonnes en 1990 puis 154,1 millions de tonnes en 2008 (FAO, 2010). A la fin des années 80, les captures marines totales atteignirent 80 à 85 millions de tonnes (Pauly et al., 2003) puis déclinèrent jusqu'à 79.5 millions de tonnes en 2008. Les pêcheries ont probablement atteint leur Rendement Maximal Soutenable (RMS), après avoir atteint leur maximum (Pauly et al., 2002). La croissance récente de la production de poisson est liée à la croissance de la production d'aquaculture marine et lacustre, qui assurait en 2008 46% de la production mondiale totale (FAO, 2010).

Le déclin des captures mondiales, corrélé à l'augmentation de la production aquacole, est lié à la surexploitation de la plupart des stocks de poisson. Selon la FAO, la proportion des stocks mondiaux de poissons sous-exploités ou modérément exploités a chuté de 40% à 15 % entre le milieu des années 70 et 2008, tandis que la proportion des stocks de poissons surexploités, épuisés ou en phase de reconstitution passait de 10% à 32% pour les 600 stocks de poissons, suivis par la FAO (FAO, 2010).

En 2008, 3% seulement des stocks étaient estimés sous-exploités, et 1% et 12% exploités modérément. 52% des stocks sont surexploités et par conséquent on atteint leur rendement maximum soutenable.

L'analyse des données mondiales de pêche montre une diminution de la taille moyenne des individus capturés (FAO, 2010), ce qui résulte de la pression importante exercée par la pêche sur les stocks mondiaux de poisson. L'effondrement du stock de morue dans l'atlantique Nord en 1992 (Myers et al., 1997) ou le fort déclin des populations de thon rouge en Méditerranée (MacKenzie et al., 2009) sont des exemples bien connus de conséquences de la surpêche.

Il est devenu nécessaire de mieux gérer les stocks afin de restaurer les populations surexploitées à des niveaux soutenables à long terme. Si 90 % des captures mondiales ont lieu dans les zones économiques exclusives (ZEE) (Rothschild, 1996), dans une région de

200 miles au larges des côtes, toutes les zones côtières ne sont pas équivalentes en terme de production (Fig. 1.1).

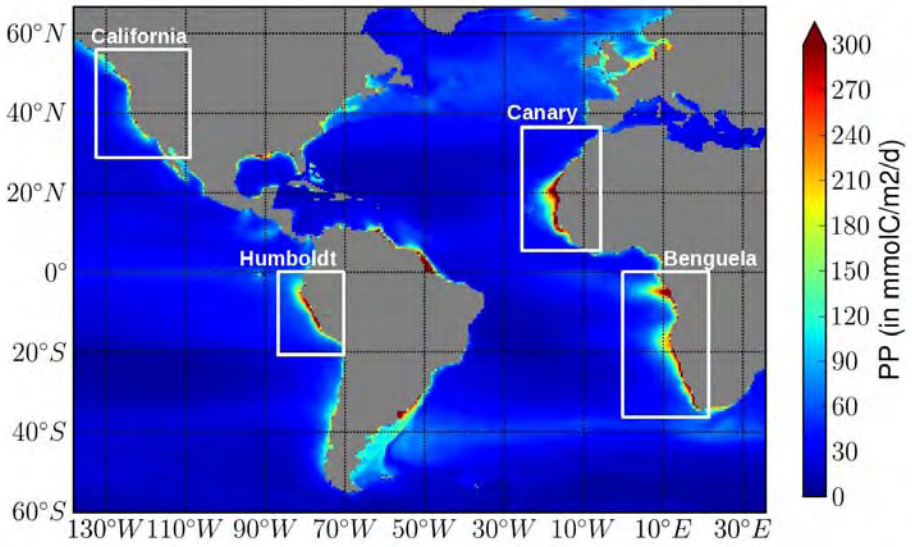

Figure 1.1: Moyenne annuelle de la production primaire tirée de données satellitaires (basée sur l'algorithme VGPM de Behrenfeld and Falkowski (1997). Moyenne calculée sur la période allant d'Octobre 1997 à Décembre 2008. Les 4 principaux Système d'Upwelling de Bord Est (EBUS) sont caractérisés par des vents alizés persistants entraînant la formation d'un upwelling côtier, d'une forte advection le long de la côte vers l'équateur, d'un contre-courant en profondeur dirigé vers les pôles, de précipitations faibles à modérées ainsi que d'une très forte production de plancton et poissons, plus particulièrement de petits pélagiques (Fréon et al., 2009).

Il y a 4 principaux systèmes d'upwelling de bord est (EBUS): le système du courant de Humboldt (HCS) dans le Pacifique Sud, le système du Courant des Canaries (CanCS) dans l'Atlantique Nord, le système du courant du Benguela (BCS) dans l'atlantique Sud, et le système du Courant de Californie (CalCS) dans le Pacifique Nord. Ils représentent moins de 0,1% de la surface des océans mais concentrent 30% des captures mondiales de pêche (Durand et al., 1998). La productivité de ces régions est dûe à la remontée dans la couche de mélange d'eaux riches en nutriments, un phénomène généré par les alizés qui soufflent vers l'équateur le long de la côte. Les écosystèmes de ces systèmes d'upwelling sont dominés par les petits poissons pélagiques, majoritairement les anchois et les sardines (Bakun, 1996).

Parmis ces systèmes d'upwelling de bord Est, **le système d'upwelling du courant de Humboldt est le troisième système le plus productif en terme de production primaire, et le plus productif en terme de biomasse de poisson** (Bakun and Broad, 2003; Chavez et al., 2008, Fig. 1.2). Le système est dominé par l'anchois du Pérou, connu selon le nom d'anchoveta, actuellement exploité par la plus grande flottille commerciale au monde ciblant une seule espèce. Les captures annuelles atteignent 12 millions de tonnes

les meilleures années. En 2008, les captures totales d'anchois représentèrent 7,4% des captures mondiales de poisson, constituant ainsi la plus grande pêcherie au monde exploitant une seule espèce (FAO, 2010). 92 % des débarquements d'anchois sont utilisés pour la production de farines et huiles de poissons, majoritairement à destination de l'aquaculture. 30 à 35 % de la production de farines et huiles provient de la pêche à l'anchois au Pérou (International Fishmeal and Fish Oil Organisation, IFFO, 2009). Cette industrie contribue à 1.2% du PIB du Pérou (en 2007), et emploie directement et indirectement plus de 125000 personnes (Bertrand et al., 2010). Il est ainsi primordial pour les Péruviens de maintenir et de bien gérer cette précieuse ressource.

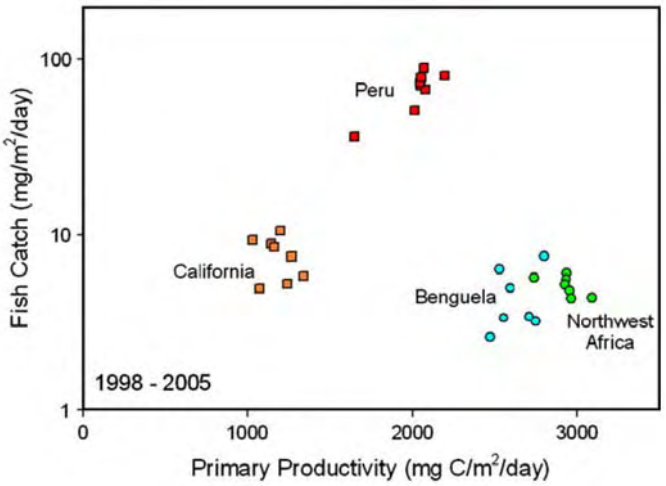

Figure 1.2: Captures de poisson en fonction de la production primaire dérivée de données satellitaires (Behrenfeld and Falkowski, 1997) pour les 4 principaux systèmes d'upwelling de bord Est entre 1998 et 2005. Source: Chavez et al., 2008.

Ceci illustre l'importance de l'anchois en tant qu'espèce clé du système d'upwelling de Humboldt, ainsi que son rôle majeur dans le monde de la pêche et de l'industrie aquacole. Avec une demande croissante de petits poissons pélagiques pour nourrir en partie les poissons d'élevages (Checkley Jr et al., 2009), le désir d'exploiter cette ressource est élevé. Cependant, selon la FAO, **le stock d'anchois est d'ores et déjà surexploité.** Avec des armements modernes et suréquipés, dont environ 1200 senneurs (Fréon et al., 2008), les pêcheurs péruviens peuvent pêcher jusqu'à 170 000 tonnes d'anchois en un seul jour. Sans le contrôle du gouvernement et des scientifiques, la pêcherie de l'anchois du Pérou ne pourrait être durable. Après l'effondrement des stocks d'anchois dans les années 1970, un quota annuel global fut introduit et les décisions concernant la pêcherie sont depuis prises quasiment en temps réel en utilisant les observations des scientifiques les plus récentes. L'ouverture de la saison de la pêche varie d'année en année selon les recommandations des scientifiques. Généralement, la pêcherie ouvre deux fois par an, d'avril à juillet et d'octobre à janvier hors des principales saisons de ponte. La pêcherie est aussi sujète à différentes restrictions, et est arrêtée aussi tôt que les juvéniles atteignent plus de 10% des débarquements ou si la taille de l'anchois est inférieure à 12. La gestion de cette pêcherie dans le nord du HCS est ainsi reconnue comme la plus

rapide et adaptative au monde (Chavez et al., 2008).

Malgré cela, la surcapacité de la flottille de pêche péruvienne est estimé à 300%, en raison de la course à la pêche des quotas annuels, et de l'incapacité des gestionnaires à contrôler la croissance de la flottille et des industries transformatrices (Fréon et al., 2008). Avec ce système, les quotas annuels furent atteint en 50 jours en 2007 (Fréon et al., 2008), contre 270 jours en 1986. En 2009 fut introduit un système de gestion individuelle permettant la réduction du nombre de bateaux participant à la pêcherie et permettant de mieux contrôler la course à la pêche tout en prolongeant la saison de pêche (Tveteras et al., 2011).

Dans les EBUS, les petits poissons occupent une position clé dans l'écosystème marin puisqu'ils dominent le niveau trophique moyen. Bien que composées de relativement peu d'espèces, les populations atteignent de grandes abondances qui peuvent varier drastiquement en taille (Checkley Jr et al., 2009).

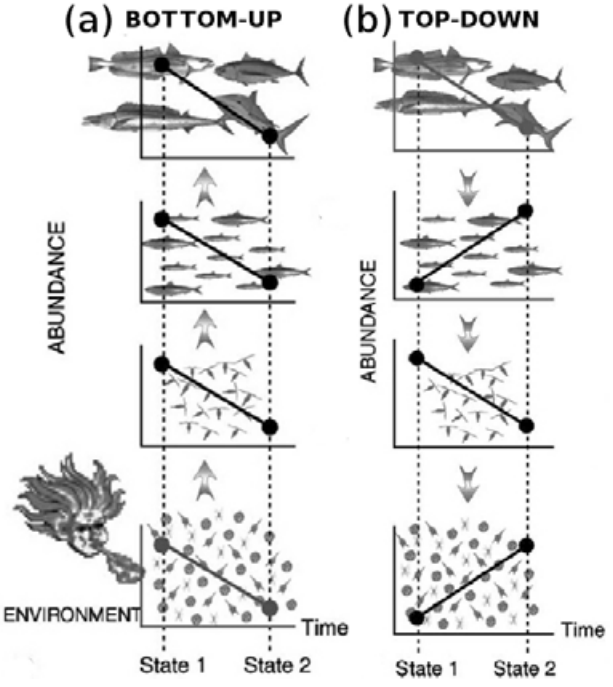

Figure 1.3: (a) Contrôle "Bottom-up" et (b) "Top Down", dans un système trophique simplifié à 4 niveaux. Dans le contrôle "Bottom-up", l'environnement abiotique perturbe le phytoplancton et cette perturbation se propage vers le haut menant a des évolutions d'abondances corrélées positivement à tous les niveaux trophiques. Dans le contrôle "Top-Down", la pêche perturbe les top-prédateurs, effet qui se propage vers le bas, menant à des évolutions d'abondances corrélées négativement. Adapté de Cury and Shannon (2004).

Leurs stocks peuvent répondre à un changement par le bas, comme un forçage environnemental (contrôle "bottom-up"), ou par le haut, comme le forçage de la pêche (contrôle "top-down") (Cury et al., 2000, Fig. 1.3).

Cependant, le forçage "bottom-up" apparait comme le forçage principal dans le HCS aux échelles de temps interannuelles, multidécénnales et centennales (Ayón et al., 2008; Bertrand et al., 2008; Chavez et al., 2008). En effet, les stocks de petits pélagiques dans le HCS varient de manière importante à différentes échelles de temps, en raison de contraintes climatiques à différentes échelles de temps: les variations océanographiques à l'échelle séculaire, l'oscillation décennale du Pacifique (PDO) à l'échelle multidécennale,

l'ENSO (Oscillation Nino) à l'échelle interannuelle, et la variabilité saisonnière à l'échelle intra-annuelle (Chavez et al., 2008).

Cette variabilité climatique affecte directement les stocks d'anchois et sardines, qui fluctuent au cours du temps. Ces fluctuations, observées dans d'autres systèmes EBUS (Lluch-belda et al., 1992; Alheit and Niquen, 2004), ont posé de nombreuses questions aux océanographes et scientifiques des pêches au cours des dernières décennies, et leur origine est encore débattue (Chavez et al., 2008). De plus, cette forte variabilité rend le management de ces pêcheries difficiles. Les fortes variations des captures de pêche ont des conséquences socio-économiques très importantes (voir section 1.1.3). Dans ce contexte, le développement d'un modèle qui puisse aider à mieux comprendre l'impact de la variabilité environnementale sur les populations de poissons exploités représenterait un pas en avant très significatif.

De plus, dans le contexte du changement climatique en cours à l'échelle globale et régionale, il devient crucial de développer et intégrer des outils scientifiques afin d'étudier l'impact du changement climatique sur les ressources marines. C'est l'objectif du projet Peru Ecosystem Projection Scenarios (PEPS), financé par l'Agence Nationale de la Recherche (ANR). Ce projet est réalisé en collaboration avec l'Institut de la Mer du Pérou (IMARPE), l'Institut de Recherche pour le Développement (IRD), le laboratoire Mixte International "Dynamiques du système du courant de Humboldt" (LMI DISCOH) et l'entreprise CLS (Collecte Localisation Satellites). Ce projet a pour objectif d'évaluer l'impact du changement climatique sur les deux plus grandes pêcheries de la partie Nord du HCS: l'anchois et la sardine. Cependant, avant d'effectuer des prédictions, nous devons construire un modèle capable de reproduire la variabilité des stocks aux cours de ces dernières 50 années. Ce sera notre premier et principal objectif.

Des modèles incluant une représentation de l'écosystème dans son ensemble avec ses forçages physiques et biogéochimiques sont nécessaires afin de tester en même temps les scénarios de changements climatiques et la pression de pêche. Le développement d'un modèle qui puisse expliquer la variabilité environnementale des stocks, en prenant en compte rigoureusement l'impact de la pêche et les paramètres décrivant la dynamique des stocks, constituerait une avancée vers la mise en place d'une approche écosystémique pour le management des pêches.

Une approche de modélisation avec des objectifs similaires a été développée récemment pour de grands poissons pélagiques (comme le thon ou les espèces similaires au thon) à l'échelle du bassin (Lehodey et al., 2008, 2010) avec un modèle de de dynamique spatiale des populations et des écosystèmes (SEAPODYM). L'idée initiale de ce projet est

d'adapter le modèle SEAPODYM à l'anchois et à la sardine dans le HCS, des oeufs aux adultes, en prenant compte l'impact de la pêche.
Cependant, les défis pour adapter SEAPODYM à l'échelle régionale d'une part et à des espèces de petits poissons pélagiques d'autre part, sont multiples. Cette thèse se focalise donc sur les premiers stades de vie des populations d'anchois et sardines, qui contrôlent largement la dynamique de toute la population.
De plus, le second objectif majeur est de développer une méthode d'assimilation pour le modèle d'écosystème dans le but de mieux représenter la réalité des observations.

1.1.2 Le système d'upwelling de Humboldt: Océanographie

Le système du Courant de Humboldt (HCS) du Pérou et Chili est remarquable pour différentes raisons. Suivant la description de Chavez et al. (2008) nous résumons les caractéristiques de ce système en 4 points.

Premièrement, comme mentionné antérieurement, le système est caractérisé par la présence d'eaux froides et riches en nutriments qui circulent à des latitudes où les températures devraient être plus élevées (Chavez et al., 2008). La figure 1.4 présente le schéma de la circulation océanique dans cette zone. L'océanographie générale du HCS est caractérisée par une circulation prédominante vers le nord des eaux de surface d'origine subantarctique et par un fort upwelling des eaux de subsurface froides et riches en nutriments d'origine équatoriale (Thiel et al., 2007).

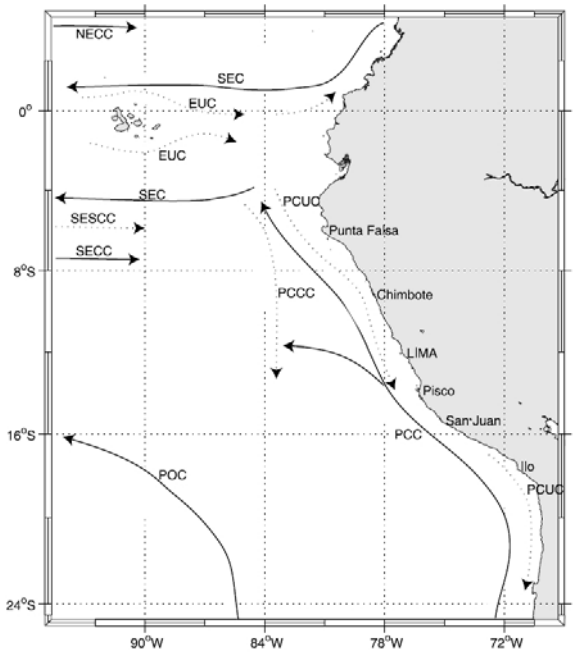

Figure 1.4: Schéma de la circulation océanique dans le Système du Courant du Pérou (Penven et al., 2005.

Le long de la cote Péruvienne et Chilienne, les alizés, vents de sud-est, sont forcés par les montagnes andines de souffler vers le nord (Penven et al., 2005). Ces vents soufflants vers le nord, l'eau superficielle est transportée vers l'ouest par le transport d'Ekman, conséquence de l'effet de Coriolis. Les vents soufflant de manière continue, l'upwelling est présent tout au long de l'année.

La circulation de surface est dominée le long de la côte par le Courant Côtier du Pérou, également appelé Courant de Humboldt, associé aux eaux salées et froides de l'upwelling côtier. Sous cette couche de surface ($\sim$ $20m$) circulant en direction de l'équateur, le sous courant du Pérou-Chili (PCUC) circule en direction du Sud et domine la circulation de sub-surface et du plateau. A 5°S, le PCUC a été observé à des profondeurs de 50 à 100m et s'approfondit en direction du pôle.

Deuxièmement, le HCS du Pérou, comme décrit précédemment, est le système d'upwelling de bord Est le plus productif en terme de biomasse de poisson. Il n'y a cependant pas de relation directe entre la quantité de poisson capturée et la production primaire estimée par satellite (Fig. 1.2), et ce paradoxe qui a a interpellé les scientifiques au cours des dernières décennies (Bakun and Weeks, 2008; Chavez et al., 2008; Ballón et al., 2011; Brochier et al., 2011) reste inexpliqué.

Troisièmement, le HCS contient la zone de minimum d'oxygène (OMZ) la plus intense et la plus superficielle de tous les océans, résultat de l'export vers les profondeurs de la production primaire de surface, ainsi que d'une ventilation faible.

Quatrièmement, étant donné la proximité avec l'Equateur, la dynamique du Nord du HCS est intimement liée à la dynamique du Pacifique équatorial, et plus particulièrement au phénomène d'oscillation australe El Niño (ENSO), le signal climatique le plus important à l'échelle inter-annuelle (McPhaden et al., 2006).

De manière générale, le HCS est soumis à des contraintes climatiques à différentes échelles de temps:

- **Variabilité séculaire**
  Les études des séries temporelles des sédiments marins au large du Pérou ont montré "un changement abrupt du régime biogéochimique à l'échelle centennale, de plus grande magnitude et durée que la variabilité actuelle multi-décénnale" (Gutiérrez et al., 2009). Ce changement à l'échelle centennale qui commença aux alentours de 1820 AD (Fig 1.5), fut probablement dû à un déplacement vers le nord de la Zone de Convergence Inter-tropicale (ITCZ) et de l'Anticyclone Subtropical du Pacifique Sud-Est jusqu'à leur position actuelle, couplé à un accroissement de la circulation

de Walker. Ceci provoqua une intensification de l'upwelling qui augmenta la productivité de la zone et par conséquent la biomasse de petits pélagiques (Fig 1.5).

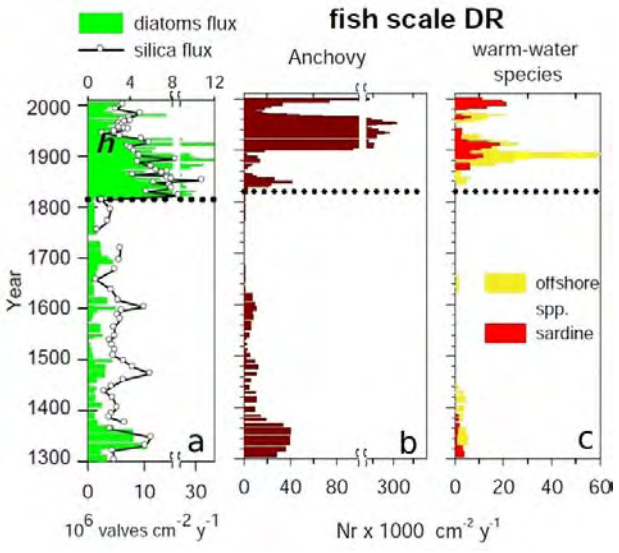

Figure 1.5: Proxy de la production primaire dans l'écosystème (a), biomasse d'anchois (b) et biomasse de sardine (c). (a) barres: taux d'accumulation de diatomées (10^6 valves $cm^{-2}.y^{-1}$), cercles blancs: flux de silice biogénique ($mg.cm^{-2}.y^{-1}$); (b) Moyenne glissante sur 3 ans des taux de déposition des écailles d'anchois ($Nr \times 1000$ $cm^{-2}.y^{-1}$); (c) Moyenne glissante sur 3 ans des taux de déposition des écailles de sardines et autres petits pélagiques (maquereaux et chinchards) ($Nr \times 1000$ $cm^{-2}.y^{-1}$). Données provenant des sédiments de Callao. Source: Gutiérrez et al., 2009

- **Variabilité décennale**

  L'oscillation décennale du Pacifique (ODP, "PDO en anglais") est une fluctuation décennale océan-atmosphère dans l'océan Pacifique. Elle a été décrite comme une analogie au phénomène climatique du "Niño", mais de plus longue durée, entre 20 à 30 ans. Les phases décennales chaudes et froides se différencient les unes des autres par une anomalie de la température de surface de 1°C à 2°C.

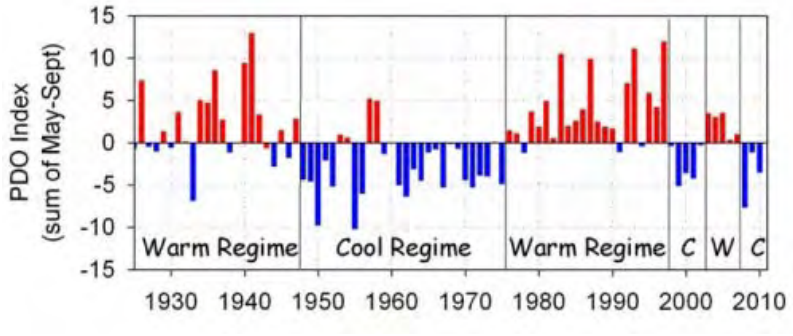

Figure 1.6: Série temporelle des changements de phases de l'oscillation décennale du Pacifique (PDO), de 1925 à 2010. Les barres rouges indiquent les années d'anomalie positive (années chaudes), les barres bleues indiquent les années d'anomalie négative (années froides). Le régime froid eut lieu de 1950 à 1971 et de 1999 jusqu'à aujourd'hui, le régime chaud de 1972 à 1998. Un léger régime chaud est observé de 2003 à 2008. Crédit: NOAA PMEL.

Les causes de la PDO restent méconnues. L'indice de la PDO est défini comme la variabilité de la température de surface dans le nord du Pacifique, et est

représenté dans la figure 1.6. Au cours des périodes "chaudes" (rouge), l'intensité de l'upwelling péruvien ainsi que la productivité sont faibles, et la thermocline est plus profonde. Dans les périodes "froides" (bleu), la situation inverse se produit.

- **Variabilité interannuelle**
  Le phénomène d'Oscillation australe El Niño (ENSO) est dû à une instabilité de la dynamique ocean-atmosphère dans le bassin de l'océan Pacique. Les évènements "El Niño" et "La Niña" représentent les phases opposées du cycle ENSO. L'évènement du Niño représente la phase chaude, avec une température de surface plus chaude dans la zone est-Pacifique qu'en temps normal. L'évenement "Niña" représente la phase froide. La fréquence d'alternance des deux phases est irrégulière et varie entre 2 et 7 ans.

  L'indice d'oscillation australe (IOS, "SOI" en anglais) est une mesure de la fluctuation à grande échelle de la pression atmosphérique entre l'est et l'ouest du Pacifique. Il est défini pour mesurer l'intensité et la durée de l'oscillation. L'indice est calculé à partir de la différence de pression atmosphérique entre Tahiti (Polynésie Française) et Darwin (Australie). Un fort indice négatif (valeurs rouges) indique un évenement "El Niño" et un indice positif (valeurs bleues) indique un évènement "La Niña" (Fig. 1.7).

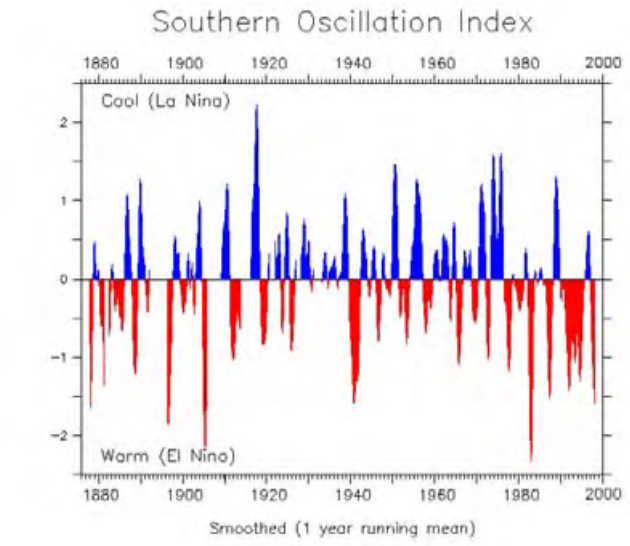

Figure 1.7: Indice d'Oscillation Australe (IOS). Les plus forts évènements "El Niño" ont eu lieu dans les années 1982-83 et 1997-98. Source: NOAA

- **Variabilité Saisonnière**

  A l'échelle intra-annuelle, la variabilité est déterminée par la variabilité saisonnière de l'intensité de l'upwelling, de luminosité et de température, ainsi que par la variabilité des structures de mésoéchelles (tourbillons, panaches). Le système d'upwelling est présent tout au long de l'année mais varie en intensité avec

les saisons. Paradoxalement, intensité du transport et production primaire sont en opposition de phase (Chavez and Messié, 2009). En effet, l'intensité maximale de l'upwelling de l'HCS est atteinte au cours de l'hiver austral et le minimum d'intensité a lieu pendant l'été austral. La concentration en surface en chlorophylle atteint en revanche son minimum lors de l'hiver austral et son maximum lors de l'été austral (Fig. 1.8). Echevin et al. (2008) suggèrent que le cycle saisonnier pourrait être expliqué par la variabilité saisonnière de la couche de mélange: en hiver, quand l'intensité de l'upwelling est maximale, la couche de mélange est plus profonde, la dilution des nutriments est plus forte et la luminosité plus faible, ce qui résulte en un minimum de production primaire, et donc de chlorophylle. En été, la couche de mélange est plus superficielle, la luminosité est plus forte et la biomasse de phytoplancton est concentrée dans le niveau de surface, ce qui résulte en un maximum de chlorophylle.

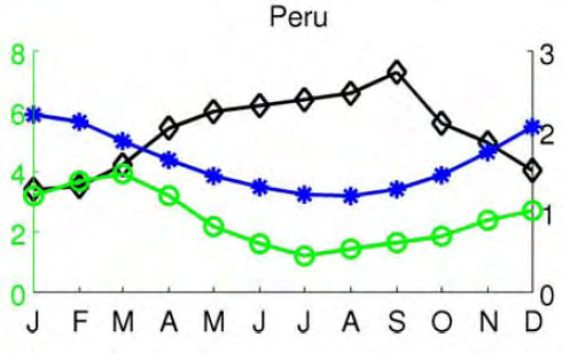

Figure 1.8: Cycle saisonnier de la concentration en chlorophylle (gris, SeaWiFS 1997-2007), température de surface (bleu, Reynolds 1981- 2007) et transport vertical totale (noir, transport d'Ekman plus pompage d'Ekman, estimé à partir des vents QuikSCAT 1999-2008) moyenné entre 6°S et 16°S dans une bande de 150km offshore. Source: Chavez and Messié (2009)

1.1.3 Populations d'anchois et sardines

Nous nous focalisons sur les deux principales populations de petits pélagiques du Pérou et du Chili: l'anchois (*Engraulis ringens*) et la sardine (*Sardinops sagax*). Ces deux espèces appartiennent à l'ordre des Clupéiformes, l'anchois appartenant à la famille Engraulidae et la sardine à la famille Clupeidae.
Ces deux espèces ont une espérance de vie relativement courte (environ 4 ans et 8 ans, respectivement) avec une taille maximale de 20 cm et 40 cm respectivement, une croissance rapide et atteignent tôt leur maturité (un an et deux ans respectivement). L'anchois et la sardine se nourrissent principalement de zooplancton, mais l'anchois se nourrit directement de larges copépodes et euphausiacés, tandis que la sardine semble se nourrir principalement par filtration d'éléments plus petits du zooplancton même si elle est capable de se nourrir de macrozooplancton (Espinoza and Bertrand, 2008; Espinoza

et al., 2009; Ayón et al., 2011, Fig. 1.10).

Trés fécondes et capables de pondre toute l'année, ces espèces sont très bien adaptées à la variabilité du HCS (Bertrand et al., 2004; Gutiérrez et al., 2007; Espinoza and Bertrand, 2008; Swartzman et al., 2008), mais restent cependant très sensibles aux forçages environnementaux. Leurs stocks répondent ainsi rapidement et parfois de manière dramatique aux changements climatiques de l'océan.

Au cours des dernières décennies, différentes périodes de fortes abondances d'anchois suivies par des périodes de fortes abondances de sardines (et vice-versa) ont été observées (Fig 1.9) sur l'écosystème du HCS.

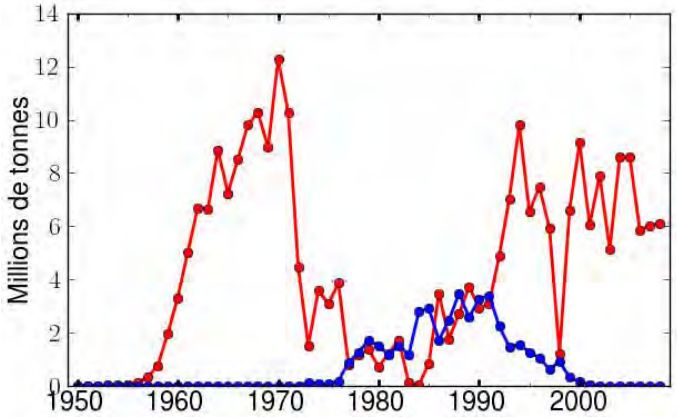

Figure 1.9: Débarquements d'anchois (*Engraulis ringens*) et sardines (*Sardinops sagax*) dans le HCS entre 1950-2008 (Données de l'IMARPE). Les captures totales dans l'année sont utilisées ici comme indice d'abondance. Cependant, cette hypothèse est vraie uniquement si l'effort de pêche et la capturabilité demeurent constantes pendant toute la période d'étude. Dans le cas de la pêcherie de l'anchois, la capture est estimée être un bon indice de biomasse depuis les années 1965, année des débuts de la pêche industrielle à l'anchois, et même si l'effort de pêche a augmenté de façon continue depuis lors en raison de l'amélioration de capacité de la flottille (Fréon et al., 2008).

Les débarquements d'anchois peuvent varier de 1 à 12 millions de tonnes par an. Dans les années 1970, la pêcherie de l'anchois s'effondra en raison de la conjonction de plusieurs facteurs: la pression de pêche, un fort évènement "El Niño", et la fin d'une période PDO froide. La biomasse d'anchois chuta, et il fallut plusieurs années pour que les stocks retrouvent leurs niveaux d'avant-crise. La flotte commerciale chuta de 1400 à 400 bateaux en quelques années. Dans les années 1980, le gouvernement péruvien comprit l'importance de mieux gérer les ressources marines, et établit un système de quota global afin de contrôler la pression de pêche sur les stocks. Des campagnes acoustiques pour évaluer les stocks furent mises en place. Grâce à ces mesures, lorsqu'en 1982, un fort évènement "El Niño" eut lieu, les stocks de poissons se rétablirent rapidement.

Il est à noter qu'un évènement "El Niño" seul ne peut pas expliquer les variations d'abondance des anchois ou des sardines. Les populations de sardines ont ainsi disparu de la côte péruvienne depuis la fin des années 1999 et n'ont pas réapparu depuis. Swartzman et al., 2008 proposèrent que l'épuisement de la sardine pouvait être associé au rôle du plateau continental et de la rupture du plateau continental: *"Quand l'habitat préférentiel se déplace offshore (ie quand l'upwelling augmente), ce qui arrive dans les périodes froides, la rétention des oeufs et larves de sardines est alors probablement réduite, diminuant la capacité des larves à s'alimenter et leur survie."*

Comme le montre la figure 1.9, durant les dernières décennies, la biomasse de sardines n'atteint jamais plus d'un tiers de la biomasse maximale d'anchois. Depuis les années 1960 on a vu une période dominée par l'anchois puis durant les années 1970-1980, une période durant laquelle la sardine est devenu plus abondante, mais il n'y a pas un régime régulier d'alternances. En fait, le paradigme des alternances d'anchois et sardines à l'échelle décennale a été questionné (ou même invalidé) par des études récentes en paléoécologie (Valdés et al., 2008; Gutiérrez et al., 2009, voir Fig. 1.5). Les régimes inter-décennaux (Périodes chaudes et froides de la PDO), la variabilité des événements "El Niño" (en termes de puissance et de durée), la pression de pêche, la variabilité de l'upwelling local, sont autant de facteurs qui expliquent probablement les périodes de variations d'abondance des anchois et sardines (Bertrand et al., 2004). Malgré de nombreuses hypothèses proposées pour expliquer ces différentes périodes de forte variabilité, leurs origines restent encore débattues.

Bertrand et al. (2011) ont montré récemment que l'oxygène pourrait être le lien manquant expliquant les régimes et la dynamique spatiotemporelle des petits pélagiques dans le HCS, plus particulièrement pour l'habitat de la sardine à l'échelle décennale. En effet, en observant que la profondeur de l'oxycline a été plus superficielle au cours de cette dernière décennie (Stramma et al., 2008), Bertrand et al., 2011 ont proposé que l'anchois pourrait vivre dans un habitat vertical plus superficiel tandis que la sardine nécessiterait une extension verticale plus importante de la couche de surface.

La profondeur de l'oxycline peut en effet être très superficielle le long de la côte péruvienne (jusqu'à seulement 10 mètres de profondeur). Les auteurs montrent que les densités d'anchois sont plus fortes quand la profondeur de l'oxycline est plus superficielle, et la relation inverse est observée dans le cas de la sardine. Par conséquent, si la sardine ne peut pas entrer dans les eaux du plateau continental, le processus de rétention des larves et moins efficace, la quantité disponible de nourriture chute, diminuant le succès à s'alimenter et leur survie. Dans les périodes plus chaudes, la sardine pourrait approcher la côte, et les oeufs et larves pourraient être mieux retenues (Fig. 1.10).
Ainsi, comme résumé sur la figure 1.10, une faible oxygénation des eaux de surface est favorable à l'anchois, lui permettant d'accéder aux fortes concentrations de macrozoo-

plancton à la rupture du plateau et permettrait d"expulser les espèces nécessitants des eaux plus oxygénées.

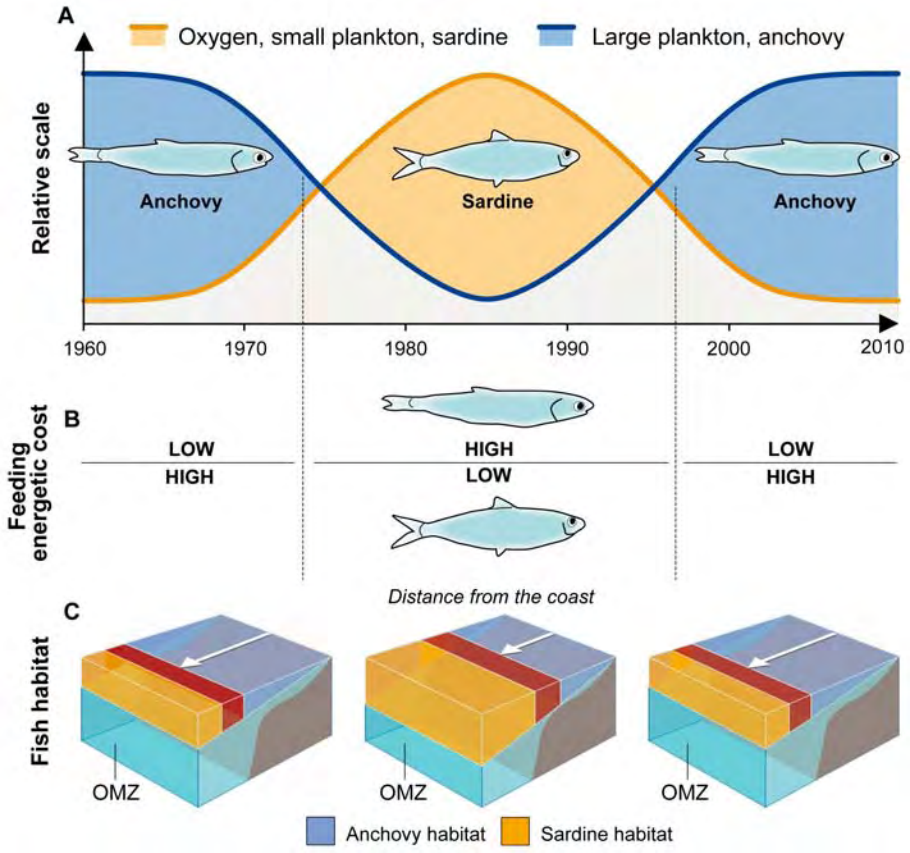

Figure 1.10: Modèle conceptuel des changements décennaux des populations d'anchois et sardines dans le Pacifique tropical sud-est, issu de Bertrand et al. (2011): *"A. Schéma de l'évolution temporelle du gros plancton et de l'anchois (ligne bleue), et oxygène, petit plancton et sardine (ligne rouge) entre 1960 et 2010.; B. Coût énergétique de l'alimentation en plancton de grandes tailles pour l'anchois et la sardine en accord avec les scénarios décrit en A.; C. Schéma de l'habitat disponible pour l'anchois (zones en bleue) et sardines (zones en rouge)."*

1.1.4 Approche par modélisation numérique

Pour modéliser le comportement et le mouvement des poissons, deux approches peuvent etre utilisées:

- L'approche lagrangienne suit individuellement les poissons qui sont dirigés par des lois comportementales (conditions "si/alors"). Le principal avantage de cette approche est qu'elle permet de décrire de façon explicite le comportement du poisson (en réponse à un forçage biologique ou environnemental). Ainsi, différents poissons d'une même cohorte peuvent avoir différents comportements.

  Dans l'approche Lagrangienne, le modèle tient compte explicitement des déplacements de chaque individu dans le temps et l'espace, et calcule la croissance de

l'individu, le comportement et le changement d'état (par exemple, de la larve au juvénile) comme une fonction de variables environnementales et physiologiques. Cette approche présente l'avantage de permettre une résolution spatiale supérieure à celle du forçage d'entrée en raison de la différentiation de chaque poisson. Ainsi, elle peut permettre de représenter des phénomènes de forte aggrégation d'indivus. Cependant, les modèles individus-centrés (IBM) peuvent vite s'avérer très coûteux en temps de calcul. En outre, il est plus difficile avec ces modèles de modéliser le cycle entier de la population.

- L'approche eulérienne calcule les flux de densités de poissons en utilisant des équations d'advection-diffusion-réaction. Elle consiste à étudier en un point fixe les modifications des propriétés de la densité de poisson en ce point. Elle permet ainsi de calculer facilement la variabilité spatiale de la densité en fonction du temps. Le principal désavantage de cette méthode est que les propriétés individuelles des poissons ne peuvent pas être prises en compte. En effet, on considère ainsi qu'au sein d'une même cohorte d'âge tous les poissons sont identiques. Néanmoins, si cette approche permet moins de détails, elle présente l'avantage en étant plus intégrative de nécessiter un nombre plus faible de paramètres. Le principal avantage de cette méthode est que l' implémentation des méthodes d'assimilation de données est beaucoup plus simple puisque les équations sont continues. De plus, le cycle entier (des oeufs aux adultes) est plus facile à modéliser.

Les modèles spatiaux développés pour étudier les relations des petits pélagiques avec leur environnement (simulé) sont principalement basés sur l'approche individu-centrée (IBM) (i.e., Lagrangienne) soit se focalisant sur les stades larvaires (e.g., Mullon et al., 2002; Lett et al., 2007; Brochier et al., 2008) ou sur la croissance (Megrey et al., 2007; Rose et al., 2007), soit en essayant de décrire le cycle entier de la population (Ito et al., 2007; Okunishi et al., 2009; Xi, 2009), ou encore les interactions entre espèces multiples (Travers et al., 2009).

En revanche, les tentatives d'utilisation de l'approche eulérienne pour calculer les flux de densités de poissons avec des équations d'advection-diffusion-réaction (ADR) sont rares (MacCall, 1990; Magnússon et al., 2004). Elles peuvent pourtant offrir un cadre de travail très avantageux pour modéliser la dynamique spatiale des populations et étudier la gestion des stocks, puisqu'elles contiennent un nombre limité de paramètres et des fonctions continues permettant le développement de modèles adjoints pour l'optimisation (Senina et al., 2008). Elles sont de plus un outil idéal pour modéliser les populations des oeufs jusqu'aux adultes.

Une approche de modélisation de dynamique des populations utilisant ce point de vue eulérien a été développée avec succès pour les grands prédateurs pélagiques

(i.e., thon et espèces proches) à l'échelle du bassin Pacifique (Lehodey et al., 2008, 2010). Les principales caractéristiques du modèle "Spatial Ecosystem And POpulation DYnamics Model" (SEAPODYM) sont: (i) un forçage par des données environnementales (température, courants, production primaire, concentration en oxygène dissous), (ii) des prédictions de la distribution spatiale et temporelle des groupes fonctionnels correspondant au niveau trophique moyen (micronecton) (Lehodey et al., 2010), (iii) des prédictions spatiales et temporelles des populations de thon en fonction de leur âge, (iv) des prédictions des captures totales et des fréquences de tailles par flottille quand l'effort de pêche est connu, et (v) l'optimisation des paramètres basées sur les techniques d'assimilation de données (Senina et al., 2008). De plus, puisque l'anchois et la sardine semblent être contrôlés par un forçage "bottom-up" (Chavez et al., 2008; Bertrand et al., 2008; Ayón et al., 2008), un modèle conduit par un forçage environnemental devrait être capable d'expliquer la variabilité de ces populations.

Bien que l'approche lagrangienne soit susceptible de mieux représenter les fortes aggrégations des petits pélagiques, nous avons choisi d'utiliser l'approche eulérienne via le modèle SEAPODYM, car elle permet d'optimiser les paramètres du modèle via des méthodes d'optimisation ainsi que de modéliser toute la population de l'espèce, des oeufs aux adultes.

Le schéma général du modèle SEAPODYM adapté à l'anchois et à la sardine est présenté en figure 1.11. Des détails supplémentaires concernant le modèle appliqué au thon sont disponibles dans Lehodey et al. (2008) et Senina et al. (2008). En résumé, la dynamique spatiale des poissons est basée sur des équations d'advection-diffusion-réaction simulant des mouvements aléatoires et orientés. Les courants transportent passivement les larves et les juvéniles, et les mouvements des jeunes et des poissons adultes sont dirigés par des indices d'habitats. L'advection et le taux de diffusion sont proportionnels à la taille du poisson, et le taux d'advection est proportionnel au gradient d'habitat, alors que le taux de diffusion décroit quand l'indice d'habitat augmente.

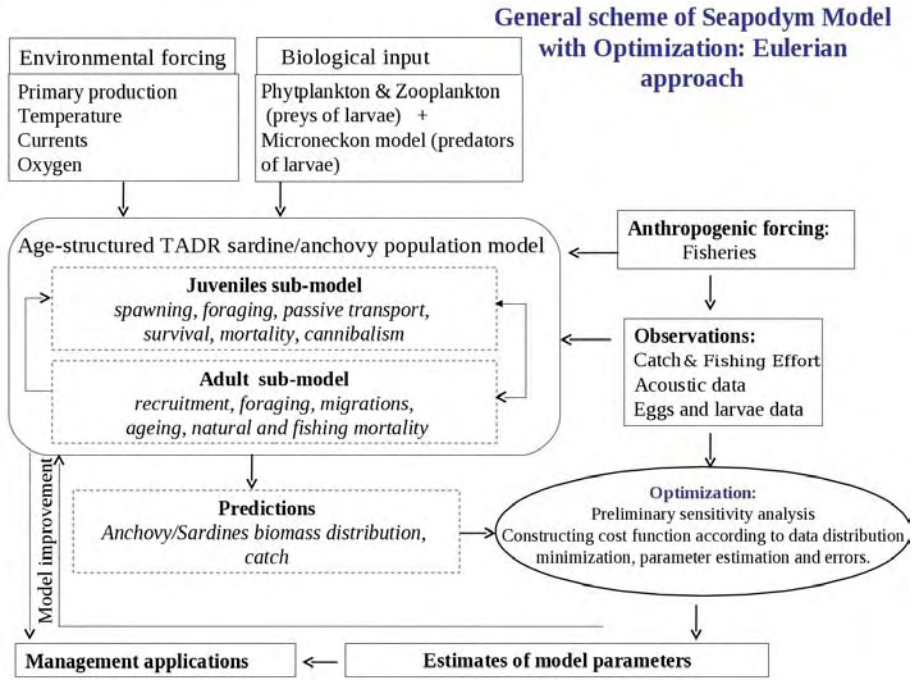

Figure 1.11: Schéma général du modèle SEAPODYM incluant l'approche d'optimisation, modifié d'après Senina et al. (2008)

1.1.5 Objectifs de la thèse

L'objectif principal de cette thèse est de développer un cadre de modélisation afin de tester les mécanismes expliquant les périodes de fortes abondances d'anchois et de sardines avec un modèle incluant la dynamique spatiale des deux espèces sous la contrainte de l'environnement. Les objectifs sont doubles : en premier lieu, adapter le modèle SEAPODYM à un domaine côtier pour les petits poissons pélagiques; ensuite rechercher si la variabilité environnementale et les dynamiques propres à chaque espèce (et le choix de leur habitat) peuvent expliquer ces périodes d'abondances.

Selon la litérature (Bertrand et al., 2004; Gutiérrez et al., 2007; Bertrand et al., 2008; Swartzman et al., 2008), la distribution de l'anchois est reliée aux eaux cotières froides (CCW) et eaux de mélange (MCW), indépendamment de la biomasse d'anchois. La population semble déterminée par la qualité de l'habitat ("habitat-based hypothesis"), et les changements de l'habitat déterminent les variations de taille du stock et de leur extension géographique. Cependant cette hypothèse basée sur l'habitat n'explique pas pourquoi la sardine, qui peut se distribuer dans une plus ample gamme de masses d'eau que l'anchois et est donc plus ubiquiste, s'est effondrée depuis la fin des années 90. Afin d'expliquer l'habitat de la sardine, la concentration en oxygène dissous a été proposée comme étant le lien manquant (Bertrand et al., 2011) et devrait ainsi être prise en compte dans la définition de l'habitat.

Ainsi, le modèle SEAPODYM qui combine la dynamique spatiale des populations avec la définition des habitats de chaque espèce devrait permettre de simuler la dynamique de ces populations.

Il existe cependant plusieurs défis pour adapter ce modèle aux petits poissons pélagiques:

1. Le domaine du modèle dans le cadre des applications au thon est le bassin océanique dans son ensemble, et à une résolution relativement faible (typiquement 1 degré par mois) lors de l'optimisation des paramètres. En revanche, les stocks d'anchois et de sardines dans le HCS sont côtiers, et les principaux facteurs qui contrôlent l'abondance des oeufs et larves (enrichissement, concentration et rétention), comme identifié par Bakun (1996), sont fortement reliés à l'activité de mésoéchelle. Par conséquent, notre approche nécessite un modèle régional à haute résolution et avec une représentation précise des structures des phénomènes physiques et biogéochimiques de mésoéchelles.

2. Le type d'observation pouvant être utilisé pour l'optimisation des paramètres est différent. Pour la pêcherie du thon, il existe un long historique de pêche (50 ans) avec des enregistrements relativement détaillés des captures totales de pêche, de l'effort et la fréquence de taille de capture sur l'ensemble du bassin océanique. Pour l'anchois et la sardine, les données historiques de captures sont basées sur les débarquements par port (i.e., pas d'information spatialisée), et l'activité de pêche peut être intense mais concentrée sur une très courte période de temps, qui ne donnent pas par conséquent d'information sur le reste de l'année (Fréon et al., 2008). Cependant des informations supplémentaires précieuses ont été collectées par l'Institut de la Mer du Pérou (IMARPE, http://www.imarpe.pe). Oeufs, larves et biomasses acoustiques des poissons ont été enregistrés au cours de campagnes régulières ayant eu lieu le long de la côte du Pérou depuis 1961 pour les oeufs et larves, et 1983 pour le suivi acoustique. Ces données sont utilisées pour mettre en oeuvre l'approche d'assimilation. Ces modifications sont décrites dans le chapitre 6 et 7.

3. La modélisation de la dynamique spatiale de populations de poisson dans SEAPODYM nécessite un champ de proies, ainsi qu'un champ représentant la biomasse des groupes de niveau trophique moyen (i.e. micronecton) qui sont les principaux prédateurs des larves de poissons. Les champs de proies pour le thon sont le micronecton, tandis que dans le cas des petits poissons pélagiques comme l'anchois et la sardine, le champ de proies des larves est un groupe contenant phytoplancton et zooplancton. Pour le thon, les prédateurs et les proies des larves sont donc modélisés par le même groupe, tandis que pour l'anchois et la sardine, proies

et prédateurs des larves du modèle nécessitent d'être différenciés et adaptés à ces espèces. Ces modifications sont décrites dans les chapitres 4 et 5.

4. Entre le thon et l'anchois, les mécanismes définissant l'habitat de ponte peuvent être différents et nécessitent des révisions. Ceci est décrit en détail dans le chapitre 2.

Etant donné que les mécanismes de ponte et de recrutement larvaire déterminent la dynamique des populations (Beverton and Holt, 1957), et que la plupart des informations collectées correspondent aux premiers stades de vie, cette thèse se focalise sur l'adaptation du modèle SEAPODYM à l'habitat de ponte et à la dynamique larvaire de l'anchois et la sardine, suivant le schéma général du modèle pour les premiers stades de vie décrit dans la Figure 1.12.

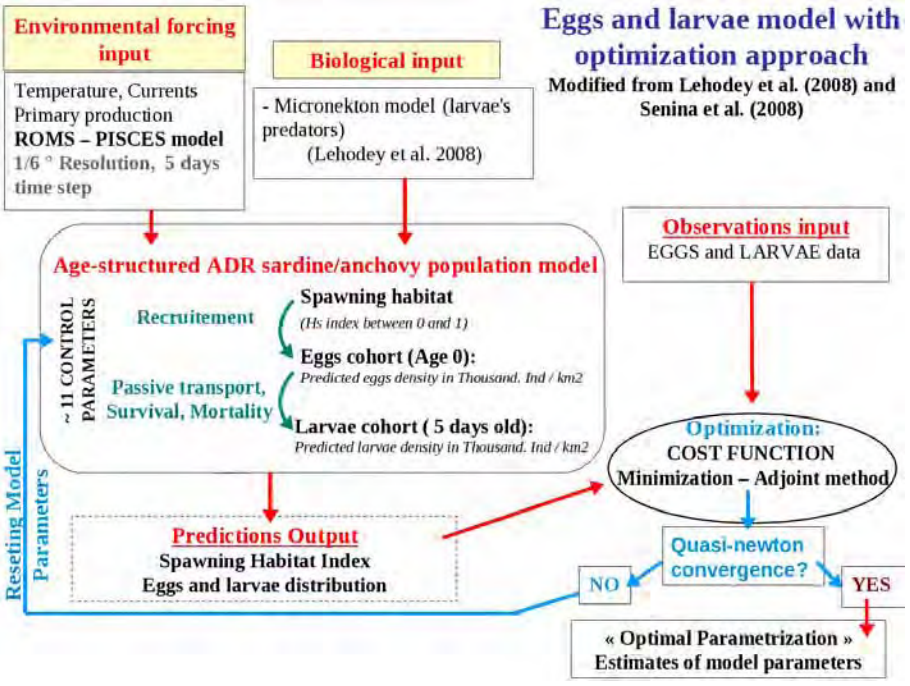

Figure 1.12: Schéma général du modèle avec l'approche d'optimisation.

1.1.6 Structure de la thèse

Cette étude est organisée en 7 chapitres qui suivent l'organisation du modèle SEAPODYM pour les premiers stades de vies, décrit dans la Figure 1.12.

- Le chapitre 2 décrit les fondamentaux de la nouvelle version de SEAPODYM adaptée aux premiers stades de vies de petits poissons pélagiques. Les différentes hypothèses et équations de cette version du modèle y sont décrites, ainsi que les différences avec la version initiale du modèle SEAPODYM pour le thon.

- Le chapitre 3 se focalise sur les observations disponibles et décrit les données d'oeufs, larves et adultes d'anchois et sardines de l'IMARPE qui seront utilisées pour valider le modèle SEAPODYM et mettre en oeuvre l'approche d'optimisation. Ce chapitre présente les principales caractéristiques que nous tenterons de reproduire avec notre approche de modélisation, ainsi que les sources d'incertitudes de ces données.

- Dans le chapitre 4, le forçage environnemental d'entrée est décrit. Pour faire tourner le modèle anchois et sardine avec SEAPODYM, nous utiliserons les champs de température, courant et production primaire de 3 différents run ROMS-PISCES : un run climatologique et un run interannuel allant de 1992 à 2008 à la résolution de 1/6° et un run interannuel allant de 2000 à 2006 à la résolution de 1/9°. Les modèles physiques de l'océan se sont beaucoup améliorés au cours de la dernière décennie, mais peuvent encore avoir des biais à haute résolution. Comme les résultats de notre modèle dépendent de ces variables d'entrée, il est nécessaire d'analyser les différentes variables physiques et biologiques utilisées pour faire tourner le modèle, ce qui inclut le modèle micronecton.

- Le chapitre 5 concerne l'application du modèle SEAPODYM aux premiers stades de vie des petits poissons pélagiques. Ce chapitre se présente sous la forme d'un article qui sera prochainement soumis. Nous montrons que la combinaison de mécanismes simples avec un nombre limité de paramètres pour décrire la dynamique de l'anchois conduit rapidement à une multitude de solutions que seule une approche rigoureuse d'optimisation des paramètres peut explorer.

- Dans le chapitre 6, je reviens brièvement sur les méthodes d'assimilation utilisées dans les sciences marines, et je présente les principaux concepts nécessaires à la compréhension de l'implémentation de l'assimilation de données d'oeufs et larves dans SEAPODYM. Dans ce chapitre, nous nous intéressons plus particulièrement aux méthodes variationnelles utilisées pour chercher la solution optimale aux problèmes d'assimilation de données. Différents estimateurs peuvent être utilisés pour définir l'optimalité d'une solution donnée. Ici, nous présentons deux estimateurs optimaux: le "Maximum A Posteriori estimation (MAP)" et le "Maximum Likelihood Estimator (MLE)". Le premier estimateur est illustré avec l'exemple du suivi de baleines bleues à partir d'enregistrements de leurs chants à l'aide de sismomètres de fonds et mer et d'un réseau d'hydrophones, travail qui a été publié au début de ma thèse (voir Annexe C). Le deuxième estimateur est celui utilisé dans SEAPODYM.

- Le chapitre 7 présente les résultats des expériences d'optimisation des paramètres régissant les premiers stades de vie des anchois et sardines, opérés sur la base du run climatologique d'entrée présenté dans le chapitre 4. Grâce à une étude de sensi-

bilité des principales sources de variabilité et d'incertitudes de notre modèle, nous conclurons sur les principaux mécanismes qui influencent fortement ses résultats. Je propose différents axes de recherche susceptibles de permettre d'améliorer le modèle. Je présente également les résultats préliminaires des optimisations réalisées à partir des deux runs interannuels présentés dans le chapitre 4.

- La conclusion générale résume les résultats principaux de ce travail de thèse, souligne les limites de l'approche de modélisation et propose quelques perspectives de développements futurs.

1.2 Introduction (English version)

1.2.1 Motivations

Within the last 60 years, the world population as grown by 4.25 billions of people, reaching today 7 billions, and causing an increasing pressure on marine resources. In many parts of the world, this demographic evolution came along with economic growth and diversification of protein sources, modifying the human food habits towards an increasing demand in products of animal origin (meat and fish). Indeed, the average annual fish consumption increased from 9.9 kg per capita in the 1960's to 17 kg in 2007 (FAO, 2010). This exacerbate demand for fish has caused an explosive increase of the world fisheries production, from 19 millions of tons in 1950, to 90 millions of tons in 1990 and 145.1 millions of tons in 2008 (FAO, 2010). At the end of the 1980's, the total marine captures peaked at 80 to 85 millions tons (Pauly et al., 2003) and then declined to 79.5 millions of tones in 2008. Marine fisheries have probably passed over their maximum sustainable yield after reaching their maximum (Pauly et al., 2002). The recent growth of fisheries production is due to the growth of the inland and marine aquaculture production, which provided in year 2008 46 % of the total fish supply (FAO, 2010).

The decline of the marine catches, and the correlated increase of aquaculture production, is related to the overexploitation of most fish stocks. According to the FAO, the proportion of world fish stocks underexploited or moderately exploited has declined from 40 % to 15% between the mid-1970's and 2008, whereas the proportion of overexploited, depleted or recovering stocks has increased from 10% to 32% of the 600 marine fish stocks monitored by the FAO (FAO, 2010)

In 2008, only 3% of the stock were underexploited, 1% recovering from depletion and 12% moderately exploited.

52% of the stocks were fully exploited, and therefore at their maximum sustainable productions.

As a result, analysis of global fisheries data show a decrease in the mean size of individual fish catches (FAO, 2010), illustrating the pressure exerted by human fishing on the world fish stocks. The collapse of the North-Atlantic cod stock in 1992 (Myers et al., 1997) or the strong decline of the Bluefin tuna in the Mediterranean sea (MacKenzie et al., 2009) are well-known consequences of overfishing.

A better management of fish stocks has became a necessity in order to restore overfished populations to sustainable levels. 90 % of the world catches take place in the Exclusive Economic Zones (EEZ) (Rothschild, 1996), a 200 miles-wide band along the coastline. However, as observed in Figure 1.13, not all coastal areas are equally productive.

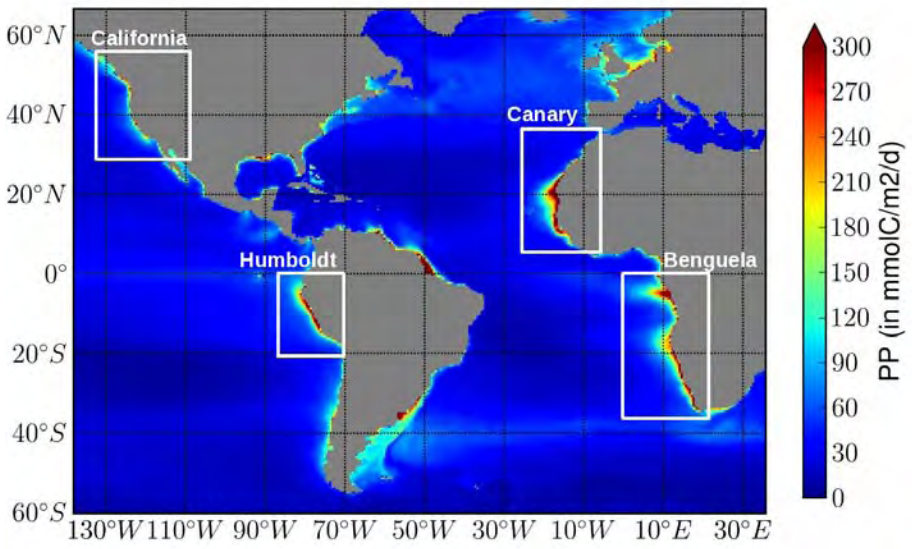

Figure 1.13: Annual mean of satellite-derived primary production (based on the VGPM algorithm Behrenfeld and Falkowski, 1997). Data from October 1997 to December 2008. The four main Eastern Boundary Upwelling System (EBUS) are characterized by "local wind-driven upwelling, strong alongshore advection, a poleward undercurrent, very low to moderate precipitation and high productivity of plankton and fish, especially pelagic fish" (Fréon et al., 2009).

There are 4 main Eastern Boundary Upwelling System (EBUS): the Humboldt Current system (HCS) in the South Pacific, the Canary Current system (CanCS) in the North Atlantic, the Benguela Current system (BCS) in the South Atlantic and the California Current system (CalCS) in the North Pacific. They represent less than 0.1% of the world surface oceans but 30% of the world's fish catches (Durand et al., 1998).The productivity of these areas is enhanced by the nutrient-rich waters upwelled in the euphotic layer, a phenomenon generated by the trade winds blowing equatorward along the coastline. The ecosystems of these upwelling systems are dominated by small pelagic fishes, mainly sardine and anchovy species (Bakun, 1996).

Among these Eastern Boundary Upwelling System, the Humboldt Current system is the third most productive in terms of primary production, and the most productive system in terms of fish biomass (Bakun and Broad, 2003; Chavez et al., 2008, Fig. 1.2). The system is dominated by Peruvian anchovy, known as anchoveta, which is nowadays exploited by the biggest commercial fleet targeting a single species, with total catches reaching up to 12 millions of tons the best years. In 2008, annual landings of anchoveta represented 7.4 % of the world fish catches, therefore representing the biggest single-species fish catch (FAO, 2010). 92 % of these landings are dedicated to the fishmeal and fish oil industries (Durand and Seminario, 2009), which are mainly exported for aquaculture. As a result, 30 to 35% of the worldwide fishmeal and fish oil comes from the Peruvian anchovy fishery (International Fishmeal and Fish Oil Organisation, IFFO, 2009).

This industry contributes to about 1.2% of Peru GDP (in 2007), Bertrand et al., 2010), and employ directly and indirectly more than 125 000 persons (Bertrand et al., 2010). It is thus primordial for the Peruvians to protect, control and manage this resource.

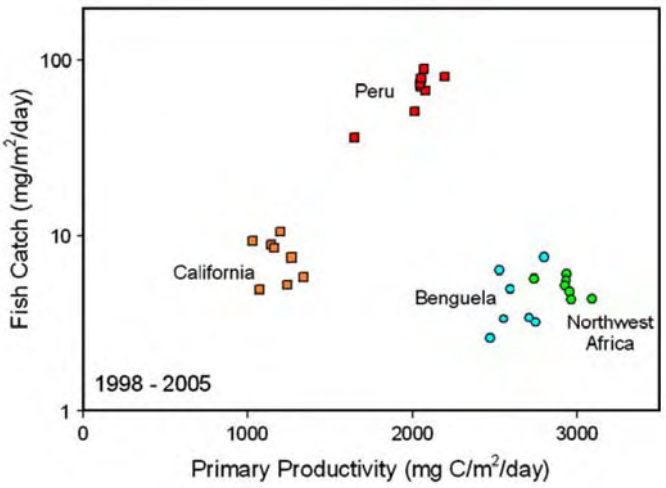

Figure 1.14: Fish catch versus satellite-derived primary productivity (Behrenfeld and Falkowski, 1997) for the four main eastern boundary coastal upwelling ecosystems for years 1998 to 2005. Source: Chavez et al., 2008.

This illustrate the importance of anchoveta both as a central species in the Humboldt Current system, as a major target for the world fishing and aquaculture industry. With the growing demand of aquaculture for small pelagic fish as food for cultured fish (Checkley Jr et al., 2009), the will to exploit this resource is high. However, according to the FAO, the anchoveta stock is already fully exploited. With a modern and well-equipped commercial fleet, with over 1200 purse seiners (Fréon et al., 2008), Peruvian fishermen can fish up to 170 000 tons of anchovy in a single day. Without scientific and governmental control, the anchoveta fisheries couldn't be sustainable. After the collapse of the anchovy in the seventies, annual global quota have been introduced and fishery decisions have been made in quasi real time using the most recent scientific observations. Open fishing season vary year to year depending on the scientific advise. Usually fishery open two times by year around April to July and October to January when the spawning is limited. Fishery is also subject to different restrictions and stops in real time when juveniles reach more than 10% of the landings, the fish size is lower then 12cm, etc. The fisheries management on the northern HCS is known as the most rapidly "adaptive" management in the world (Chavez et al., 2008).

However, the overcapacity of the Peruvian purse seiner fleet is estimated at over 300% due to the race to catch a larger share of the global annual quota and to the inability of fisheries managers to control the expansion of the fleet and fish processing plants (Fréon et al., 2008). With this system, global annual quota were reach in 50 days in 2007 (Fréon et al., 2008), as opposed to 270 days in 1986. Therefore, since 2009, a management using individual vessel quotas (IVQs) has been introduced, aiming at a reduction of the number of vessels participating in the fishery, controlling the race for fish and prolonging

the fishing season (Tveteras et al., 2011).

In the EBUS's, small pelagic fish occupy a key position in marine ecosystem since they dominate mid trophic levels and are composed by relatively few species but attaining large abundances that can vary drastically in size (Checkley Jr et al., 2009).

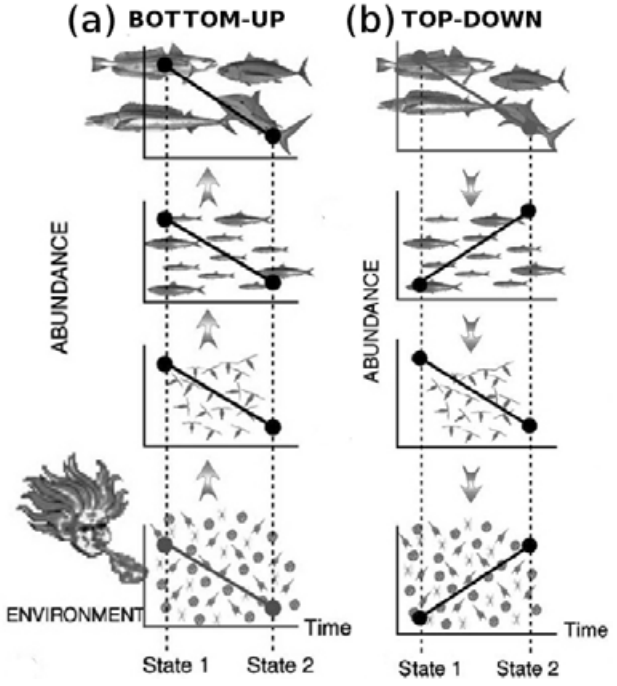

Figure 1.15: (a) Bottom-up control within a simplified four-level food web in a marine ecosystem; (b) Top-down control. Adapted from Cury and Shannon (2004)

Their stocks can respond to a change from below, as an environmental forcing (bottom-up control), or from above, as fishing forcing (top-down control) (Cury et al., 2000, Fig. 1.15).

However, bottom-up forcing appears to be the main forcing in the HCS at interannual, multidecal and centennial scales (Ayón et al., 2008; Bertrand et al., 2008; Chavez et al., 2008). Indeed, the small pelagic stocks in the HCS are highly variable at various times scales, as a result of the high variability of ocean conditions due to climatic stresses at different time scales: oceanographic variations at secular scale, the Pacific decadal oscillation (PDO) at multi-decadal scale, ENSO (El Niño Southern Oscillation) at the interannual scale, and the seasonal variability at intra-annual scale (Chavez et al., 2008).

The strong climate variability affects directly the anchovy and sardines stocks which fluctuate through time. These fluctuations, observed also in other EBUS (Lluch-belda et al., 1992; Alheit and Niquen, 2004), have questioned oceanographers and fisheries scientists for decades, leading to a considerable literature and several still debated hypotheses (Chavez et al., 2008). In addition, this strong variability makes the management of the fisheries difficult. The societal and economical impact of fish catches fluctuations are major (see section 1.3). In this context, the development of a model that could

help to better understand the impact of the environmental variability on exploited fish populations would represent a significant step forward.

Furthermore, in the context of the global and regional climate change, it is crucial to develop and integrate scientific tools to address the impact of climate change on marine resources. This is the aim of the Peru Ecosystem Projection Scenarios (PEPS) project, funded by ANR (Agence Nationale de la Recherche), the French research funding agency. This project is made with the collaboration of the Peruvian Research Institute (Instituto del Mar del Peru (IMARPE)), the Institut de Recherche pour le Developpement (IRD), the Laboratoire Mixte International (LMI DISCOH) and the company CLS (Collecte Localisation Satellites). The project aims at assessing the impact of the climate change on two major fisheries in the Northern HCS: the sardine and anchovy. However, before being able to forecast, we need to build a model which is able to reproduce the variability of the stocks over these last fifty years. This will make up our fist and main goal.

Models including a representation of the whole ecosystem with its physical and biogeochemical forcing are needed to test at the same time the climatic change and the fishing pressure scenarios. In this context, the development of a model that could explain the environmental variability of populations by accounting rigorously for the fishing mortality and stock dynamics parameters, would provide a considerable move forward to develop an ecosystem-based management approach.

One modeling approach with similar objectives has been developed recently but focusing on large pelagic species (i.e. tuna and tuna-like species) at basin-scale (Lehodey et al., 2008, 2010) with the Spatial Ecosystem And POpulation DYnamics Model (SEAPODYM). The initial idea of this PhD project was to adapt the SEAPODYM model to anchovy and sardine in the HCS, from eggs to adults, while accounting for the fishing impact. However, as challenges to adapt a basin scale model to the regional scale and small pelagic fishes are multiple, this thesis focus on the early life stages of the population of anchovy and sardine, that control largely the dynamics of the whole population. Furthermore, the second main objective was to develop data assimilation methods for ecosystem model in order to better approach the reality of the observations.

1.2.2 The Humboldt Current system: Oceanography

The Humboldt Current system (HCS) off Peru and Chile is notable for several reasons. Following the description of Chavez et al. (2008) we summarize the characteristics of this system in 4 main points.

First, as mentioned previously, this system is characterized by the presence of cold nutrient rich waters flowing at latitudes where average temperature should be higher

(Chavez et al., 2008). Figure 1.16 presents a scheme of the oceanic circulation in this area. The general oceanography of the HCS is characterized by a predominant northward flow of surface waters of subantarctic origin and by a strong upwelling of cool nutrient-rich subsurface waters of equatorial origin (Thiel et al., 2007).

Along the Peruvian and Chilean coast, the south-east trade winds are forced by the Andes mountain to blow northward (Penven et al., 2005). The winds blowing northward, the superficial water are transported to the west by the Ekman transport resulting from the Coriolis effect. The winds being almost constantly favourable, the upwelling is present all year round.

The surface layer is dominated along the coast by the Peru Coastal Current (PCC) also named Humboldt Current, that is associated with the coastal upwelling of cold and salty water. Below a shallow upper layer ($\sim 20m$) of this equatorward current, the Peru-Chile under-current (PCUC) flows polewards and dominates the subsurface and the shelf. At 5°S the PCUC has been observed at a depth of about 50-100 m and it is going deeper polewards.

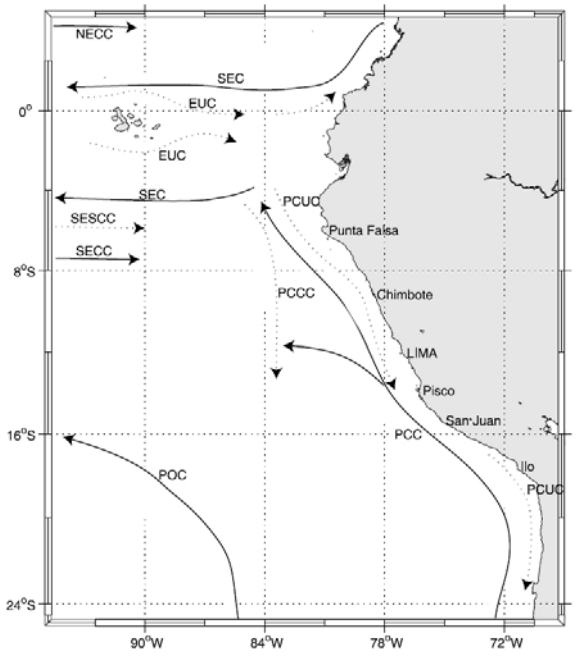

Figure 1.16: Oceanic circulation scheme for the Peruvian Current System (Penven et al., 2005.

Secondly, the HCS of Peru, as mentioned previously, is the most productive Eastern Boundary Upwelling Ecosystem in terms of fisheries. No direct relationship is observed between fish catches and the primary productivity estimated by satellite (Fig. 1.14), and this paradox which questioned scientists during this last decades (Bakun and Weeks,

2008; Chavez et al., 2008; Ballón et al., 2011; Brochier et al., 2011) remains open.

Thirdly, as a result of sinking and decay of surface-derived primary production and poor ventilation, surface oxygenated waters overlie the most shallow, intense and expanded mininum oxygen zone (OMZ) of all oceans (Chavez et al., 2008).

Fourthly, due to the proximity with the Equator, the northern HCS is intimately linked to equatorial Pacific dynamics, and particularly to the El Niño-Southern Oscillation (ENSO), the strongest climate signal on an interannual time scale (McPhaden et al., 2006).

Overall, the HCS is characterized by a high variability due to climatic processes interacting on different different time scales:

- **Secular Variability**

  Studies of time series from marine sediments off Peru showed *"an abrupt centennial-scale biogeochemical regime shift in the early nineteenth century, of much greater magnitude and duration than present day multi-decadal variability"* (Gutiérrez et al., 2009). This centennial-scale shift which begin around 1820 AD (Fig 1.17) was probably driven by a northward displacement of the Intertropical Convergence Zone (ITCZ) and the South Pacific Subtropical High to their present day locations, coupled with an enhancement of the Walker circulation. As a consequence, the upwelling was enhanced, increasing the high productivity of the area and therefore the biomass of pelagic fish (Fig 1.17).

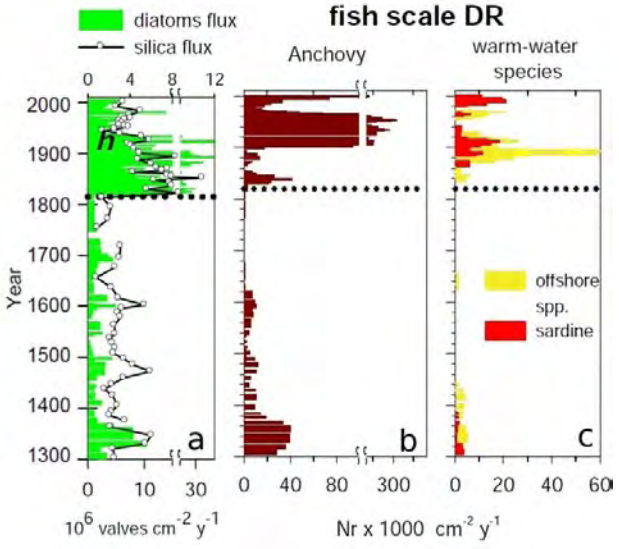

Figure 1.17: Proxies of primary production ecosystem (a), anchovy fish biomass (b) and sardine fish biomass (c). (a) bars: diatom accumulation rate (10^6 valves $cm^{-2}.y^{-1}$), white circles: flux of biogenic silica ($mg.cm^{-2}.y^{-1}$); (b) 3-term running averages of anchovy scale deposition rates ($Nr \times 1000\ cm^{-2}.y^{-1}$);(c) 3-term running averages of sardine scale deposition rates and offshore pelagic (jack mackerel + mackerel) scale deposition rates ($Nr \times 1000\ cm^{-2}.y^{-1}$). From Callao sediments. Source: Gutiérrez et al. (2009)

- **Decadal Variability**
  The Pacific decadal oscillation (PDO) is a long-term ocean/atmosphere fluctuation of the Pacific Ocean. It has been described as an analog climatic phenomenon of "El Niño", but with longer duration: 20 to 30 years. The decadal warm and cold phases distinguish oneself by an anomaly of sea surface temperature of 1°C to 2°C. Causes for the PDO are still unknown. The PDO index is defined as the variability of the sea surface temperature in the northern Pacific, as shown in Figure 1.18.

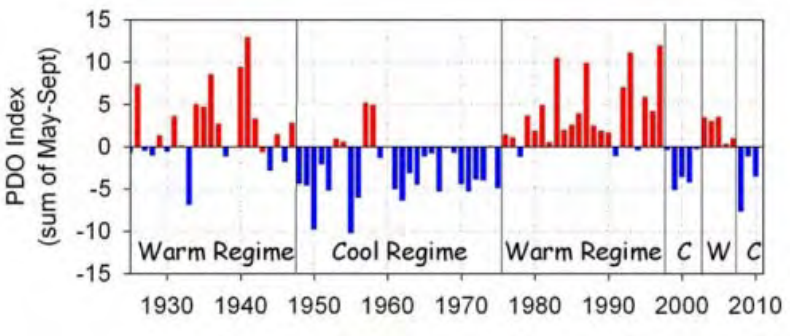

Figure 1.18: Time series of shifts in the phase of the Pacific Decadal Oscillation (PDO), from 1925 to 2010. Red bars indicate positive (warm) years; blue bars indicate negative (cool) years. The cold regime occurred from 1950 to 1971 and from 1999 to present, and the warm regime from 1972 to 1998. Credit: NOAA PMEL.

  In the "warm" periods (red), the intensity of the Peruvian upwelling and the productivity is lower and the thermocline deeper. In the "cold" periods (blue) the situation is opposite.

- **Interanual variability**
  The El Niño Southern Oscillation (ENSO) phenomenon is due to a periodic instability of the ocean-atmosphere dynamics in the Ocean Pacific basin. El Niño and la Niña events represent the opposite phases of the ENSO cycle. The El Niño event represents the warm phase, the sea-surface temperature is warmer in the eastern tropical pacific than usual. The La Niña event represents the cold phase. The frequency of ENSO is irregular and varies between 2 and 7 years.

  The southern oscillation index (SOI) is a measure of the large-scale fluctuation in air pressure occurring between the western and eastern tropical Pacific. It is designed to measure the strength and phase of the Southern Oscillation. The index is calculated based on the difference in air pressure between Tahiti (French Polynesia) and Darwin (Australia). A strong negative index (red values) indicates an El Niño event while a positive index (blue values) reveals a La Niña event (Fig. 1.19).

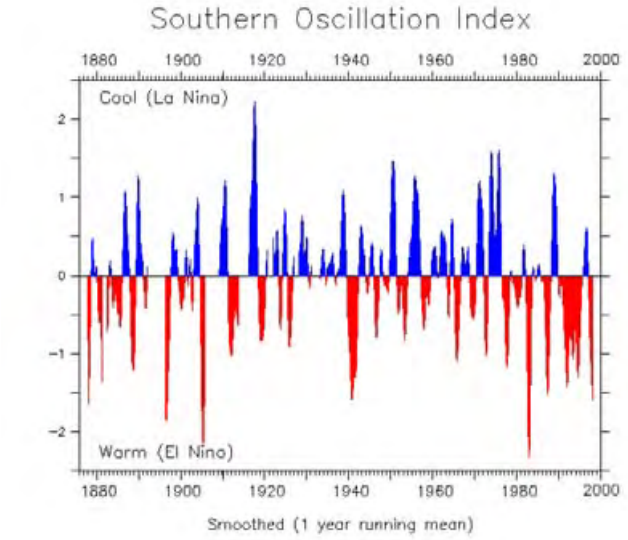

Figure 1.19: Southern Oscillation Index (SOI). Strongest El Niño events occurred in the years 1982-83 and 1997-98. Source:NOAA

- **Seasonal variability**

  At the intra-annual scale, the variability is driven by the seasonal variability of upwelling intensity, luminosity and temperature and by the variability of mesoscale structure (eddies, plumes).

  The upwelling system is present all year round but varies in intensity between seasons. This system is perplexing since transport and chlorophyll are completely out of phase (Chavez and Messié, 2009). Indeed, the maximum intensity of the HCS upwelling is during the Austral Winter while the minimum intensity is in Austral Summer. This is in opposite phase with the surface chlorophyll concentration, which is minimum in austral winter and maximal in austral summer (Fig. 1.20).

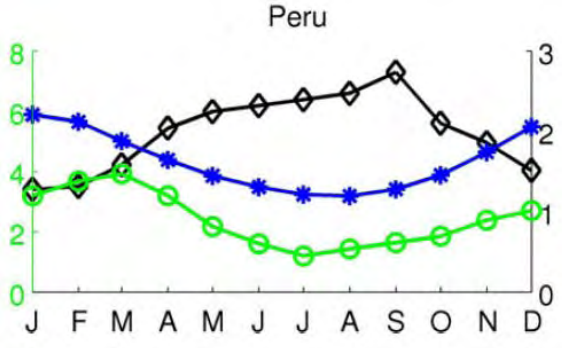

Figure 1.20: Seasonal cycles of chlorophyll concentration (gray, SeaWiFS 1997-2007), sea surface temperature (blue, Reynolds 1981- 2007) and total vertical transport (black, Ekman transport plus Ekman pumping, estimated from QuikSCAT winds 1999-2008) averaged between 6°S and 16°S in a band of 150km offshore. Source: Chavez and Messié (2009)

Echevin et al. (2008) suggested that this parodoxal seasonal cycle could be explained by the seasonal variability of the mixed layer depth: in winter, when the intensity of the upwelling is maximal, the mixed layer is deep, the dilution of the nutrients

is higher and the luminosity lower, resulting in a minimum of chlorophyll. In summer, the mixed layer is shallower, the luminosity higher and all the phytoplankton biomass is concentrated in the surface layer, resulting in a chlorophyll maximum.

1.2.3 Anchovy and Sardines populations

We focus on the two main population of small pelagic fish in Peru and Chile: Anchoveta (*Engraulis ringens*) and Sardine (*Sardinops sagax*). Both fishes belong to the Clupeiformes order, anchovy comes from Engraulidae family and sardine from Clupeidae family. These two species have a relatively short life span (around 4 and 8 years respectively) with a maximal size of 20 cm and 40 cm respectively, a fast growth and early maturity (one and two years respectively). Anchovy and sardine feed mainly on zooplankton, but anchovy appears to be more effective feeding by direct biting (particulate feeding) on large copepods and euphausids wheres sardines appears as primarily filter-feeders, feeding on smaller zooplankton (small copepods and fewer euphausiids) (Espinoza and Bertrand, 2008; Espinoza et al., 2009; Ayón et al., 2011, Fig. 1.22).

Highly fecund and being able to spawn all year-round, these species are well adapted to the variability of the HCS (Bertrand et al., 2004; Gutiérrez et al., 2007; Espinoza and Bertrand, 2008; Swartzman et al., 2008), but still highly sensitive to environmental forcing and therefore responding quickly and sometimes dramatically to changes in ocean climate.

Within the last decades, different periods with high abundances of anchovy followed by high abundances of sardines (and vice-versa) have been observed (Fig 1.21). Anchovy landings can vary from 1 to 12 millions of tons per year. In the seventies, anchovy fisheries collapsed due to several factors: fishing pressure, one strong El Niño event, and the end of a cold PDO period. The biomass of the anchovy collapsed and it took several years to recover the previous stock levels. The commercial fleet collapsed from 1400 to 400 boats in a few years. In the 1980's, the government understood the importance to better manage that resource and established global quotas to control the fishing pressure. Acoustic campaigns to evaluate the stocks were set up. As a result, in 1982, a powerful El Niño event occurred, but the fish stock recovered relatively quickly after. Indeed, El Niño events alone cannot explain the variability of anchovy and sardine abundances. The sardine population disappeared from the Peruvian coast after 1999 and did not reappear since. Swartzman et al., 2008 proposed that the depletion of sardine could be associated with the shelf and the shelf break: *"When its preferred habitat moves further offshore (i.e. increased upwelling), as happens during cool periods, retention of eggs and larvae of sardine may be reduced, weakening larval feeding success and survival"*.

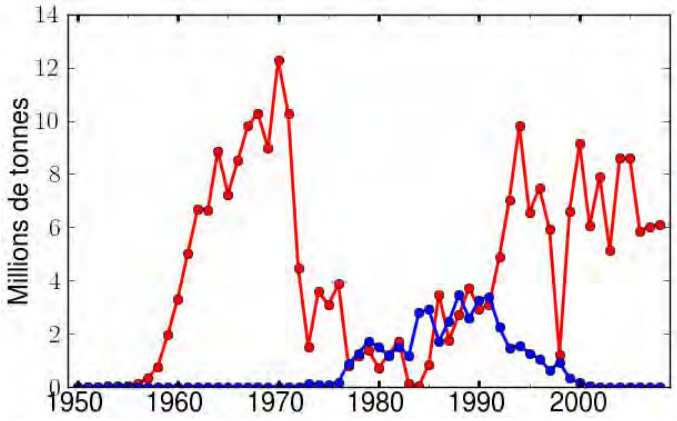

Figure 1.21: Historical landings for anchoveta (*Engraulis ringens*) and sardines (*Sardinops sagax*) in the HCS between 1950-2008 (IMARPE data). Total catch in a year is used here as an abundance index. However, this assumption is true only if the fishing effort and the catchability remain constant over the study period. Therefore, for anchovy fishery, catch is supposed a good index of biomass after 1965, year when the industrial anchovy fishery began, despite fishing effort was regularly increasing due to the improvement of the fleet capacity (Fréon et al., 2008).

As seen in Figure 1.21, in the last decades, the biomass of sardines never reached more than a third of the maximal anchovy biomass. Since 1960, we have seen a period dominated by anchovy, then between 1970 and 1980, a period where sardine began more abundant. In fact, the paradigm of regular alternation between anchovy and sardine regimes at a decadal scale was questioned (or even invalidated) by recent studies in paleoecology (Valdés et al., 2008; Gutiérrez et al., 2009, see Fig. 1.17). Interdecadal regime (Warm and cold periods from the PDO), variability in the Niño events (strength and duration are different for each Niño), fishing pressure, variability of the local upwelling are together factors which are likely to explain the high variability of anchovy and sardines abundances (Bertrand et al., 2004). Despite all the hypotheses proposed to explain the different periods of high variability of abundances, their origin are still debated. However, Bertrand et al. (2011) has showed recently that oxygen could be the missing link explaining pelagic fish spatiotemporal dynamics and regimes in the HCS, specially for sardines habitat at decennal scale.

Observing that the oxycline depth has been shallower during this last decade (Stramma et al., 2008); Bertrand et al. (2011) proposed the hypothesis that anchovies can live in a shallow vertical habitat while sardines need a higher vertical extension. Indeed, the oxycline depth could be as shallow as 10m along the Peruvian coast. The authors showed that anchovy concentrations were higher when the oxycline depth was shallow and an inverse relationship existed in the sardine's case. Thus, if the sardine population can not enter the waters of the continental shelf edge, the larvae retention's process is less efficient, the quantity of food drops, weakening larval feeding success and survival. In the warmer periods, the sardine population could approach the coast, and the eggs

and larvae could be retained (Fig. 1.22).

Therefore, as observed on Fig. 1.22, a large offshore low oxygen is favorable for anchovy, allowing it access to the high concentration of macrozooplankton on the shelf break and reducing the "expulsion" of species that need more oxygenated water.

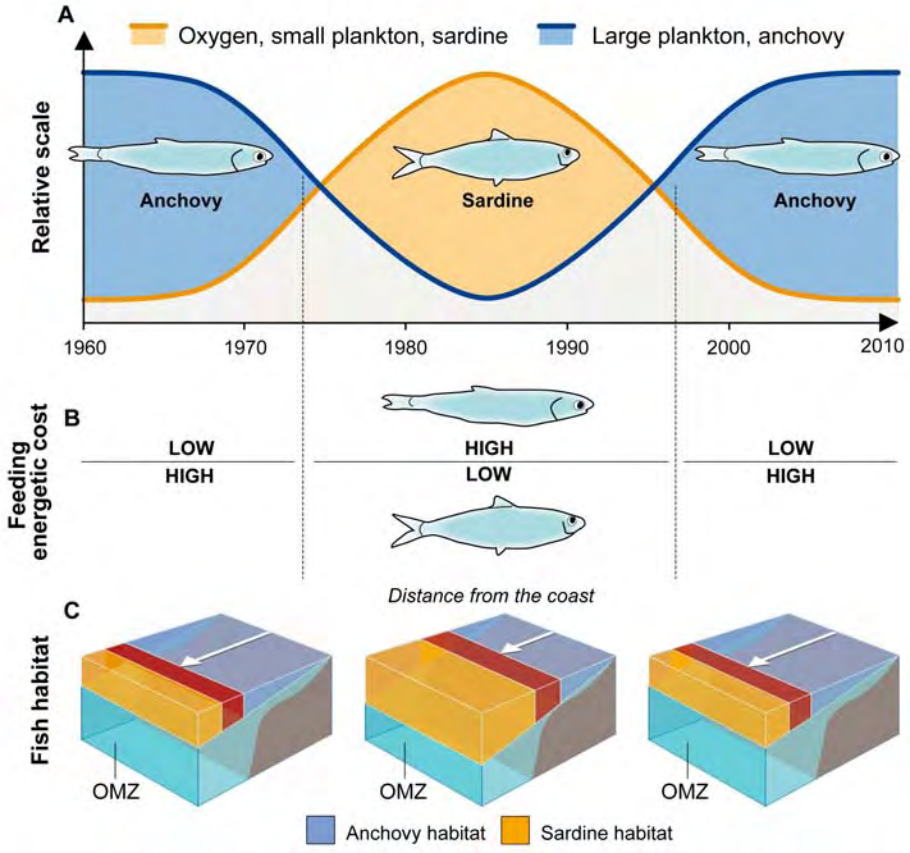

Figure 1.22: Conceptual model of decadal changes in anchovy and sardine populations in the southeastern tropical Pacific from Bertrand et al. (2011): *"A. Schematic of the temporal evolution of large plankton and anchovy (blue solid line), and oxygen, small plankton and sardine (red solid line) between 1960 and 2010. B. Energetic costs of feeding on dominant plankton size-spectra for anchovy and sardine according to the scenarios from A. C. Schematic of the available habitat for anchovy (blue shaded area) and sardine (red shaded area)."*

1.2.4 Approach by numerical modelling

To model fish behaviour and movement, two approaches can be used:

- The Lagrangian approach tracks individual fishes governed by behavioral rules (i.e. if/else conditions). The main advantage of this approach is that it allows to explicitly describe the fish behaviour (in response to biological or environmental forcing). As a result, different fishes of the same cohort may have different behaviours.

In the Lagrangian approach, the model explicitly accounts for the displacement through space and time of each individual, and computes the individual growth, behaviour, and change of state (e.g larvae to juveniles) as a function of environmental variables. This approach also allows to use a spatial resolution higher than that of the physical input, thanks to the independent modelling of each individual. It therefore allows to represent highly aggregated populations. However, the main drawback of Individual-Based Models (IBM) is that they quickly become computing-intensive. In addition, it is hard with these models to represent the full population cycle.

- The Eulerian approach computes fluxes of fish density using advection-diffusion-reaction equations. This description consist in studying at a fixed point the modifications of the properties of the fish density at this point. This approach allows to easily calculate the spatial variability of the fish as a function of time. The main drawback is that individual properties of fishes cannot be taken into account. Indeed, with this method, we consider that inside a given cohort, all fishes are identical and have the same behaviour. Although this approach allows for less details, it has the advantage of being more integrative, and therefore requires less parameters. Another advantage of this method is that the implementation of data assimilation method is easier since it is possible to use continuous equations. Furthermore, the modelling of full cycle from eggs to adults is more straigtforward.

Spatial models developed to investigate the relationships of small pelagic species with their (simulated) environment are mostly based on individual-based model (i.e., Lagrangian approach), either focusing on larval stages (e.g., Mullon et al., 2002; Lett et al., 2007; Brochier et al., 2008) or growth (Megrey et al., 2007; Rose et al., 2007), or trying to describe the full cycle of the population (Ito et al., 2007; Okunishi et al., 2009; Xi, 2009) or even multi-species interactions (Travers et al., 2009) after reaching the number of elements by using meta-individuals.

Comparatively, attempts to use the Eulerian approach for computing fluxes of fish density with advection-diffusion-reaction (ADR) equations are scarce (MacCall, 1990; Magnússon et al., 2004), despite that they can offer an advantageous framework for spatial population dynamics modeling and stock assessment studies due to limited number of parameters and continuous functions allowing the development of inverse models for optimization (Senina et al., 2008). They are a tool to model the full population cycle from eggs to adults.

One such Eulerian modeling approach of population dynamics has been successfully developed for large pelagic species (i.e. tuna and tuna-like species) at basin-scale (Lehodey et al., 2008, 2010). The main features of this Spatial Ecosystem And POpulation

DYnamics Model (SEAPODYM) are: (i) forcing by environmental data (temperature, currents, primary production and dissolved oxygen concentration), (ii) prediction of both temporal and spatial distribution of mid-trophic (micronekton) functional groups (Lehodey et al., 2010), (iii) prediction of both temporal and spatial distribution of age-structured predator populations, (iv) prediction of total catch and size frequency by fleet when fishing effort is available, and (v) parameter optimization based on fishing data assimilation techniques (Senina et al., 2008). Furthermore, since anchoveta and sardine seems to be controlled by a bottom-up forcing (Chavez et al., 2008; Bertrand et al., 2008; Ayón et al., 2008), a model driven by environmental forcing should be able to explain variability of these populations.

Although the lagrangian approach is more likely to be able to represent the high patchiness of observed data, we choose an eulerian approach implemented through the SEAPODYM model, since it allows the optimization of model parameters, as well as the representation of the full population cycle.

The general scheme of the SEAPODYM model for anchovy and sardine is presented in Figure 1.23.

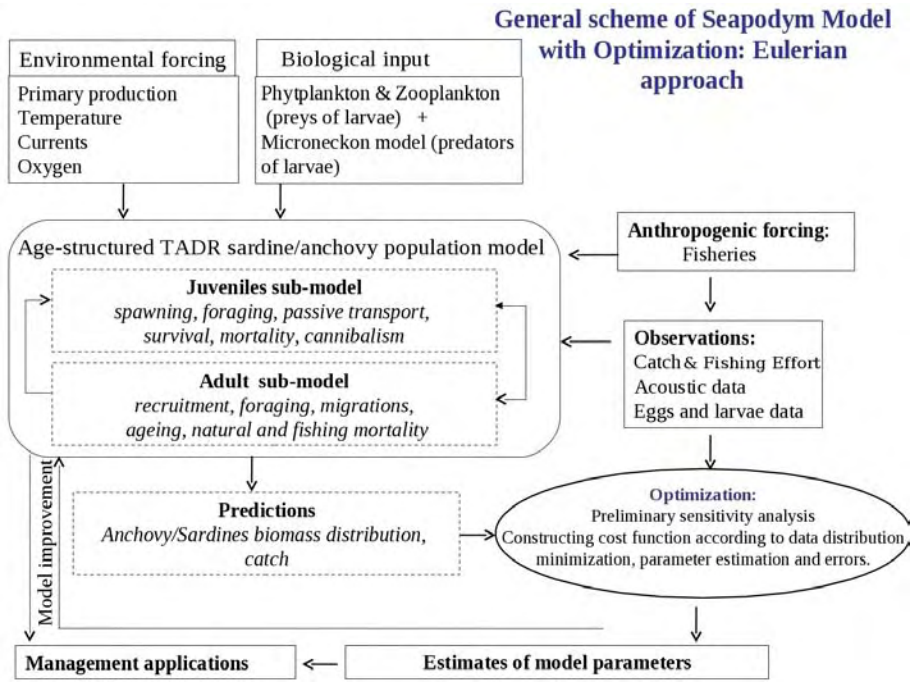

Figure 1.23: General scheme of the model with optimization approach, modified from Senina et al. (2008)

More details of the model are available in Lehodey et al. (2008) and Senina et al. (2008) for tuna application. In brief, the spatial dynamics of fish is based on advection-diffusion-reaction equation simulating random and oriented movements. Currents passively trans-

port larvae and juveniles, and young and adult fish movements are driven by habitat indices. Advection and diffusion rate are proportional to the size of the fish, and the advection rate is proportional to the gradient of the habitat, while the diffusion rate decreases when the habitat index increases.

1.2.5 Thesis aims and objectives

The main scientific objective of this PhD thesis is to develop the modeling framework allowing to test the mechanisms leading to anchovy or sardines abundances, with a model including spatial dynamics of both species under the constraints of their environment. The aims are twofolds: first, to adapt the SEAPODYM model to a regional coastal domain for small pelagic fish; second, to investigate whether the environmental variability and the different internal population dynamics (and habitat selection) can explain periods of high abundances of anchovy or sardine.

From the literature (Bertrand et al., 2004; Gutiérrez et al., 2007; Bertrand et al., 2008; Swartzman et al., 2008), we know that anchovy distribution is strongly linked to cold coastal water (CCW) and mixed water (MCW), independently of anchovy's biomass. The anchovy population seems to follow the habitat-based hypothesis, i.e in which the fish population is determined by the quality of its habitat, and the changes of this habitat determine variations of the stock size and its geographical extension. However, this hypothesis do not account for why sardines were restricted to nearshore waters and collapsed in the late 1990s (Bertrand et al., 2011). To explain sardine habitat, oxygen has been proposed to be the missing link (Bertrand et al., 2011) and should be taken into account for habitat definition.

Therefore, the SEAPODYM model that combines spatial population dynamics with definition of species habitats should be suitable to simulate these fishes population dynamics.

However, there are several challenging issues to adapt it to small pelagics.

1. The model domain for tuna applications is at basin-scale and can run at relatively low resolution (typically 1 degree x month) for parameter optimization. Conversely, anchovy and sardine stocks in the HCS are coastal, and the main factors driving the variability in eggs and larvae abundances (enrichment, concentration and retention), as identified by Bakun (1996), are likely strongly influenced by mesoscale activity. Thus, the **approach requires a regional model with higher resolution and accurate representation of mesoscale physical and biogeochemical patterns**.

2. The type of available observation that can be used in the parameter estimation approach is different. For tuna fisheries, there is a long historical (50 years) fishing period with relatively detailed (spatially-disaggregated) records of catch, effort and

size frequency of catch over all the oceanic basins. For anchovies and sardines, historical catch data is often based on port landings (i.e., no detailed spatial information), and the fishing activity can be huge but on a short period of time, thus providing no information over the rest of the year (Fréon et al., 2008). However, critical additional information has been collected by the Peruvian marine research institute IMARPE (Instituto del Mar del Peru, http://www.imarpe.pe). Eggs, larvae and fish acoustic biomass have been recorded during regular scientific cruises along the coast of Peru since 1961 for eggs and larvae, and 1983 for acoustical monitoring. This data will be used for data assimilation approach. These modifications will be described in chapters 6 and 7.

3. The modeling of spatial dynamics of a fish species in SEAPODYM requires the prey fields and the biomass of mid-trophic level groups (ie micronekton) that are the main predators of fish larvae. The prey fields for tuna is the micronekton whereas in the case of small pelagics like sardine and anchovy, forage is a group containing phytoplankton and zooplankton. For tuna, preys and larvae's predators model are the same, whereas for anchovy and sardine, preys and larvae's predators model need to be differentiate and adapted to the specificity of these species. These modifications will be described in chapters 4 and 5.

4. Between tuna and anchovy, mechanisms defining spawning habitat can be different and thus need revision. These will be described in details in chapter 2.

Therefore, given that spawning and larvae recruitment mechanisms largely determine the dynamics of the population (Beverton and Holt, 1957), and since most of the collected information is relevant for early life stages, this thesis focuses mainly on the adaptation of the SEAPODYM model to the spawning habitat and larvae dynamics of the anchovy and sardine, following therefore the general scheme of the model for early life stages in Figure 1.24.

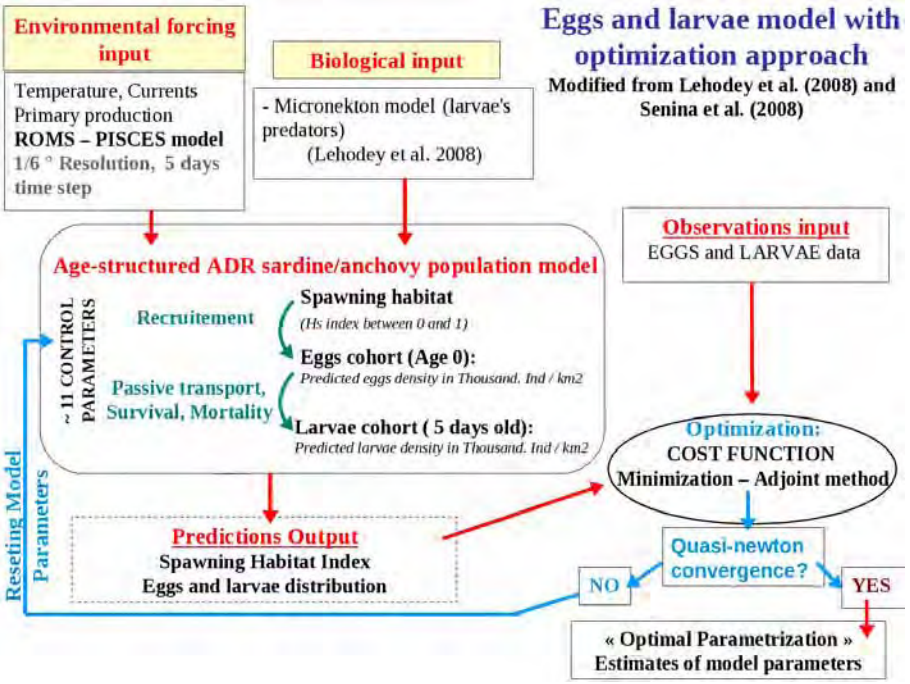

Figure 1.24: General scheme of the model with optimization approach.

1.2.6 Structure of the thesis

This study is organized into 7 chapters following the SEAPODYM model organization for early life stages described in Figure 1.12.

- Chapter 2 describes the fundamentals of the new SEAPODYM version adapted to early life stages of small pelagic fishes. The different hypothesis and equations of this new model version are described as well as the differences from the initial tuna SEAPODYM version.

- Chapter 3 focuses on **observations input** and describes the eggs, larvae and adults datasets of anchovy and sardines from the Institut del Mar del Peru (IMARPE) which will be used to validate the SEAPODYM model and to perform the optimization approach. This chapter presents the main patterns that we will try to reproduce with our modelling approach as well as describes the sources of uncertainties in the data, and the part of the data used on the optimization process.

- In chapter 4, the **environmental forcing input** is described. To drive the anchovy and sardine with SEAPODYM, we use temperature, currents and primary-production fields from different ROMS-PISCES models: a climatological run with 1/6° of resolution, a interannual run from 1992 to 2008 at 1/6° of resolution, an interannual run from 2000 to 2006 at 1/9° of resolution. Ocean physical models greatly improved during the last decade, but still can have biases at high resolution.

As our model depends on these inputs fields, the interpretation of SEAPODYM results, is linked to the analysis of the differents physical and biological fields used to drive the model. This includes micronekton model.

- In chapter 5, I present the first attempt to apply SEAPODYM for early life stages. This chapter is under the form of an article to be submitted soon. It shows that the combination of a few simple mechanisms with a limited number of parameters quickly leads to infinity of solutions that only rigorous parameter optimization approaches can help to explore.

- Thus, in chapter 6, I review briefly the data assimilation methods used in marine sciences and present the main concepts needed to implement assimilation of eggs and larvae data in SEAPODYM. We focus mainly on the variational methods used to find the optimal solution of the data assimilation problem. As detailed in this chapter, different optimal estimators can be used to define optimality of a solution. Here I present two optimal estimators: the "Maximum A Posteriori estimation (MAP)" and the "Maximum Likelihood Estimator (MLE)". The first estimator is illustrated with an application to track blue whales calls with ocean-bottom seismometers and an hydrophone array, a work that has been published at the beginning of my PhD (see Annexe C). The second estimator is the one used in SEAPODYM.

- Chapter 7 presents the results of optimization experiments for early life stages parameters using a climatological run presented in chapter 4. A sensitivity study of the main sources of variability and uncertainty of our model allows to determine the main mechanisms which strongly affect the results of our model. Based on this study, I propose different options to possibly improve the results of the model. I will also present preliminary results of optimisation experiments based on two interanual runs presented in chapter 4.

- The general conclusion summarizes the main results of this thesis. We outline the limits of the modelisation approach, and propose some perspectives of future developments.

Bibliography

Alheit, J. and M. Niquen, 2004: Regime shifts in the humboldt current ecosystem. *Progress in oceanography*, **60**(2-4), 201–222.

Ayón, P., G. Swartzman, A. Bertrand, M. Gutiérrez, and S. Bertrand, 2008: Zooplankton and forage fish species off Peru: Large-scale bottom-up forcing and local-scale depletion. *Progress in Oceanography*, **79**(2-4), 208–214.

Ayón, P., G. Swartzman, P. Espinoza, and A. Bertrand, 2011: Long-term changes in zooplankton size distribution in the Peruvian Humboldt Current System: conditions favouring sardine or anchovy. *Mar. Ecol. Prog. Ser*, **422**, 211–222.

Bakun, A., 1996: Patterns in the Ocean: Ocean Processes and Marine Population Dynamics. *University of California Sea Grant, San Diego, California, USA, in cooperation with Centro de Investigaciones Biológicas de Noroeste, La Paz, Baja California Sur, Mexico*.

Bakun, A. and K. Broad, 2003: Environmental 'loopholes' and fish population dynamics: comparative pattern recognition with focus on El Niño effects in the Pacific. *Fisheries Oceanography*, **12**(4-5), 458–473.

Bakun, A. and S. J. Weeks, 2008: The marine ecosystem off Peru: What are the secrets of its fishery productivity and what might its future hold ? . *Progress in Oceanography*, **79**(2-4), 290–299.

Ballón, M., A. Bertrand, A. Lebourges-Dhaussy, M. Gutiérrez, P. Ayón, D. Grados, and F. Gerlotto, 2011: Is there enough zooplankton to feed forage fish populations off Peru? An acoustic (positive) answer. *Progress In Oceanography*, **91**(4), 360–381.

Behrenfeld, M. J. and P. G. Falkowski, 1997: Photosynthetic rates derived from satellite-based chlorophyll concentration. *Limnology and oceanography* 1–20.

Bertrand, A., A. Chaigneau, S. Peraltilla, J. Ledesma, M. Graco, F. Monetti, and F. Chavez, 2011: Oxygen, a fundamental property regulating pelagic ecosystem structure in the coastal Southeastern tropical Pacific. *PlosOne*, **6**(12), e29558.

Bertrand, A., P. Fréon, A. Chaigneau, V. Echevin, C. Estrella, H. Demarcq, D. Gutiérrez, and J. C. Sueiro, 2010: Climate change impact on ocean dynamics, ecosytem functioning and fisheries of peru: projection scenarios and socioeconomic impacts. *Report for the British Embassy in Peru, IRD, IMARPE, 42 pp. http://ukinperu.fco.gov.uk/resources/es/pdf/5933461/cambio-climatico-pesqueras*.

Bertrand, A., F. Gerlotto, S. Bertrand, M. Gutiérrez, L. Alza, A. Chipollini, E. Diaz, P. Espinoza, J. Ledesma, and R. Quesquon, 2008: Schooling behaviour and environmental forcing in relation to anchoveta distribution: An analysis across multiple spatial scales. *Progress in Oceanography*, **79**(2-4), 264–277.

Bertrand, A., M. Segura, M. Gutiérrez, and L. Vasquez, 2004: From small-scale habitat loopholes to decadal cycles: a habitat-based hypothesis explaining fluctuation in pelagic fish populations off Peru. *Fish and Fisheries*, **5**(4), 296–316.

Beverton, R. and S. Holt, 1957: On the dynamics of exploited fish populations. *Fish. Invest. Ser. II, XIX HMSO, London* 1–533.

Brochier, T., C. Lett, and P. Fréon, 2011: Investigating the "northern humboldt paradox" from model comparisons of small pelagic fish reproductive strategies in eastern boundary upwelling ecosystems. *Fish and Fisheries*, **12**(1), 94–109.

Brochier, T., C. Lett, J. Tam, P. Fréon, F. Colas, and P. Ayón, 2008: An individual-based model study of anchovy early life history in the northern Humboldt Current system. *Progress in Oceanography*, **79**(2-4), 313–325.

Chavez, F. P., A. Bertrand, R. Guevara-Carrasco, P. Soler, and J. Csirke, 2008: The northern Humboldt Current System: Brief history, present status and a view towards the future. *Progress in Oceanography*, **79**(2-4), 95–105.

Chavez, F. P. and M. Messié, 2009: A comparison of Eastern Boundary Upwelling Ecosystems. *Progress in Oceanography*, **83**, 80–96.

Checkley Jr, D., P. Ayón, T. Baumgartner, M. Bernal, J. Coetzee, R. Emmett, R. Guevara-Carrasco, L. Hutching, L. Ibaibarriaga, H. Nakata, et al., 2009: *Habitats. In: D.M. Checkley Jr., J. Alheit, Y. Oozeki and C. Roy, Editors, Climate Change and Small Pelagic Fish, Cambridge University Press, Cambridge, pp. 12-44 (Chapter 3).*

Cury, P., A. Bakun, R. Crawford, A. Jarre, R. Quiñones, L. Shannon, and H. Verheye, 2000: Small pelagics in upwelling systems: patterns of interaction and structural changes in "wasp-waist" ecosystems. *ICES Journal of Marine Science*, **57**(3), 603–618.

Cury, P. and L. Shannon, 2004: Regime shifts in upwelling ecosystems: observed changes and possible mechanisms in the northern and southern benguela. *Progress in Oceanography*, **60**, 223–243.

Durand, M., Climate, and E. O. S. Project, 1998: *Global versus local changes in upwelling systems*. ORSTOM.

Durand, N. S. and G. Seminario, 2009: Status of and trends in the use of small pelagic fish species for reduction fisheries and for human consumption in Peru. In M.R. Hasan and M. Halwart (eds). Fish as feed inputs for aquaculture: practices, sustainability and implications. Technical report, FAO Fisheries and Aquaculture Technical Paper.

Echevin, V., O. Aumont, J. Ledesma, and G. Flores, 2008: The seasonal cycle of surface chlorophyll in the Peruvian upwelling system: A modelling study. *Progress in Oceanography*, **79**(2-4), 167–176.

Espinoza, P. and A. Bertrand, 2008: Revisiting Peruvian anchovy (*Engraulis ringens*) trophodynamics provides a new vision of the Humboldt Current system. *Progress in Oceanography*, **79**(2-4), 215–227.

Espinoza, P., A. Bertrand, C. D. van der Lingen, S. Garrido, and B. R. de Mendiola, 2009: Diet of sardine (*Sardinops sagax*) in the northern Humboldt Current system and comparison with the diets of clupeoids in this and other eastern boundary upwelling systems. *Progress In Oceanography*, **83**, 242–250.

FAO, 2010: The state of world fisheries and aquaculture. Technical report, FAO Fisheries Department, Rome, Italy.

Fréon, P., J. Aristegui, A. Bertrand, R. Crawford, J. Field, M. Gibbons, J. Tam, L. Hutchings, H. Masski, C. Mullon, et al., 2009: Functional group biodiversity in Eastern Boundary Upwelling Ecosystems questions the wasp-waist trophic structure. *Progress in Oceanography*, **83**(1-4), 97–106.

Fréon, P., M. Bouchon, C. Mullon, C. GarcÃa, and M. Ñiquen, 2008: Interdecadal variability of anchoveta abundance and overcapacity of the fishery in Peru. *Progress in Oceanography*, **79**(2-4), 401–412.

Gutiérrez, D., A. Sifeddine, D. B. Field, L. Ortlieb, G. Vargas, F. P. ChÃ¡vez, F. Velazco, V. Ferreira, P. Tapia, R. Salvatteci, et al., 2009: Rapid reorganization in ocean biogeochemistry off peru towards the end of the little ice age. *Biogeosciences*, **6**(5), 835–848.

Gutiérrez, M., G. Swartzman, A. Bertrand, and S. Bertrand, 2007: Anchovy (*Engraulis ringens*) and sardine (*Sardinops sagax*) spatial dynamics and aggregation patterns in the Humboldt Current ecosystem, Peru, from 1983-2003. *Fisheries Oceanography*, **16**(2), 155–168.

IFFO, 2009: (International Fishmeal and Fish Oil Organisation), The production of fishmeal and fish oil from Peruvian anchovy. Technical report.

Ito, S., B. Megrey, M. Kishi, D. Mukai, Y. Kurita, Y. Ueno, and Y. Yamanaka, 2007: On the interannual variability of the growth of Pacific saury (*Cololabis saira*): a simple 3-box model using NEMURO. FISH . *Ecological modelling*, **202**(1-2), 174–183.

Lehodey, P., R. Murtugudde, and I. Senina, 2010: Bridging the gap from ocean models to population dynamics of large marine predators: A model of mid-trophic functional groups. *Progress In Oceanography*, **86**(1-2), 302–315.

Lehodey, P., I. Senina, and R. Murtugudde, 2008: A spatial ecosystem and populations dynamics model (SEAPODYM) - Modeling of tuna and tuna-like populations. *Progress in Oceanography*, **78**(4), 304–318.

Lehodey, P., I. Senina, J. Sibert, L. Bopp, B. Calmettes, J. Hampton, and R. Murtugudde, 2010: Preliminary forecasts of Pacific bigeye tuna population trends under the A2 IPCC scenario. *Progress In Oceanography*, **86**(1-2), 302–315.

Lett, C., P. Penven, P. Ayón, and P. Fron, 2007: Enrichment, concentration and retention processes in relation to anchovy (*Engraulis ringens*) eggs and larvae distributions in the northern Humboldt upwelling ecosystem. *Journal of Marine Systems*, **64**(1-4), 189–200.

Lluch-belda, D., R. Schwartzlose, R. Serra, R. Parrish, T. Kawasaki, D. Hedgecock, and R. Crawford, 1992: Sardine and anchovy regime fluctuations of abundance in four regions of the world oceans: a workshop report. *Fisheries Oceanography*, **1**(4), 339–347.

MacCall, A., 1990: *Dynamic geography of marine fish populations*. Washington Sea Grant Program Seattle, WA.

MacKenzie, B., H. Mosegaard, and A. Rosenberg, 2009: Impending collapse of bluefin tuna in the northeast Atlantic and Mediterranean. *Conservation Letters*, **2**(1), 26–35.

Magnússon, K., S. Sigurdsson, P. Babak, S. Gudmundsson, and E. Dereksdottir, 2004: A continuous density Kolmogorov type model for a migrating fish stock. *Discrete and Continuous Dynamical Systems Series B*, **4**, 695–704.

McPhaden, M. J., S. E. Zebiak, and M. H. Glantz, 2006: ENSO as an Integrating Concept in Earth Science. *Science*, **314**(5806), 1740–1745.

Megrey, B., K. Rose, R. Klumb, D. Hay, F. Werner, D. Eslinger, and S. Smith, 2007: A bioenergetics-based population dynamics model of pacific herring (*clupea harengus pallasi*) coupled to a lower trophic level nutrient-phytoplankton-zooplankton model: description, calibration, and sensitivity analysis. *Ecological Modelling*, **202**(1-2), 144–164.

Mullon, C., P. Cury, and P. Penven, 2002: Evolutionary individual-based model for the recruitment of anchovy (*engraulis capensis*) in the southern benguela. *Canadian Journal of Fisheries and Aquatic Sciences*, **59**(5), 910–922.

Myers, R., J. Hutchings, and N. Barrowman, 1997: Why do fish stocks collapse? The example of cod in Atlantic Canada. *Ecological Applications*, **7**(1), 91–106.

Okunishi, T., Y. Yamanaka, and S. Ito, 2009: A simulation model for japanese sardine (*sardinops melanostictus*) migrations in the western north pacific. *Ecological Modelling*, **220**(4), 462–479.

Pauly, D., J. Alder, E. Bennett, V. Christensen, P. Tyedmers, and R. Watson, 2003: The future for fisheries. *Science*, **302**(5649), 1359–1361.

Pauly, D., V. Christensen, S. Guénette, T. J. Pitcher, U. R. Sumaila, C. J. Walters, R. Watson, and D. Zeller, 2002: Towards sustainability in world fisheries. *Nature*, **418**(6898), 689–695.

Penven, P., V. Echevin, J. Pasapera, F. Colas, and J. Tam, 2005: Average circulation, seasonal cycle, and mesoscale dynamics of the 3 Peru Current System: A modeling approach. *Journal of Geophysical Research*, **110**.

Rose, K., F. Werner, B. Megrey, M. Aita, Y. Yamanaka, D. Hay, J. Schweigert, and M. Foster, 2007: Simulated herring growth responses in the Northeastern Pacific to historic temperature and zooplankton conditions generated by the 3-dimensional NEMURO nutrient-phytoplankton-zooplankton model. *Ecological modelling*, **202**(1-2), 184–195.

Rothschild, B., 1996: How Bountiful are Ocean Fisheries? Consequences 2: 1 [Online]. Available HTTP: http://www.gcrio.org/CONSEQUENCES/winter96/oceanfish.html.

Senina, I., J. Sibert, and P. Lehodey, 2008: Parameter estimation for basin-scale ecosystem-linked population models of large pelagic predators: Application to skipjack tuna. *Progress in Oceanography*, **78**(4), 319–335.

Stramma, L., G. C. Johnson, J. Sprintall, and V. Mohrholz, 2008: Expanding oxygen-minimum zones in the tropical oceans. *Science*, **320**(5876), 655–658.

Swartzman, G., A. Bertrand, M. Gutiérrez, S. Bertrand, and L. Vasquez, 2008: The relationship of anchovy and sardine to water masses in the Peruvian Humboldt Current System from 1983 to 2005. *Progress in Oceanography*, **79**(2-4), 228–237.

Thiel, M., E. C. Macaya, E. Acuna, W. E. Arnitz, H. Bastias, K.Brokordt, P. A. Camus, J. C. Castilla, L. R. Castro, M.Cortes, et al., 2007: The Humboldt Current System of northern and central Chile: oceanographic processes, ecological interactions and socioeconomic feedback. *Oceanography and Marine Biology*, **45**, 195–344.

Travers, M., Y. Shin, S. Jennings, E. Machu, J. Huggett, J. Field, and P. Cury, 2009: Two-way coupling versus one-way forcing of plankton and fish models to predict ecosystem changes in the Benguela. *Ecological Modelling*, **220**(21), 3089–3099.

Tveteras, S., C. Paredes, and J. Peña-Torres, 2011: Individual vessel quotas in peru: Stopping the race for anchovies. *Marine Resource Economics, Forthcoming*, **26**(3), 225–232.

Valdés, J., L. Ortlieb, D. Gutiérrez, L. Marinovic, G. Vargas, and A. Sifeddine, 2008: 250 years of sardine and anchovy scale deposition record in mejillones bay, northern chile. *Progress in Oceanography*, **79**(2-4), 198–207.

Xi, Y., 2009: *Modeling the seasonal and interannual variability of Peruvian anchovy (Engraulis ringens) population dynamics: linking environmental conditions with fish.* PhD thesis, The University of Maine.

Chapter 2

Fundamentals of the small pelagic fish SEAPODYM model version

Sommaire

2.1 Introduction **46**

2.2 Main processes controlling spawning habitat and eggs and larvae dynamics **46**

2.3 Definition of spawning habitat and larvae recruitement in the SEAPODYM standard version **47**

2.4 Revision from SEAPODYM standard version for small pelagics **47**

2.4.1 Spawning habitat definition 48

2.4.1.1 Impact of temperature 49

2.4.1.2 Effect of a relationship to prey abundance 49

2.4.1.3 Effect of a relationship to predator abundance 50

2.4.1.4 Accessibility of adults to spawning areas 50

2.4.1.5 Illustration of spawning habitat definition 50

2.4.2 Predicted eggs density 51

2.4.3 Predicted larvae density 52

2.5 Conclusion **53**

Bibliography **54**

2.1 Introduction

Sea surface temperature, currents and primary production fields of the bio-physical model are used to drive the spawning habitat and larvae movement in SEAPODYM (Fig. 1.24). This chapter briefly reviews the main processes proposed in the literature to control spawning habitat and eggs and larvae dynamics. Then the mechanisms used in the SEAPODYM standard version and the changes proposed to adapt SEAPODYM to small pelagic fishes are presented. This SEAPODYM version will be termed therafter the "SEAPODYM small pelagics version".

2.2 Main processes controlling spawning habitat and eggs and larvae dynamics

In the abundant literature devoted to the problem of fish recruitment (e.g., see reviews in Rothschild, 2000 and Govoni, 2005), the main processes that are proposed include the effect of temperature, the temporal and spatial availability of food for the larvae, and the predation of eggs and larvae. Finally, it is also proposed that the redistribution of larvae by the oceanic circulation can create retention of larvae in favorable areas (with lower natural mortality) or conversely move the larvae to unfavorable zones where the natural mortality will be higher (Parrish et al., 1981; Bakun, 1996).

Of course, recruitment is also linked to the spawning stock (Beverton and Holt, 1957), but for small pelagic fishes in upwelling regions, the variability in spawning success and larvae survival is thought to be largely driven by environmental conditions (Cury and Roy, 1989; Cury et al., 1995; McFarlane et al., 2002; Brochier et al., 2008). Therefore, in this thesis, the stock-recruitment relationship was not included and we simply assume that mature adults are present everywhere. Although the adult stock is not modelled, it is important to consider the accessibility of adults to spawning areas. Here, this process is introduced only by taking into account the possible effect of dissolved oxgen concentration on accessibility, since oxygene concentration that seems to play a role on anchovy and sardine dynamics (Bertrand et al., 2011).

Different mechanisms have been included in SEAPODYM to simulate the spawning habitat (Hs) and larvae dynamics through functional relationships. The result is a relative index normalized between 0 and 1, thus independent of absolute calibration of primary production or energy transfer to micronekton groups. These mechanisms have been revised from Lehodey et al. (2008) for small pelagic species.

2.3 Definition of spawning habitat and larvae recruitement in the SEAPODYM standard version

In SEAPODYM standard version, spawning habitat is given by the following equation:

$$H_s = f_1(T) * f(PP, PRED) = f_1(T) * \frac{\Lambda}{\Lambda + \alpha} \tag{2.1}$$

with:

$f_1(T)$, the spawning temperature index in the epipelagic layer, defined by a gaussian function with two parameter to be estimated from the data.

$\Lambda = (PP/Pred + 10^{-5})$ is the ratio between food abundance of larvae (micro-zooplankton approximated by primary production) and larvae predator density (i.e., the sum of biomass of mid-trophic groups present in the epipelagic layer during day-time and sunrise and sunset periods).

α is the curvature parameter which needs to be estimated from the data. If $\alpha = 0$, only temperature has an effect on the spawning, while the trade-off effect between densities of food and predators increases relatively to the temperature effect with increasing α.

"The number of larvae recruited in each cell of the grid at a given time is the product between H_s and a number R_S, with the primary condition that adult fish, i.e., potentially mature, are present in this cell. R_S can be fixed or linked to the adult spawning biomass, e.g., with a Beverton-Holt relationship. After spawning, currents in the surface layer redistribute larvae, and a natural mortality coefficient is applied before entering in the juvenile cohorts" (Lehodey et al., 2008).

2.4 Revision from SEAPODYM standard version for small pelagics

To model eggs and larvae of small pelagic fish we have revised spawning habitat definition from Lehodey et al. (2008). As noted above, the spawning temperature index ($f_1(T)$) is a relative index, but the ratio between food abundance of larvae and predator density, $f(PP, PRED)$, highly depends on the estimated values of primary production and predator density. In this relationship, the effet of low predator density is more important than the effect of high food abundance. To avoid this biais between food abundance and predator density, as well as to test the impact of these two mechanisms separately, we modified this definition of spawning habitat by computing two different relative indices, one for food abundance ($f_2(PP)$) and another for predator abundance

($f_3(Pred)$): $f(PP, PRED) = f_2(PP) * f_3(Pred)$. A third relative index is used to consider the accessibility of adults to spawning areas, based on the concentration of dissolved oxgen ($f_4(O_2)$).

The standard version of SEAPODYM does not model eggs density. In this version, we added an eggscohort in order to compare it with observations of eggs density.

A summary of the underlying hypotheses and equations of the SEAPODYM small pelagics version is provided in the following sections.

2.4.1 Spawning habitat definition

In its more complex form, the spawning habitat index (H_S) is the product of a temperature function ($f_1(T)$, Eq. 2.3), a prey function ($f_2(PP)$, Eq. 2.4), a predator function ($f_3(Pred)$, Eq. 2.5) and a function of the accessibility of adults to spawning areas, represented by a relationship to the oxygen concentration ($f_4(O_2)$, Eq. 2.6):

$$H_S = f_1(T) * f_2(PP) * f_3(Pred) * f_4(O_2) \tag{2.2}$$

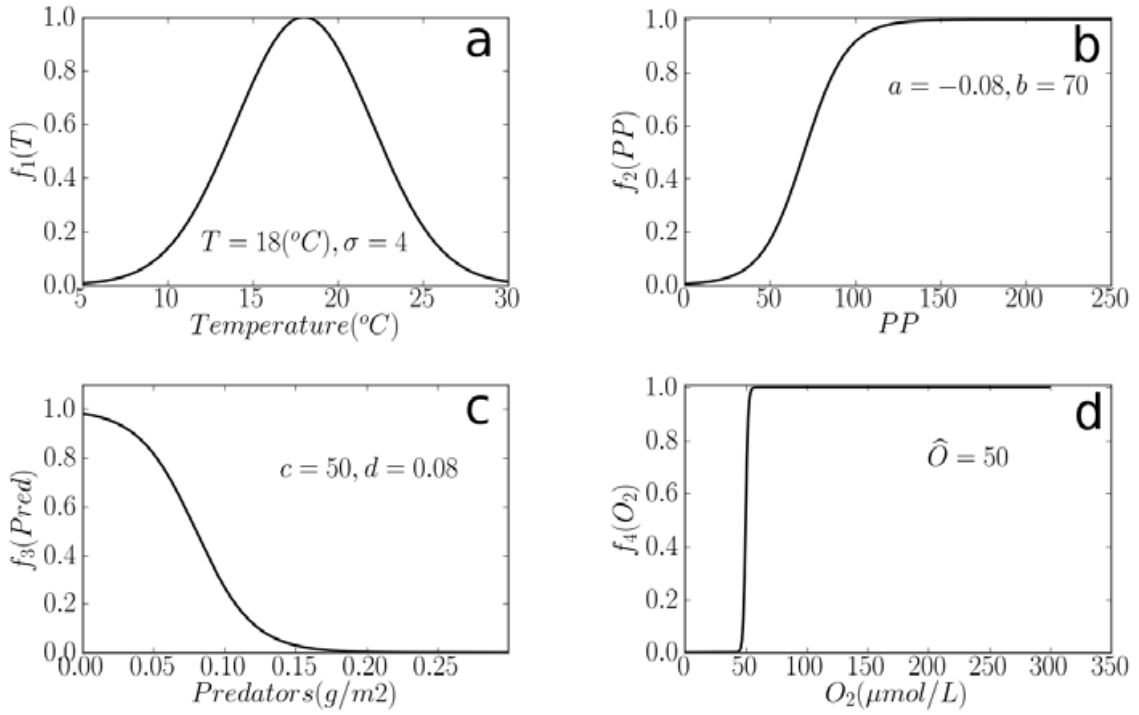

Figure 2.1: Functional relationships used to define the spawning habitat: functions of temperature (a), prey availability (b), predation (c) and oxygen (d).

2.4.1.1 Impact of temperature

The thermal spawning habitat that controls changes in the spatial extent of the favorable temperature for larval growth was estimated using a Gaussian distribution (Eq. 2.3):

$$f_1(T) = N(T_0^*, \sigma_0) \tag{2.3}$$

with T_0^* the optimal mean temperature and σ_0 the width of tolerance interval (standard deviation) for eggs and larvae stage (Fig. 2.1a). We consider that eggs and larvae temperature tolerance are identical.

Sea surface temperature is used to define the thermal spawning habitat of eggs and larvae of anchovy and sardine.

2.4.1.2 Effect of a relationship to prey abundance

We consider here the spatial and temporal availability of the density of suitable preys. Larvae certainly need a minimum of accessible food to survive, but the relationship between survival and food density is likely not linear due to a saturation effect at high levels of prey density. These mechanisms were described by Holling (1959) and more particularly in the so-called Holling's type III functional response. In this study we used a sligtly different function. We use a sigmoid function defined with two parameters (a and b) (Eq. 2.4, Fig. 2.1b) allowing more flexibility to define the minimum of accessible food to larvae survival. When there is little food, the food spawning habitat index is null. It increases with food availability, until a level where the quantity of food is enough for all larvae.

$$f_2(PP) = \frac{1}{1 + \exp^{a*(PP-b)}} \tag{2.4}$$

Since the distribution of the small zooplankton is closely linked to the phytoplankton (within the 5 day time step of the simulation), primary production is used as a proxy of larvae food.

We could also have used the functional groups of micro-zooplankton for larvae's food given by the PISCES biogeochemical model. Indeed, the PISCES model has two zooplankton size classes (microzooplankton and mesozooplankton) (see section 4.3.1), but the predictions are not necessary realistic since they correspond to the closure term of the NPZD cycle (see section 4.3.1 for more details of the NPZD model). Furthermore these outputs are difficult to validate with data. We consider that the approximation of using primary production as a proxy of larvae food is more accurate than the direct use of microzooplankton output from PISCES.

2.4.1.3 Effect of a relationship to predator abundance

To take into account the predators of eggs and larvae, we use a decreasing sigmoid function (Eq. 2.5, Fig. 2.1c) linked to the biomass of eggs and larvae predators (described in details in section 4.4):

$$f_3(Pred) = \frac{1}{1 + \exp^{c*(Pred-d)}} \tag{2.5}$$

When the predators density is low, predators habitat index is high. It decreases with predators density until a threshold where predators are too abundant and the favorability of the habitat quickly drop to zero.

We define the prey-predator trade-off as the product of the prey function (Eq. 2.4) and the predator function (Eq. 2.5).

2.4.1.4 Accessibility of adults to spawning areas

Bertrand et al. (2011) noted the importance of oxygen concentration in sardine's habitat (see Fig. 1.22). Despite the fact that oxygen concentration predicted by the ROMS-PISCES model still shows disagreements with observations, we wish to test if our spawning habitat definition is sensitive to this variable. As in Lehodey et al. (2008) for the definition of feeding habitat, the effect of oxygen concentration in the spawning habitat was modeled by a sigmoid function (Eq. 2.6, Fig. 2.1d).

$$f_4(O_2) = \frac{1}{1 + \exp^{e*(O_2-\widehat{O})}} \tag{2.6}$$

with $\widehat{O}$, the threshold value of dissolved oxygen in $\mu mol.L^{-1}$ and e, the slope coefficient in the oxygen function (see Fig. 2.1d) fixed to -1. Only the oxygen threshold value matters for fish accessibility to spawning area and $\widehat{O}$ is the only adjustable parameter to be estimated.

Note that this mechanism was only added at the end of this study, and it was only tested in Section 7.2.4. The impact of oxygen should thus further explored.

2.4.1.5 Illustration of spawning habitat definition

Figure 2.2 illustrates the idea of spawning habitat definition. When the habitat definition includes more mechanisms, the favorable habitat is restricted to more limited favorables areas.

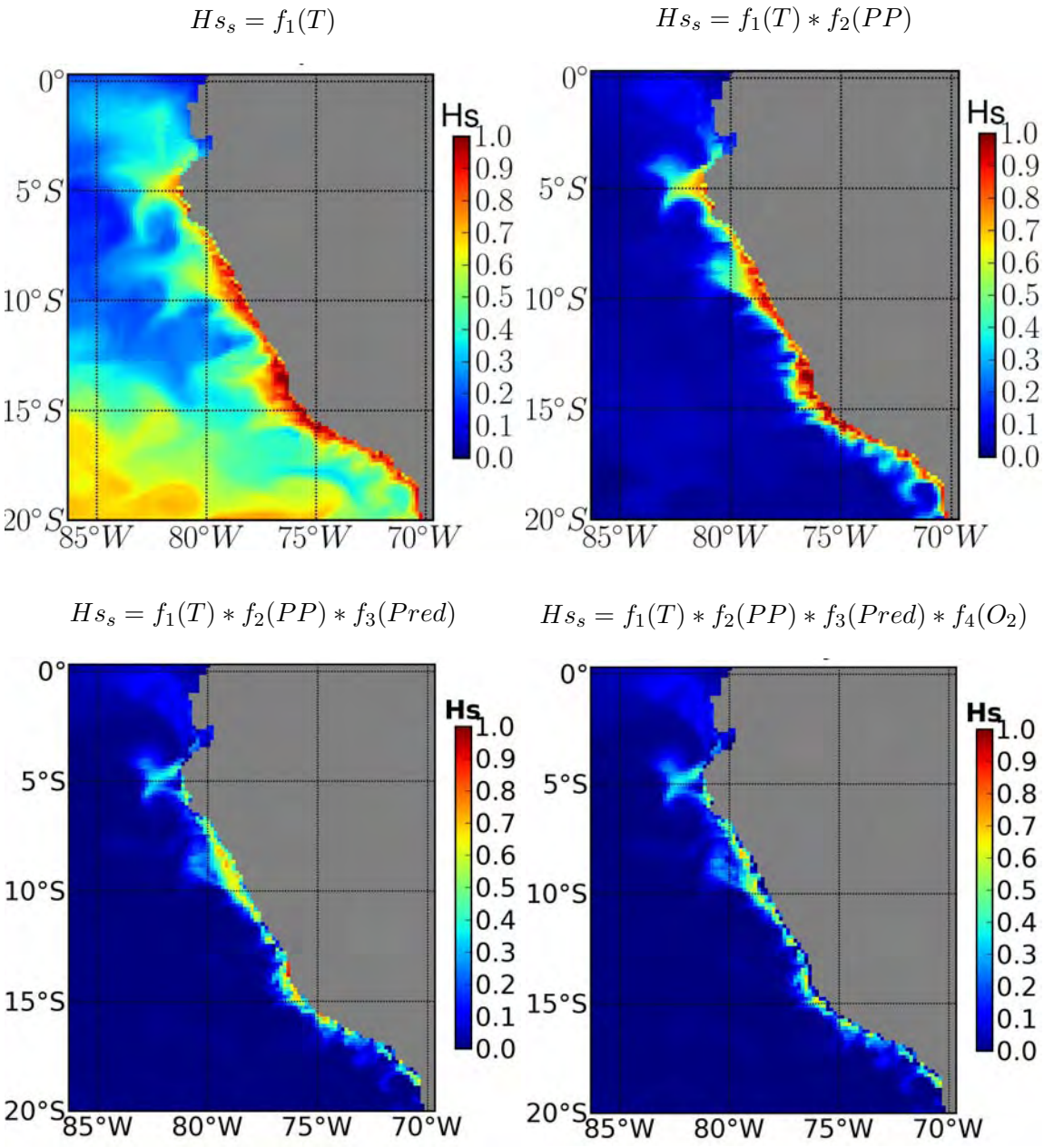

Figure 2.2: Illustration of the favorability of spawning areas when complexifyng the spawning habitat definition

2.4.2 Predicted eggs density

We define the number of eggs in each cell of the grid at a given time as the product between H_s and a number R (Eq. 2.7). Theoretically, R, is linked to the number of mature fish present in the cell. In the SEAPODYM standard model, R is linked to the adult spawning biomass with a Beverton-Holt relationship (Lehodey et al., 2008).

However, in this study, as no adults are modelled, as a first hypothesis we considered that adult biomass is the same everywhere. R is therefore considered as a constant value which needs to be estimated. It allows in other terms to fit the absolute values of eggs density observed.

$$N_0 = R * H_S \tag{2.7}$$

In the model, density of eggs is predicted in thousands of individuals per km^2.

2.4.3 Predicted larvae density

After spawning, currents in the surface layer redistribute eggs, and a natural mortality coefficient is applied before they enter in the larvae cohorts.

The redistribution of eggs and larvae by currents leading to higher or lower mortality according to the retention in favorable habitat or the drift in unfavorable habitat is included in the treatment of the spatial dynamics using a system of Advection-Diffusion-Reaction equations (ADR) (Eq. 2.8).

Let us denote N the larvae density determined at point (x,y) and time t (here after we will omit the notations of space and time). For brevity, we use divergence operator of a vector field $div(v^*) = \delta_x u + \delta_y v$ and $\Delta = divgrad$ for Laplacian of scalar field of population density. v^* denotes average of oceanic currents over the mixed layer depth, u the corresponding zonal current and v the corresponding meridional current.

The ADR equation used to describe dynamics of the anchovy larvae is:

$$dtN = -div(Nv^*) + d(\Delta N) - mN + S_0 \tag{2.8}$$

where d is a constant diffusion coefficient, fixed at $198\ m^2/s$, m the mortality rate, S_0 the source/sink (S/P) term. The larvae stage is divided in different cohorts of 5 days (time resolution of ROMS-PISCES input). At each time step surviving eggs move in the first larvae cohort, while a new recruitmeent of eggs occurs in the egg cohort. At each 5 days time step (time resolution of physical input model), surviving larvae are transferred in the following oldest cohort. In the main part of this study, larvae transport by currents with associated mortality is computed for a time step of five days only after spawning, i.e., roughly corresponding to the estimated mean age of larvae collected in a size range between 3 and 6 mm (P. Ayón pers. com.), and based on a length of hatching of 2 mm and a growth function (Marzloff et al., 2009).

The average mortality coefficient for anchovy larvae μ was set to the rate of 0.378 $month^{-1}$, i.e. 5 times the average coefficient used for the exploited adult population by Cubillos et al. (2002). However, to account for the effects of environmental variability, the average mortality-at-age varies locally with spawning habitat index (Eq. 2.9).

$$m = \mu(1+\epsilon)^{1-2H_S} \tag{2.9}$$

The implicit hypothesis here is that spawning habitat matches closely the favorable habitat of larvae. From an ecological point of view, it means that adults spawn where the habitat is favorable for eggs and larvae survival. Indeed, adult fish tend to search for environmental conditions identical to those of thir birth place (Cury, 1994).

Equation 2.8 is complemented with Neumann zero-flux boundary conditions, discretized (see Sibert et al., 1999) and solved on a 10 nmi squared cell-centered grid on a 5-days time step with help of an alternate direction implicit method.

2.5 Conclusion

To describe the spawning habitat and eggs and larvae dynamics, 10 biological parameters are needed. Although this is a rather small number of parameters, it is enough to generate a multitude of solutions. Therefore, only rigorous parameter optimization approaches can help to identify the best solution. The optimization method developed in this model is described in Chapter 6.

It should be noted that in the most of the work presented in this thesis, the accessibility of adults to spawning areas (as represented by an oxygen dependent function $f_4(O_2)$) was not taken into account. This mechanism was only tested in section 7.2.4 and should be further explored.
Therefore, unless otherwise specified, the spawning habitat index will be defined in a restrictive form which does not take into account the influence of oxygen concentration:

$$\boxed{H_S = f_1(T) * f_2(PP) * f_3(Pred)} \tag{2.10}$$

Bibliography

Bakun, A., 1996: Patterns in the ocean: Ocean processes and marine population dynamics. *University of California Sea Grant, San Diego, California, USA, in cooperation with Centro de Investigaciones Biológicas de Noroeste, La Paz, Baja California Sur, Mexico*.

Bertrand, A., A. Chaigneau, S. Peraltilla, J. Ledesma, M. Graco, F. Monetti, and F. Chavez, 2011: Oxygen, a fundamental property regulating pelagic ecosystem structure in the coastal Southeastern tropical Pacific. *PlosOne*, **6**(12), e29558.

Beverton, R. and S. Holt, 1957: On the dynamics of exploited fish populations. *Fish. Invest. Ser. II, XIX HMSO, London* 1–533.

Brochier, T., C. Lett, J. Tam, P. Fréon, F. Colas, and P. Ayon, 2008: An individual-based model study of anchovy early life history in the northern Humboldt Current system. *Progress in Oceanography*, **79**(2-4), 313–325.

Cubillos, L., D. Bucarey, and M. Canales, 2002: Monthly abundance estimation for common sardine *Strangomera bentincki* and anchovy *Engraulis ringens* in the central southern area off Chile (34-40S). *Fish Res*, **57**, 117–130.

Cury, P., 1994: Obstinate nature: an ecology of individuals: thoughts on reproductive biology and biodiversity. *Canadian Journal of Fisheries and Aquatic Sciences*, **51**, 1664–1673.

Cury, P. and C. Roy, 1989: Optimal environmental windows and fish recruitment success in upwelling areas. *Can. J. Fish. Aquat. Sci.*, **44**, 670–680.

Cury, P., C. Roy, R. Mendelssohn, A. Bakun, D. Usby, and R. Parrish, 1995: Moderate is better: exploring nonlinear climatic effects on the Californian anchovy (*Engraulis mordax*). *Can. J. Fish. Aquat. Sci.*, **121**, 417–424.

Govoni, J., 2005: Fisheries oceanography and the ecology of early life histories of fishes: a perspective over fifty years. *Sci. Mar.*, **69 (Suppl. 1)**, 125–137.

Holling, C., 1959: Some characteristics of simple type of predation and parasitism. *Can. Entomol.*, **91**, 385–398.

Lehodey, P., I. Senina, and R. Murtugudde, 2008: A spatial ecosystem and populations dynamics model (SEAPODYM) - Modeling of tuna and tuna-like populations. *Progress in Oceanography*, **78**(4), 304–318.

Marzloff, M., Y. J. Shin, J. Tam, M. Travers, and A. Bertrand, 2009: Trophic structure of the peruvian marine ecosystem in 2000-2006: Insights on the effects of management scenarios for the hake fishery using the IBM trophic model osmose. *Journal of Marine Systems*, **75**(1-2), 290–304.

McFarlane, G., P. Smith, T. Baumgartner, and J. Hunter, 2002: Climate variability and Pacific sardine population and fisheries. *Am. Fisheries Soc. Symp.*, **32**, 195–214.

Parrish, R., C. Nelson, and A. Bakun, 1981: Transport mechanisms and reproductive success of fishes in the california current. *Biolog. Oceanogr.*, **2**, 175–203.

Rothschild, B., 2000: Fish stocks and recruitment?: the past thirty years. *ICES J. Mar. Sci.*, **57**, 191–201.

Sibert, J. R., J. Hampton, D. A. Fournier, and P. J. Bills, 1999: An advection-diffusion-reaction model for the estimation of fish movement parameters from tagging data, with application to skipjack tuna (*Katsuwonus pelamis*). *Canadian Journal of Fisheries and Aquatic Sciences*, **56**(6), 925–938.

Chapter 3

Observed variability on anchovy and sardine dynamics

Sommaire

3.1 Introduction **58**
3.2 Available Data **58**
3.2.1 Eggs, larvae data 58
3.2.1.1 Spatial and temporal heterogeneity of datasets 58
3.2.1.2 Statistical characteristics 61
3.2.2 Adult data 67
3.2.2.1 Spatial and temporal heterogeneity of datasets 68
3.2.2.2 Statistical characteristics 70
3.2.3 Gonadosomatic Index 71
3.3 Spatial Distribution **72**
3.3.1 All datasets 72
3.3.2 Climatology maps 76
3.4 Seasonal variability **77**
3.4.1 Anchovy 77
3.4.2 Sardine 80
3.5 Interannual Variability: El Niño events **82**
3.6 Decadal variability **85**
3.7 Physical observations for eggs, larvae and adults **89**
3.7.1 Effect of temperature 89
3.7.2 Vertical distribution of eggs and larvae 92
3.7.3 Spatial co-occurrence of eggs and adults 94
3.8 Predators of anchovy and sardines **96**
3.9 Conclusion **98**
Bibliography **100**

3.1 Introduction

Since 1961, the Instituto del Mar del Peru (IMARPE) has been conducting regular research cruises to sample anchovy/sardine eggs and larvae (171 cruises, P. Ayón pers. com.). In 1983, IMARPE also started to monitor the adult biomass using regular acoustic sampling cruises (55 cruises, R. Castillo pers. com.). In this chapter, a synthesis of the collected data is presented from different points of view. Spatial, seasonal and interannual variability are particularly detailed to allow a synthesis of the knowledge of anchovies and sardines stocks from these datasets. This first chapter presents the observations used for the optimization process (Chapter 6), describes the sources of uncertainties of the data, and finally, concludes on the main patterns that we will try to reproduce with our modelling approach.

3.2 Available Data

3.2.1 Eggs, larvae data

Eggs and larvae data were collected from two different nets:

- The Hensen net which is characterized by 0.33 m^2 mouth area and 300 μm mesh size. The net was towed vertically from 50 meters to the surface.
- The Calvet net which is characterized by 0.05 m^2 mouth area and 300 μm mesh size. The net was towed vertically from 70 meters to the surface.

All data is given in number of individuals per square meter by multiplying by 3 or 20 the volume of eggs and larvae sampled in 0.33 square meters for Hensen net, and 0.05 square meters for Calvet net.

To be consistent with the type of net, and since the mouth area of sampled water is not the same, spatial distribution and climatological maps have been studied separately.

Mean age of larvae is estimated at 5 days old, corresponding to a size range between 3 and 6 mm (P. Ayón pers. com.), and based on a length of hatching of 2 mm and a growth function (Marzloff et al., 2009). For sardines no age data was estimated and we assumed to be the same as for anchovy.

3.2.1.1 Spatial and temporal heterogeneity of datasets

The temporal distribution of eggs and larvae survey for Hensen and Calvet nets is given in Figures 3.1 and 3.2. This spatial and temporal heterogeneity of the sampling needs to be considered for the interpretation of the observed patterns.

For the Hensen net, the number of months sampled per year varied from 1 to 11 between 1961 and 2008, with an average of 5 (Fig. 3.1a). The spatial coverage was highest from 1961 to 1967 and from 1996 to 2008 (Fig. 3.1b). The number of years observed for a given month varied from 15 years to 21 years which gives a good representation of the seasonal variability (Fig. 3.1c). The spatial coverage per month (Fig. 3.1d) was higher during February and March. Therefore the other monthly abundances may be affected by higher uncertainties.

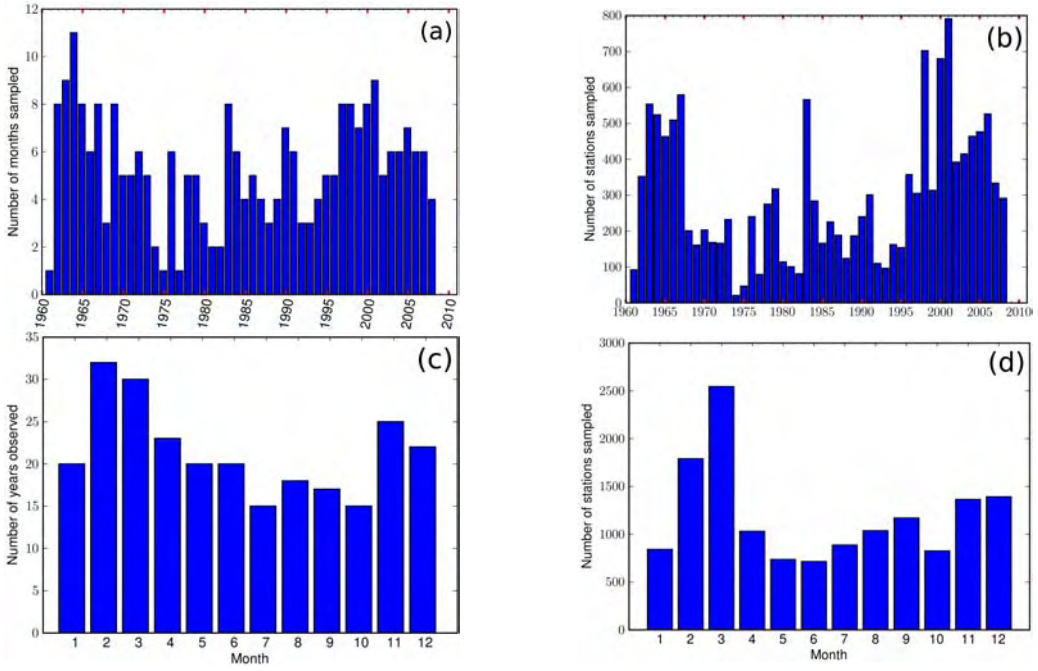

Figure 3.1: (a) Number of months sampled by year; (b) Number of stations sampled by year; (c) Number of years sampled by month; (d) Number of stations sampled by month - Dataset from Hensen net.

Sampling with the Calvet net started in 1981 during August and September, Fig. 3.2) but no cruise took place during "El Niño" events. This net was used for specific eggs and larvae surveys looking only for anchovy. Therefore, these dedicated surveys may tend to overestimate eggs and larvae density when compared to the Hensen net.

Figure 3.3 shows the total number of observations for Hensen and Calvet net on a cell of 1/6° of resolution. Coastal coverage was significantly higher than offshore coverage. For Hensen net, the spatial coverage was coarser and the number of observations per month and per cell was generally very low (<5).

It is worth mentioning that most of the survey target anchovy distribution and the sampling scheme is often adaptative and the ship perform transects from coast to offshore until anchovy disappears from the sampling. Sampling effort for anchovy can therefore be highly correlated to anchovy eggs and larvae abundances.

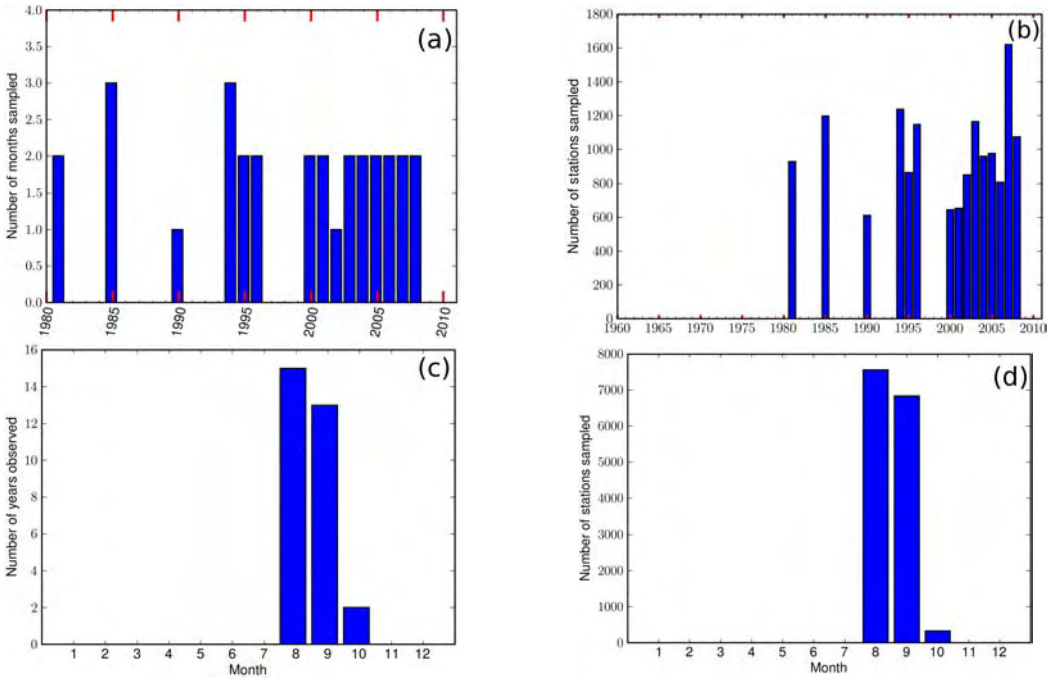

Figure 3.2: (a) Number of months sampled by year; (b) Number of stations sampled by year; (c) Number of years sampled by month; (d) Number of stations sampled by month-Dataset from Calvet net.

Finally, eggs and larvae surveys have been carried out over a long period, with different researcher teams, aims, and coverage. Furthermore, we should consider that the more eggs and larvae are collected in a net sample, the lower is the relative error due to double or missing counts. Since anchovies are more dispersed in winter than in summer (Lett et al., 2007) when the intensity of the upwelling is weaker, data could be partially biased.

Nevertheless, we will assume this bias to be small relative to the overall patterns discernible in our analysis.

Also, since we dispose of a large dataset, we will construct a climatology out of this heterogeneous dataset, assuming again that bias due to this heterogeneity is small compared to the overall seasonal and spatial patterns observed. However, we are aware that the potential bias is perhaps higher than presumed.

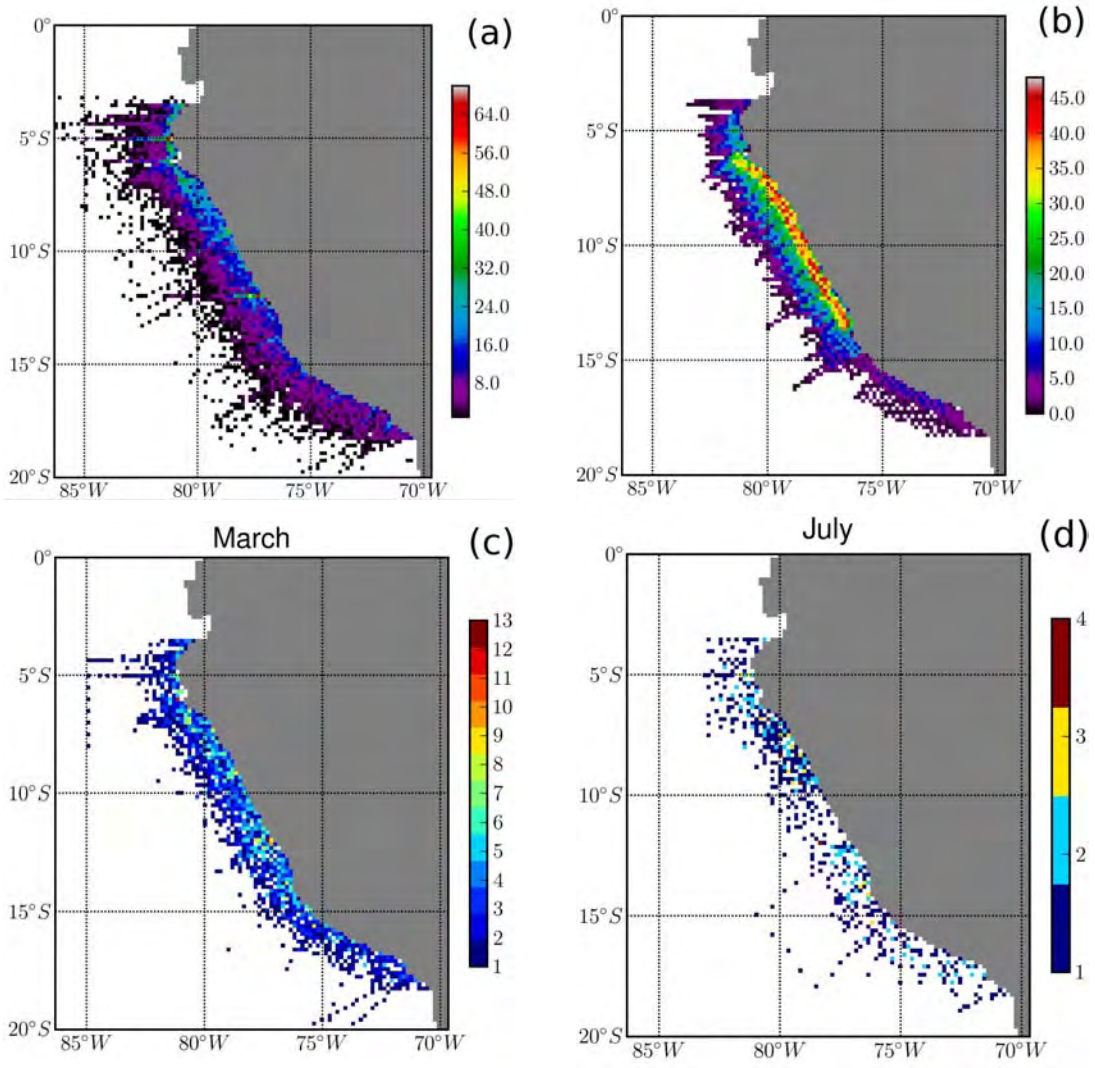

Figure 3.3: (a) Total number of observations at $1/6°$ resolution for Hensen net samples; (b) Total number of observations at $1/6°$ resolution for Calvet net samples. (c,d) respectively: Total number of observations (Hensen net) for March and July month respectively.

3.2.1.2 Statistical characteristics

Fish eggs and larvae are commonly distributed over space in a highly aggregated manner (Mangel and Smith, 1990). This highly aggregated distribution means that a high proportion of samples will have no eggs, whereas when one sample has one egg, it is likely to have many (Mangel and Smith, 1990).

Let's consider the example of anchovy eggs from 1992 to 2008. From 6567 samples, 36.6% showed the presence of eggs. The range of values varied from 0 to 76944 $individuals.m^{-2}$. The median was 0, the mean was 439 eggs per sample and the standard deviation was 2532 (i.e., coefficient of variation of about 577%). Without zeros, the mean increases to 1200 eggs per sample with a coefficient of variation of 339%. During the same period, which corresponds to a rise and a fall for anchovy and sardine populations, re-

spectively (Gutiérrez et al., 2007), only 6% of the samples from 1992 to 2000 had sardine eggs. This dataset had a mean of 19.0 and a standard deviation of 230.2 (i.e., coefficient of variation of 1200%).

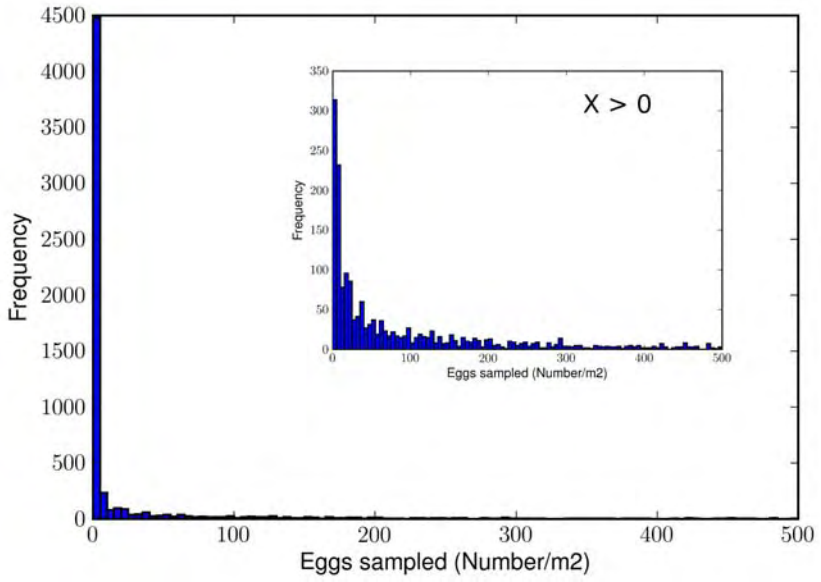

Figure 3.4: Frequency distribution of anchovy eggs from 1992 to 2008 for Hensen net.

Finally, anchovy and sardine eggs and larvae datasets are characterized by "excess zeros" and large outcomes due to their natural distribution over space. Mathematically, these datasets have the following characteristics:

Let X_i be the number of eggs or larvae taken at the i^{th} location.

- $X_i \geq 0$
- The variance of the data (X) exceeds the mean: $VarX_i > EX_i$, which is a "commonly accepted definition of an overdispersed or aggregated population" (Mangel and Smith, 1990).
- The probability of the data to have zero eggs (or larvae) is considerable.

These characteristics are common to the Negative Binomial Distribution and Zero Inflated Negative Binomial Distribution (Mangel and Smith, 1990; Minami et al., 2007).

Zeros in ecological data can be described as a "true zero", i.e., a result of unsuitable habitat condition, or due to species rarity, or as a "false zero", i.e., a failure of the observer to detect the patch data (Martin et al., 2005). The true zeros are more common for overdispersed data as eggs and larvae, and the second one for cryptic or secretive species (MacKenzie et al., 2002; Martin et al., 2005). Indeed, Mangel and Smith (1990) consider that samples with no eggs occur because they are taken beyond the current

habitat and therefore, use more than 50% of the samples with no eggs to help to delineate the boundaries of the habitat. Indeed, a clear delimitation between a favorable coastal habitat and unfavorable coastal habitat for anchovy are observed in Figures 3.8 and 3.15.

To validate this distribution in a second manner, we plotted the Quantile-Quantile plot (Q-Q plot) (Fig. 3.5 and Fig. 3.6).

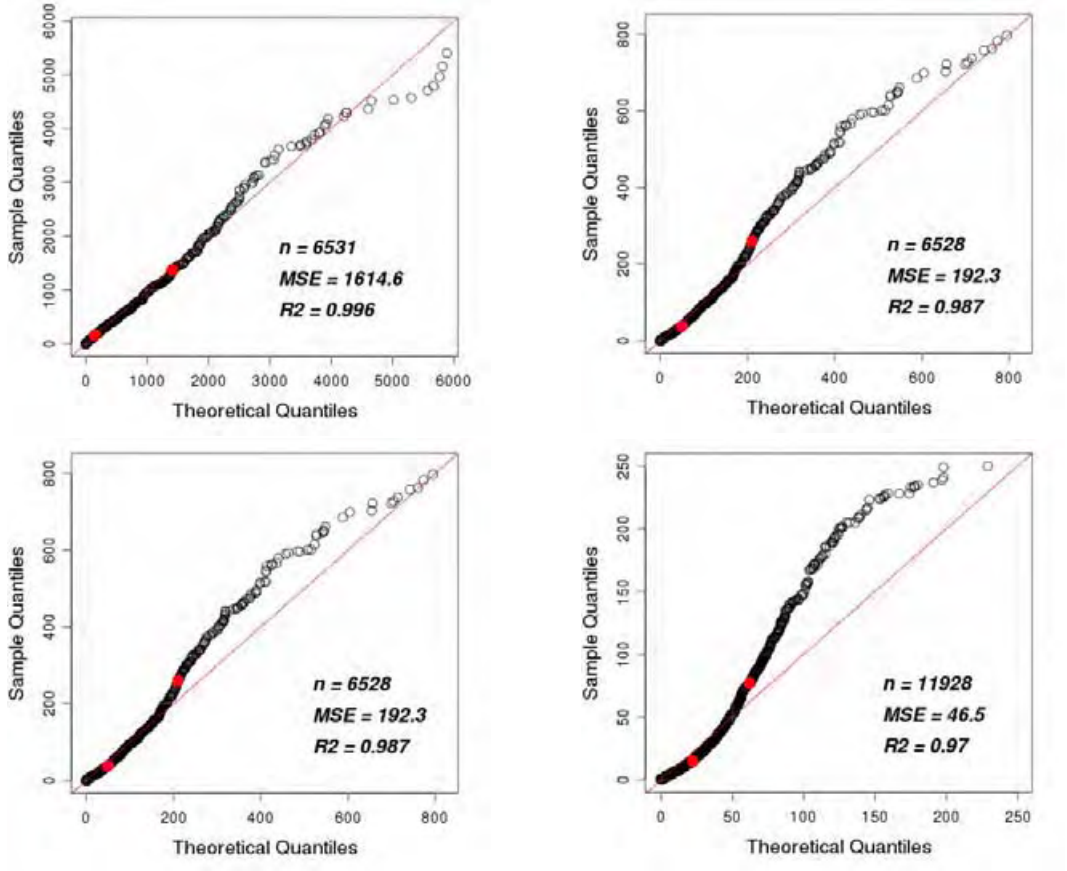

Figure 3.5: A Q-Q plot of eggs (left) and larvae (right) samples from 1992 to 2008 for Hensen net (top) and Calvet net (bottom) versus a Negative binomial distribution. The 95 and 98 quantile are shown in red, n corresponds to the number of observations used, r is the correlation coefficient, and MSE the mean square error.

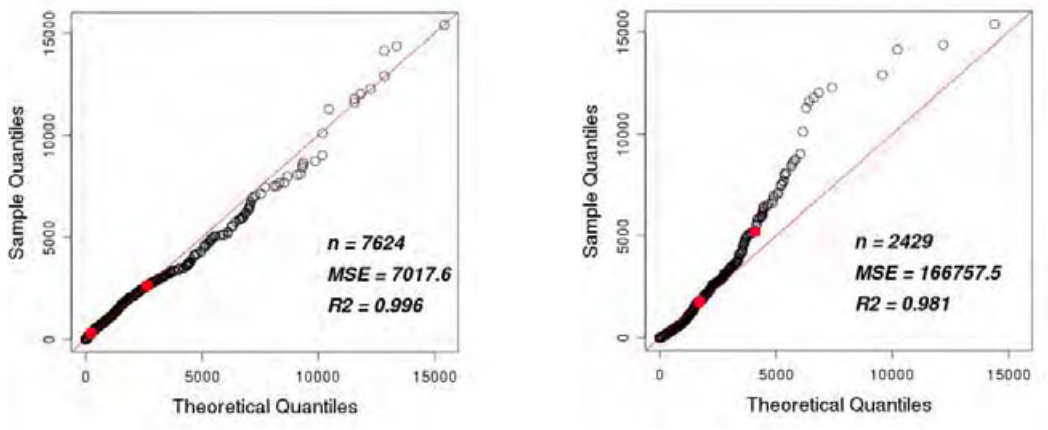

Figure 3.6: A Q-Q plot of a sample of eggs climatology versus a Negative binomial distribution (left) and Zero Inflated Negative Binomial Distribution (right). The 95 and 98 quantile are shown in red, n corresponds to the number of observations used, r is the correlation coefficient, and MSE the mean square error.

In this case, if data distribution and tested theoretical distribution are identical, the Q-Q plot follows the 45° line $y = x$. Normal, Log Normal, Poisson, Exponential, Negative binomial and Zero Inflated Negative binomial distribution were tested. Negative binomial was found to be the distribution fitting best the data with the minimum mean square error (MSE) and higher correlation R^2. For example, Figure 3.7 shows an example of bad fit using the normal distribution.

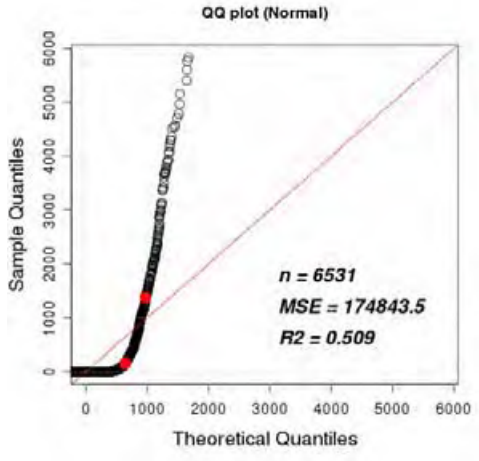

Figure 3.7: A Q-Q plot of eggs (a) samples from 1992 to 2008 for Hensen net versus a Normal distribution. The 95 and 98 quantile are shown in red, n corresponds to the number of observations used, r is the correlation coefficient, and MSE the mean square error.

For the anchovy climatology made with all datasets, the Zero Inflated Negative Binomial distribution and the Negative Binomial distribution gave the best fit with the data, the latter providing the best visual correlation. The choice of the model Negative Binomial or Zero Inflated Negative Binomial is quite controversial. Some publications conclude that for counts with high levels of zeros, Zero Inflated Negative Binomial distribution is more appropriate (Lewin et al., 2010), and others prefer the Negative Binomial (Warton, 2005; Vaudor et al., 2011). These last authors observed that even if the

proportion of zeros was high, data was generally well explained with models without zero inflation.

The model distribution gives us information about how the data is distributed and what we can expect from this data. NB or ZINB distributions indicate that the data contains high overdispersion, and therefore it is difficult to quantify the real amount of eggs or larvae when abundances are high.
Indeed, sampling scale can give information at different spatial scales. For anchovy and sardine the relevant spatial scales are presented in Table 3.1.

Units	Spatial Scales
Individual fish	cm
School	100 cm
Eggs Patches	1 km
School groups	10 km

Table 3.1: Relevant spatial scale for different fish life stages. Adapted from Mangel and Smith (1990).

Eggs and larvae sampling contain all these different scales. However, at $1/6°$ ($\sim$ 18 km) of resolution or $1/12°$ of resolution ($\sim$ 9 km), we cannot expect to model scales smaller than school groups. Therefore, modeling the extreme patchiness of data characterizing fish schools or eggs distributions would not be feasible. Nevertheless, these data will be used to optimize the parameters of our model, assuming that they contain sufficient information to predict the first order of variability.

Figure 3.8 presents the distribution of eggs, larvae and adults for one campaign from March to April 2001. As observed, for eggs and larvae in a given cell of $1/6$ degrees, we may have at the same time a sample with zeros and another one with high values. At the grid scale, heterogeneity of data is observed, but a "coherent" spatial distribution for eggs, larvae and adults can be observed in his entirety.

We may not expect to have a perfect fit of the individual data, but at least to represent the absence/presence patterns and the low amplitude signal of the data. This is illustrated on Figure 3.9, where for the same campaign, abundance patterns are "visible" only at "log normal" scale.

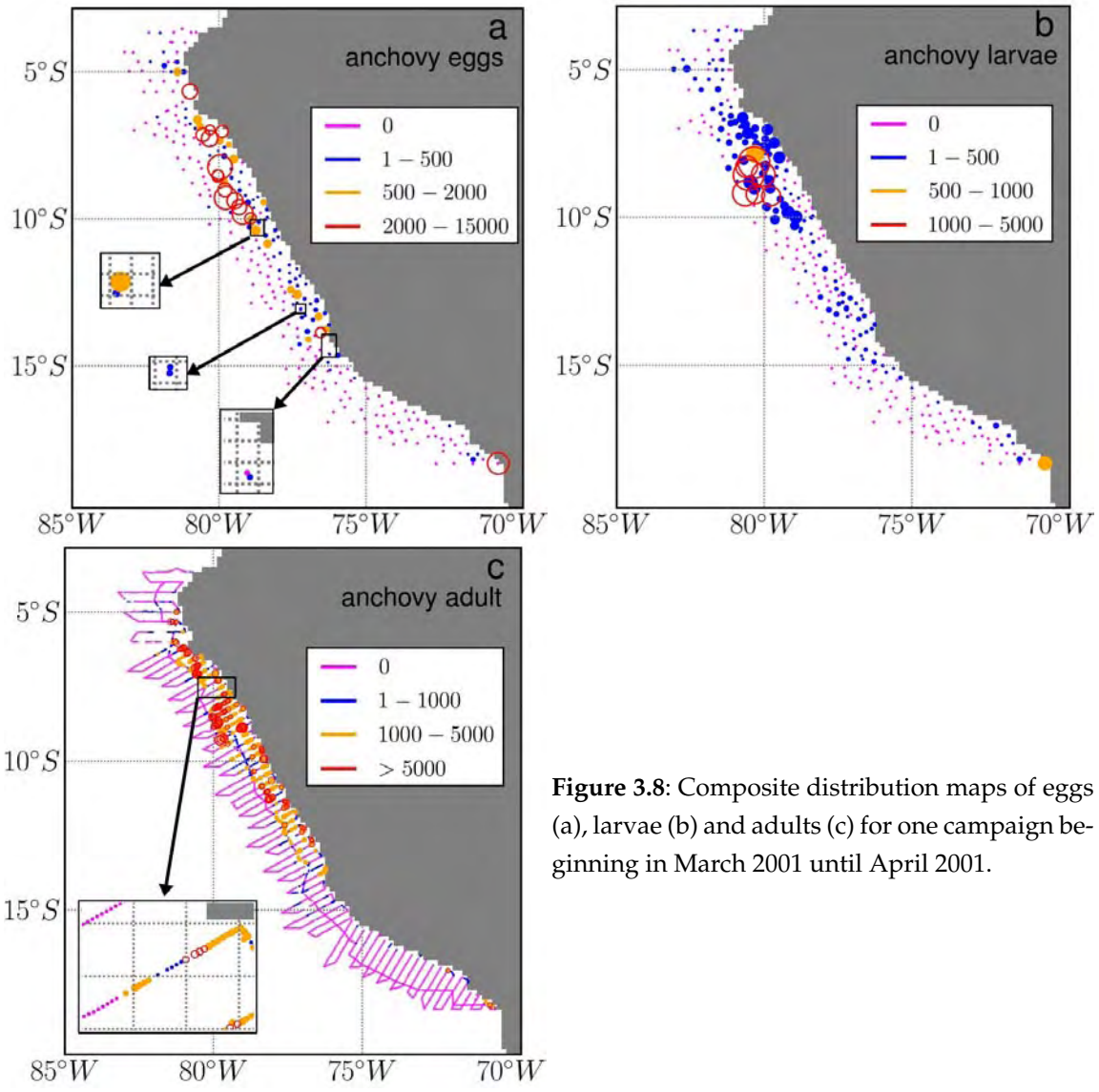

Figure 3.8: Composite distribution maps of eggs (a), larvae (b) and adults (c) for one campaign beginning in March 2001 until April 2001.

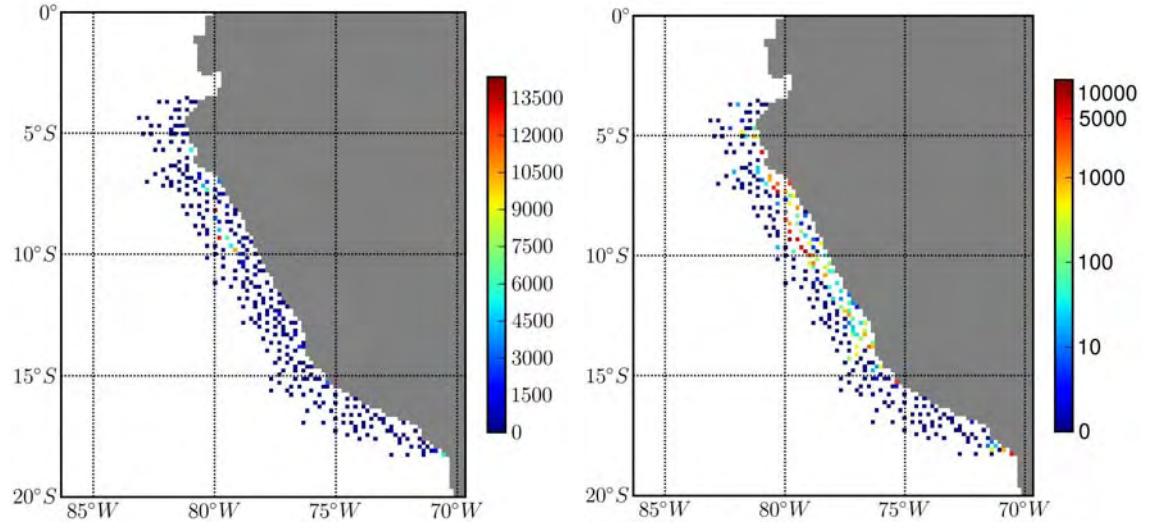

Figure 3.9: Map of eggs from March to April 2001 at the 1/6° resolution. At left, normal colorbar; at right log normal colorbar.

3.2.2 Adult data

Acoustical data of anchovy and sardine have been collected by the Instituto del Mar del Peru (IMARPE) since 1983. Here, we analyze 55 cruises from 1983 to 2009 which were collected from several vessels, most commonly the R/V Humboldt (76-m long), the R/V Olaya (41-m long) and the R/V SNP-1 (36-m long) with an inter-transect distance varying between 14 and 16 nautical miles (26-30 km) depending on the cruise. The acoustical data have been described elsewhere (Gutiérrez et al., 2007), here we only report the main points.

Over the years 1983-2008, the acoustic surveys have been performed using different types of Simrad echosounders (EK, EKS, EK400, EY500, EK500 and EK60), different types of frequencies (38 kHz between 1992 to 1996 and 120 kHz for the others surveys), and different types of calibration: "Until 1992, calibration was done by using hydrophones; after 1992, a standard procedure of calibration with a sphere (Foote and Sea, 1987) was used". According to Gutiérrez et al. (2007) and unpublished reviews, there is no bias in acoustic biomass values because of differences in calibration methods. However, they don't exclude bias caused by the large number of systems used over these 27 years, but since it cannot be quantified and retrospectively corrected, the authors consider the bias smaller in comparison to the overall patterns discernible during the analysis of observed data. We also made this hypothesis when we built a climatology with the 27 years of records.

Acoustic back-scattered energy by surface unit (s_A or NASC) was recorded in each geo-referenced elementary sampling distance unit (ESDU). Therefore, acoustic echo identification was performed by using fishing trawl results and echotrace characteristics. Between 55 and 660 trawls for each survey were done, with a mean of 190 trawls (Swartzman et al., 2008). Bias in identification of echotrace can still be high, representing one of the highest bias in the acoustical index (Castillo et al., 2009). For errors in acoustical processing see also Simmonds and Simmonds and MacLennan (2005).

The acoustical abundance is given in Nautical area backscattering coefficient in $m^2.nm^{-2}$:

$$s_A = s_a * (4\pi * 1852^2) \tag{3.1}$$

with s_a the area backscattering coefficient. On the major part of the analysis we will work directly with this acoustical abundance unit.

However, we could need to estimate the total biomass observed. The target strength TS, which is function of the length of the fish is needed to estimate this total abundance.

For anchovy, an empirical relationship built on data from 1999 to 2006 (Simmonds pers. com. and Simmonds et al., 2009) is used (eq. 3.2).

$$TS_{anchovy} = a + b * log10(L) = -36.71 dB/kg) \tag{3.2}$$

with $a = -33.81$ and $b = -2.61$, and L the mean length of anchovy sampled $L = 13\ cm$[1].

Then, to obtain a fish density (ρ_a) in $kg.nm^{-2}$:

$$\rho_a = \frac{s_A}{< \sigma_{sp} >} = \frac{s_A}{4\pi < \sigma_{bs} >} = \frac{s_A}{4\pi * 10^{TS/10}} \tag{3.3}$$

with σ_{sp} the Spherical scattering cross-section with $\sigma_{sp} = 4\pi\sigma_{bs}$ and σ_{bs} the backscattering cross-section.

The estimate of mean density of fish in $Individuals.km^{-2}$ is:

$$d = \rho_a * \frac{1/p}{1.852^2} \tag{3.4}$$

with p, the mean weight of the fish school. For an anchovy with an average mean length of 13 cm, the Von Bertalanffy growth equation fitted with coefficients from Marzloff et al. (2009), give a mean weight of 0.01485 g.

Using this approximation to build a climatology, we obtained an average biomass of 6 millions of tonnes (MT), fluctuating between 2.58 MT and 14 MT.

The conventions used here followed the review of Maclennan et al. (2002).
- Acoustical abundance is given **Nautical area scattering coefficient (NASC or s_A)** in $m^2.nmi{-}2$
- Spatial maps are plotted in **Nautical area scattering strength (S_A)** in dB.re.1($m^2.nmi{-}2$) following the equation 3.5:

$$S_A = 10 * log10(s_A) \tag{3.5}$$

3.2.2.1 Spatial and temporal heterogeneity of datasets

Spatial coverage also varied over time as observed in Figure 3.10a. Indeed, the acoustic back-scattered energy by surface unit (s_A) was recorded by geo-referenced elementary sampling distance unit (ESDU) of 2 nautical miles from 1983 to 1993, and 1 nautical mile after 1994. The months sampled per year vary from 2 to 9 between 1983 and 2008, with an average of 4.5 months per year sampled (Fig. 3.10b). The number of years sampled for a given month varied from 4 to 17 years, allowing for a fairly good representation of the seasonal variability(Fig. 3.10c). The spatial coverage by month (Fig. 3.10d) was higher in

[1] Acoustic data take into account fish larger than 4 cm and therefore our 'adult' acoustic biomass contains also juveniles.

March. Therefore the others months may be affected by higher uncertainties. For sardine the results are similar (not shown).

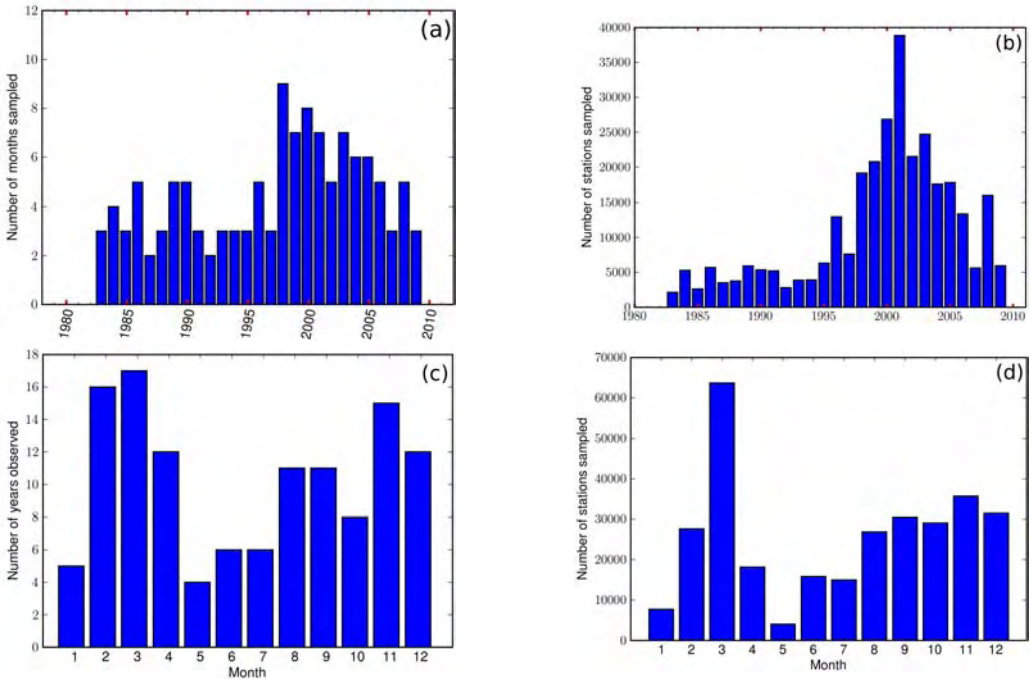

Figure 3.10: Acoustical (anchovy) data - (a) Number of months sampled by year; (b) Number of stations sampled by year; (c) Number of years sampled by month; (d) Number of stations sampled by month.

As for eggs and larvae sampling, acoustic sampling effort was also highly correlated to adult abundances (Fig. 3.11). Indeed, as seen on Figure 3.8, the sampling strategy was designed to focus on anchovy. Spatial resolution of acoustic is higher than the one for eggs and larvae with a mean of 7 samples per cell.

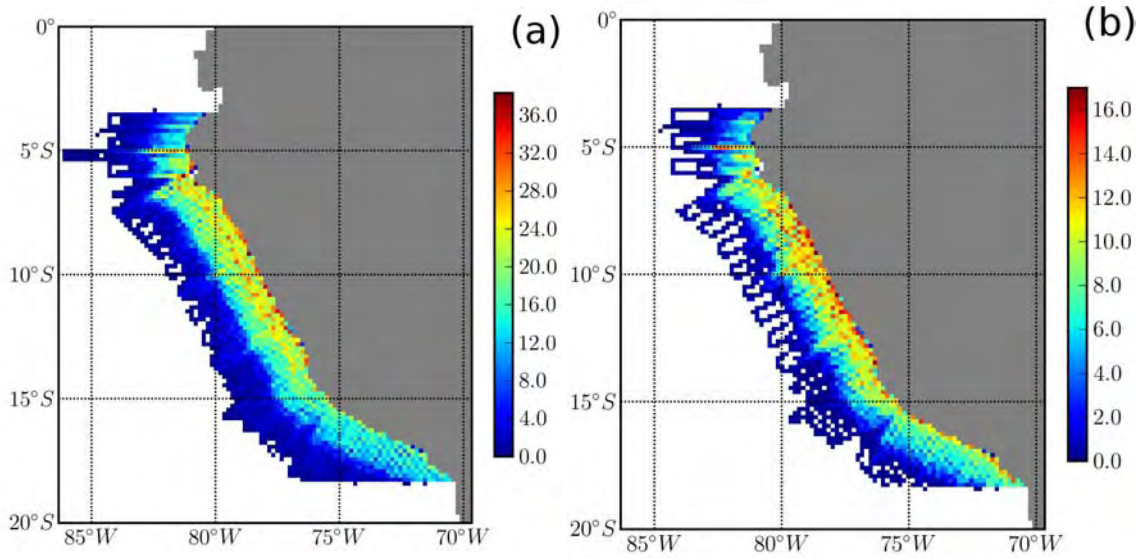

Figure 3.11: Annual mean number of observations at 1/6° resolution for anchovy (a) and sardines (b), for all data from 1983 to 2009 and 1983 to 2000 respectively.

3.2.2.2 Statistical characteristics

As eggs and larvae, distribution of adult is very patchy. Nevertheless, acoustic data is a continuous measure, thus reducing the error along-transect. But at the difference of eggs and larvae distributions that are essentially driven by environmental conditions, adults can have different levels of fish aggregation due to self-organisation (Fig. 3.12).

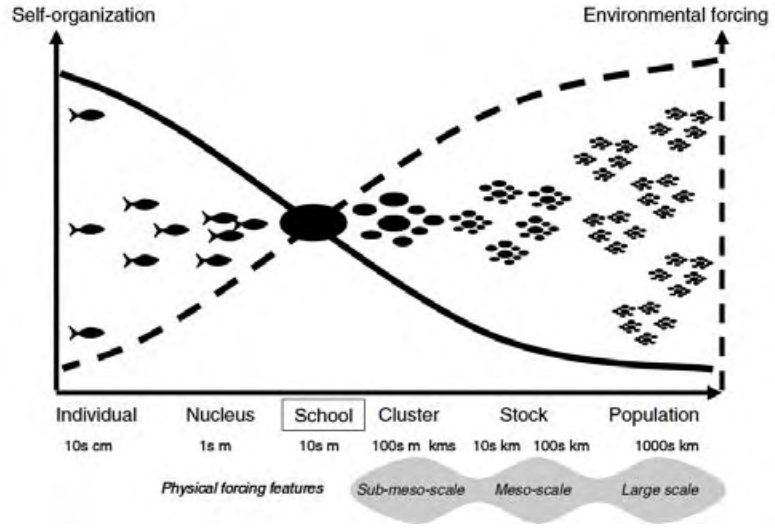

Figure 3.12: Spatial scales of fish distribution depending on self organizations and environmental forcing. At smaller scales, self organization mechanisms are likely to be dominant; at larger-scales adaptation and response to the environment are more important. Source: Bertrand et al. (2008)

Like eggs and larvae, adult acoustical data contains high number of zeros. They most often correspond to unsuitable habitat conditions and can be used to delineate the boundaries of the habitat. A clear delimitation between a favorable and unfavorable coastal habitat for anchovy is observed in Figures 3.8 and 3.15.

As an example, anchovy adult data from 1992 to 2008 included 259 935 samples, but only 25.6% of non zero data. Also the range of sA positive values varied from 0 to 438 $m^2.nm^{-2}$. The median was 0, the mean was 223 $m^2.nm^{-2}$ and the standard deviation was 1995 (i.e., coefficient of variation of about 896%). For sardines, only 5.9% of data had adults from 1992 to 2000.

These characteristics are common to the Negative Binomial Distribution and Zero Inflated Negative Binomial Distribution (Mangel and Smith, 1990; Minami et al., 2007, Fig. 3.13). However, none of these two theoretical distributions fit correctly anchovy and sardine dataset, even if the Negative Binomial was the least bad fit. This is perhaps due to the high overdispersion and at the same time self-organization.

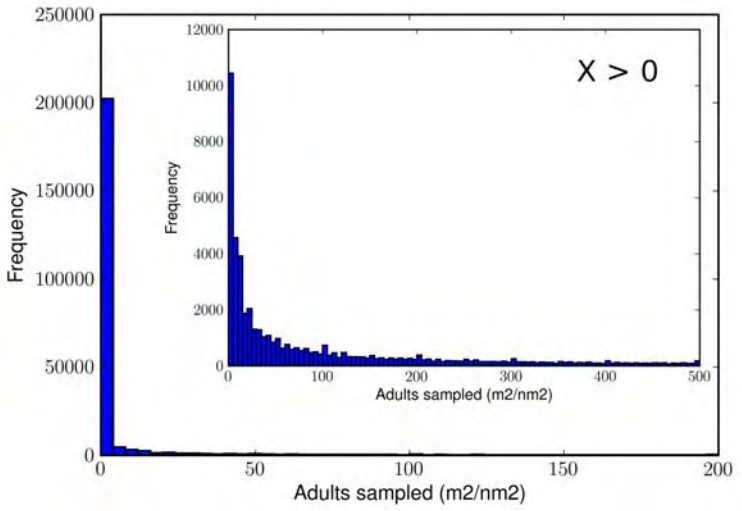

Figure 3.13: Frequency distribution of anchovy adults (in $m^2.nm^{-2}$) from 1992 to 2008.

3.2.3 Gonadosomatic Index

To study the seasonal and interannual variability we also analyzed the changes in the gonadosomatic index (GSI) of anchovy, i.e., the ratio of fish gonad weight to body weight. The GSI has been measured by IMARPE during the 1990's (B. Buitrón pers.com.). GSI is a good indicator of the spawning season since ovaries increase in weight with the maturation process of oocytes and decrease after the release of eggs. The exact peak of spawning however can be difficult to detect due to a range of variability both in the population and in relation with the environmental conditions over the sampled area. The phenomena of atresia, i.e., the resorption of non spawned oocytes, can mask the coincidence between estimated spawning peak and observed eggs density.

3.3 Spatial Distribution

3.3.1 All datasets

Anchovy and sardine eggs, larvae and adults were collected all along the Peruvian coast between 5°S and 18°S, however they were mainly found between 6°S and 14°S. To analyze the spatial distribution of the data, eggs and larvae data were separated according to Hensen (Fig. 3.15) or Calvet net sampling criteria (Fig. 3.16) to see if consistent spatial distribution appears even if only the Calvet net was used, during August and September.

Highest concentrations were observed near the coast, at the edge of the continental shelf (Fig. 3.15). Though eggs were found all along the coast without particular concentration patterns, larvae appeared in higher density in the northern region from 6°S to 9°S (Fig. 3.15, 3.16). A possible enrichment by larval drift from nearby regions or better survival rates could explain this favorable region. Lett et al. (2007) already showed with an Individual-Based Model approach that this northern region is an area of accumulation, due to northward surface transport by the Peru Coastal Current. Eggs and larvae from anchovy and sardines appear to be distributed along the coast, as for anchovy adults, whereas adult sardines locate mostly offshore. Indeed, mean cross-shore profiles for anchovy and sardine between 1983 and 2005 (Fig. 3.14) illustrate that the peak of adult anchovy abundance happens along the coast whereas for sardine, it occurs at about 100 km offshore.

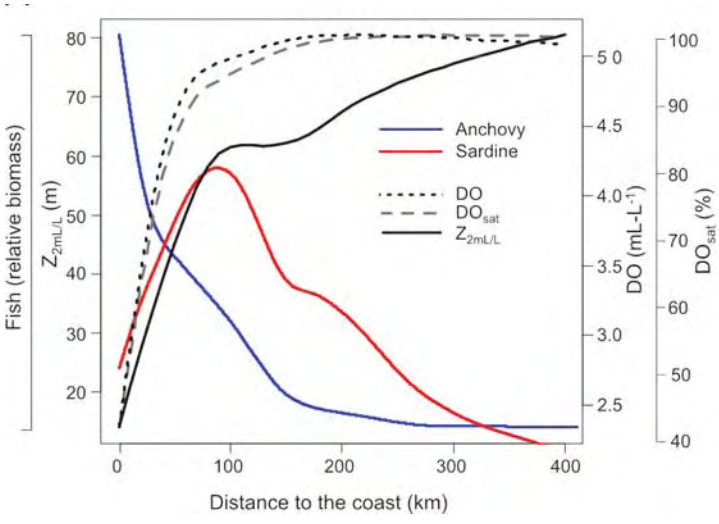

Figure 3.14: Mean cross-shore profiles of anchovy (blue solid line) and sardine (red solid line) acoustic biomass, and near surface dissolved oxygen (DO), percentage of dissolved - oxygen saturation (DO_{sat}) and oxycline depth corresponding to a concentration of 2mL/L ($Z_{2mlL/L}$). Mean average of data from 1983 to 2005. Source: Bertrand et al. (2011).

Parrish et al. (1981) already described that "*pure schools of adults sardine often migrate long distances between their feeding and spawning grounds whereas schools of anchovies tend to remain resident in or near the habitat or reproduction*". This hypothesis is reinforced by the fact that average swimming speed of sardines is expected to be higher than the anchovy speed. Indeed, Gerlotto et al. (2006) evaluated an average of anchovy speed to 0.28

$m.s^{-1}$ without predation and indicated that anchovy is a relatively slow fish, especially compared with clupeids as sardines which speed vary from 0.56 to 1.89 $m.s^{-1}$ (*Sardinops sagax*, South Africa, Misund et al., 2003).

Finally, anchovy eggs and adults appear to share the same habitat. Feeding habitat and spawning habitat cannot therefore be separated. However since mortality is probably higher in the south, higher abundances of larvae is observed on the northern regions.

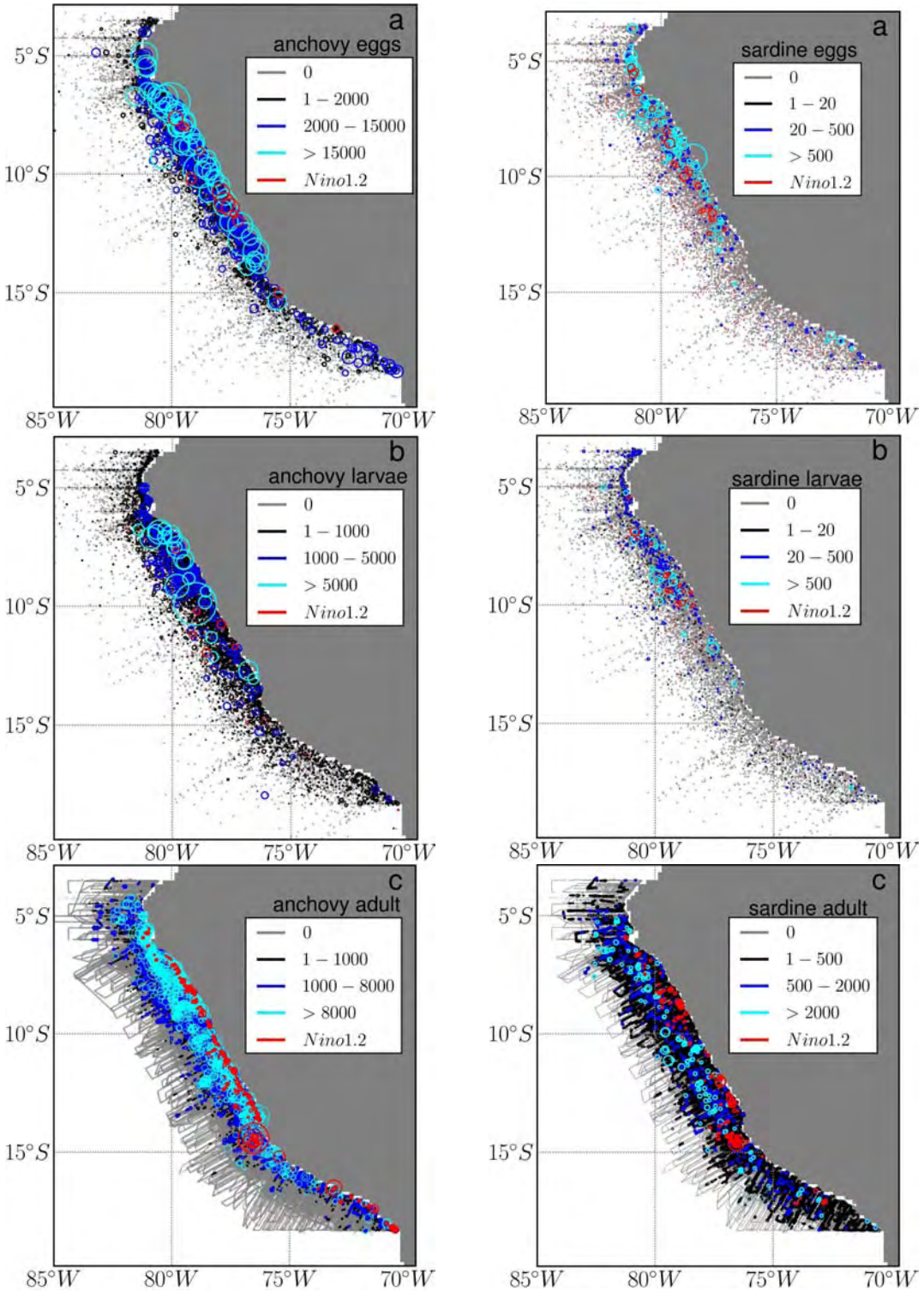

Figure 3.15: Composite distribution maps for eggs (a), larvae (b) and adult(c) of anchovy (left) and sardine (right) collected by the Instituto del Mar del Peru combining all data over the period 1961-2008 (a and b) and 1983-2009 (c). Eggs and larvae shown here were collected with the Hensen net. Circles radius are proportional to density values with the higher biggest circle corresponding to (a) 107 376 (anchovy) and 38 016 (sardine) $eggs.m^{-2}$, (b) 84 939 (anchovy) and 8856 (sardine) $larvae.m^{-2}$ and (c) s_A=438 333 (anchovy) and s_A=58 057 (sardine) $nm.m^{-2}$. All data were provided by IMARPE.

Spatial distribution of anchovy eggs and larvae from Calvet nets (Fig. 3.16) present the same spatial pattern as the one observed previously with Hensen nets. For sardines, it appears that no eggs at all were collected with Calvet nets. Possible reasons include a lack of focus with no observational effort made on sardines eggs counting, or an absence of spawning at that time period (August-September). However, this last option is not verified since eggs are observed at that time using the Hensen net. Sardine larvae are located from 6°S to 14°S.

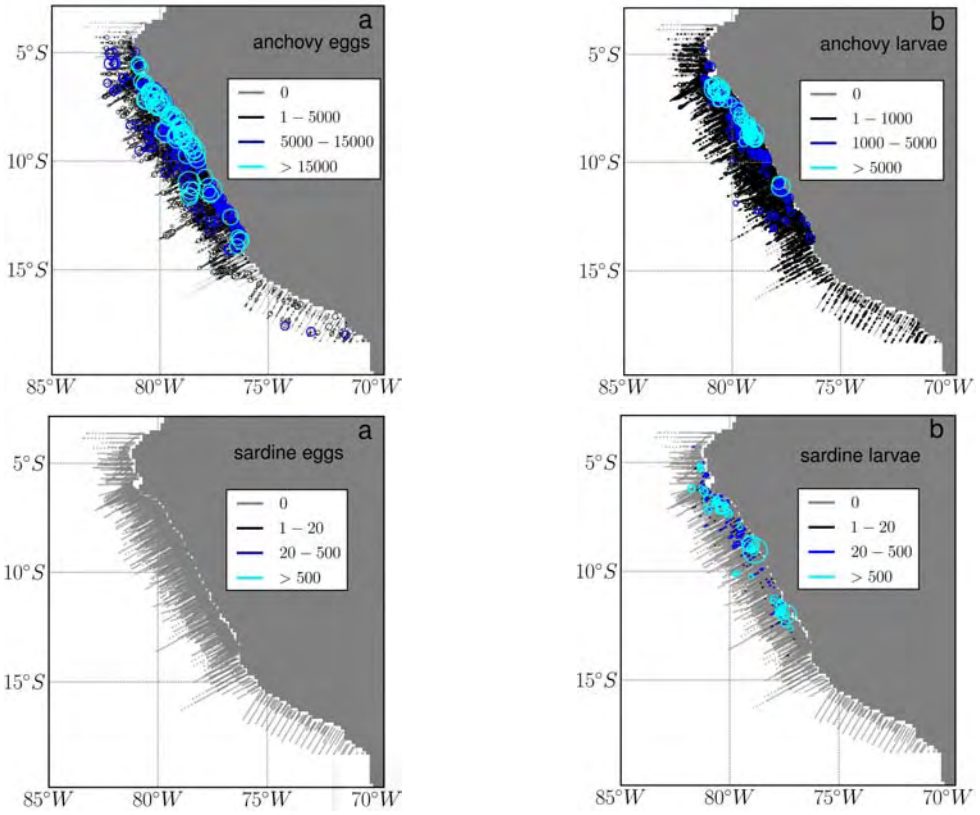

Figure 3.16: Composite distribution maps for eggs (a) and larvae (b) collected by the Instituto del Mar del Peru combining all data over the period 1961-2008 (a and b). Eggs and larvae shown here were collected with the Calvet net during August and September month. Circles radius are proportional to density values with the higher biggest circle corresponding to (a) 44 820 (anchovy) and 5136 (sardine) $eggs.m^{-2}$, (b) 29 680 (anchovy) and 33 440 (sardine) $larvae.m^{-2}$. All data were provided by IMARPE.

Based on the observed density of eggs and larvae, a spatial aggregation has been defined with a mask of five coastal and offshore regions (Fig. 2.17). The offshore regions are the offshore zone (OZ) and the transition zone (TZ), and the coastal zones are the Northern coastal zone (NCZ), the Central coastal zone (CCZ), and the Southern coastal zone (SCZ).

The northern coastal zone is the region where the highest eggs and larvae densities occurred while the offshore zone is used to characterize the null habitat, with absence of

eggs and larvae. Simulation outputs will be integrated over these areas and by month to be compared to observation.

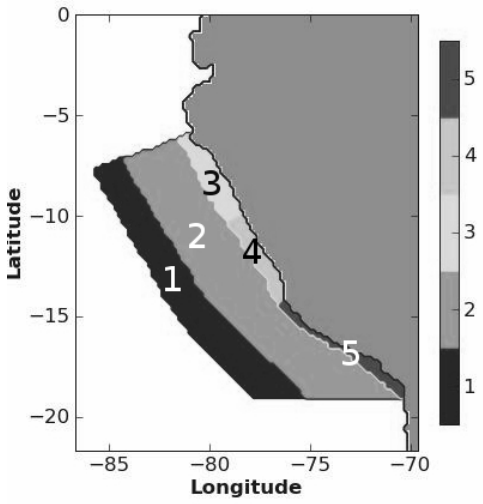

Figure 3.17: Mask used in the analysis with five coastal and offshore regions. OZ=offshore zone (1), TZ= transition zone (2), NCZ= Northern coastal zone (3), CCZ=Central coastal zone (4), SCZ=Southern coastal zone (5).

3.3.2 Climatology maps

Monthly climatology data maps were created by averaging eggs and larvae densities at the resolution of the grid model (1/6°, cf. below) after removing the data collected during the most powerful El Niño events (see all figures in Annex A). We choose to exclude these events from the data climatology, since in the modelling approach, model predictions will be forced by a climatology that do not contain Niño events. The predictions have therefore to be compared to "non-Niño" observations.

It should be noted that Calvet net was used for specific eggs and larvae survey in August-September, with a particular focus between 5° S and 15°S. Therefore, these dedicated surveys may tend to an overestimation of eggs and larvae density when compared with the other months. Thus the climatologies were produced using all available data between 1961-2008 (159 cruises) for eggs and larvae and 1983-2009 for adult density (55 cruises) without using these specific cruises. For eggs and larvae, 0.2% of higher data were not taken into account to make the climatology.

Before processing eggs, larvae and adult data, the samples collected during strong El Niño events were excluded. According to NOAA standards, El Niño and la Niña events are characterized by an ONI (Niño 3.4 region, see Fig. 3.18) greater than or equal to 0.5°C and lower than or equal to -0.5°C respectively for five consecutive overlapping 3-month seasons. For our study, we define a threshold value of +/1°C of the sea surface temperature measured on the Niño1.2 region (closest to Peruvian region) to characterized the major El Niño events (Fig. 3.18).

As a result, the following periods were removed: 04/1965-08/1965, 05/1972-01/1973, 06/1976-09/1976, 09/1982-10/1983, 02/1987- 07/1987, 02/1992-06/1992, 04/1997-08/1998.

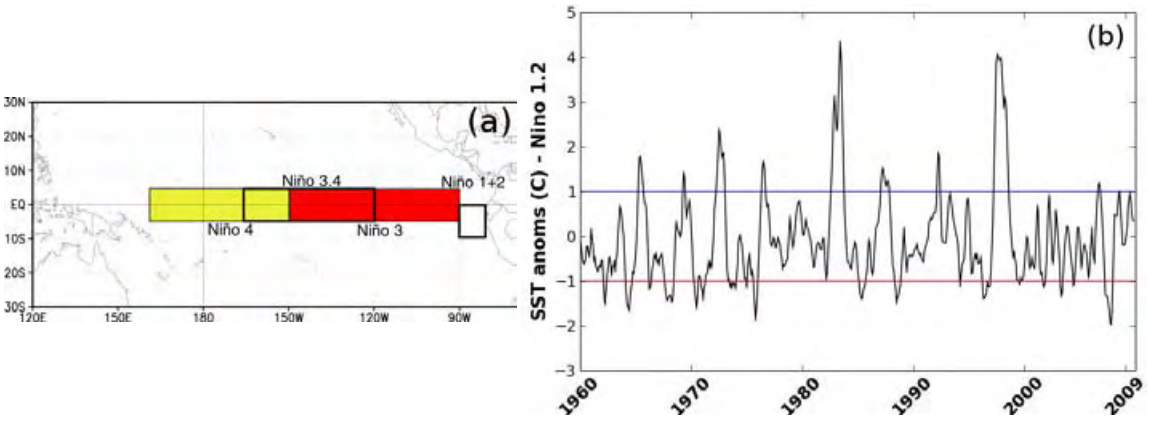

Figure 3.18: (a) Graphical description of the four Niño regions (from: http://www.cpc.ncep.noaa.gov/products/analysis_monitoring/ensostuff/nino_regions.shtml; (b) Time series (1960-2010) of oceanic Niño index (ONI), i.e area-averaged 3-month running mean SST anomalies (°C) in the Niño 1.2 region (10°S-0°N, 80°W-90°W). Data from http://www.cpc.ncep.noaa.gov/data/indices/sstoi.indices).

3.4 Seasonal variability

For the study of the seasonality only data from Hensen net is used.

3.4.1 Anchovy

Densities of eggs and larvae collected at sea are highly variable in time and space with large dispersion and very high proportion of null samples (Fig. 3.15). Due to obvious non-Gaussian distributions of anchovy data, the seasonality was explored using non-parametric statistics. Mann Whitney test (also called Wilcoxon rank-sum test) was used to make a general pairwise comparison. This test allows to assess whether one of two groups that do not follow a normal distribution tends to have larger values than the other. The null hypothesis tested here is that the two groups are significantly different. We respect the null hypothesis if the p-value is lower than 0.05. This analysis allowed to identify the periods of high and low abundance of eggs and larvae (Fig. 3.19).

We studied the monthly seasonality of all data for eggs, larvae and adults (Fig. 3.19, top) as well as the seasonal variability by regions (Fig. 3.19, bottom).

Highest abundance of eggs occurred along the coast in region NCZ and CCZ, with density starting to increase after July, decreasing in October, and increasing again from December to February to peak in March. Although a major peak occurred in September, egg abundance was not significantly higher in September than in August

(Mann-Whitney U test, p=0.4). Egg density decreased when moving towards the south (region SCZ) and was very low in the southern coastal region SCZ. No significant seasonality differences was observed (Mann-Whitney U test, p=0.7) probably because the standard error of the mean was very high. A single small peak occurred (in August) off coast in region TZ (Fig. 3.19), and almost no eggs were sampled far offshore in region OT.

The seasonal pattern for larvae was very similar to the one for eggs, with a high peak of density in September and a secondary peak between December and March (Fig. 3.19). The observed density was maximal in the north (region NCZ) and decreased southward followed by region CCZ (see also Fig. 3.15). In the southern coastal region (region SCZ) density increased in July, earlier than further north. In the transition zone (region TZ), the density of larvae was maximum in September and March.

Spatial and seasonal abundance estimates from acoustical sampling suggest that anchovy adult density was higher from February to March along the coast (Fig. 3.19), either because habitat contraction and fish aggregation or due to annual peak in recruitment after one year. In both cases, higher concentration of adults would amplify the spawning success in the coastal habitat through stock-larval recruitment relationship.

The seasonality is marked by a clear reproductive rest period during austral autumn, from April to June (Table 3.2). Indeed, Mann-Withney test was always significant between austral summer and austral autumn as well as between austral autumn and austral winter (Table 3.2). Furthermore, for most of the regions, differences were not significant between austral summer and austral winter, meaning that both periods can have the same abundances.

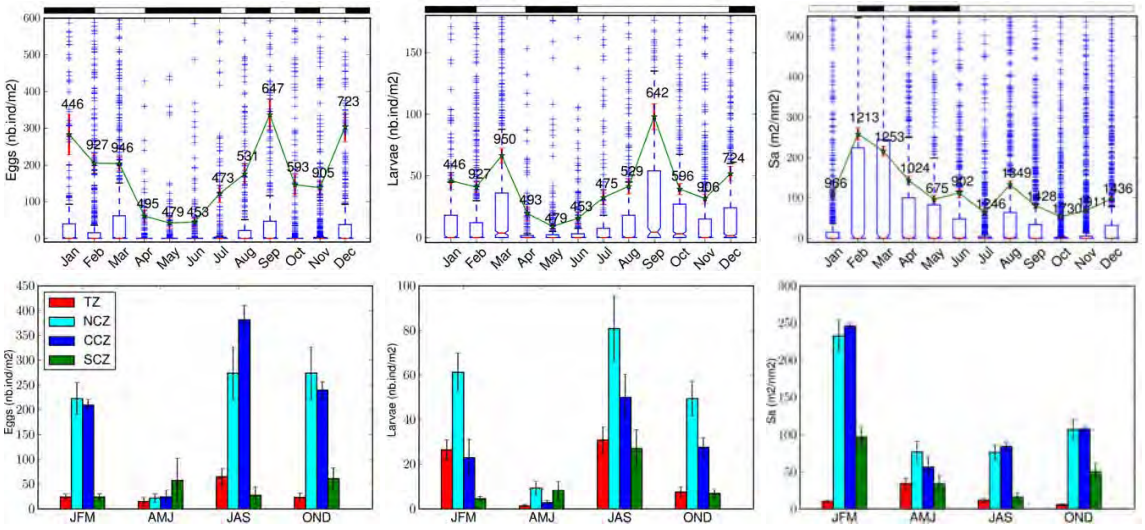

Figure 3.19: Seasonal variability in anchovy eggs, larvae and adult density. Eggs and larvae density are given in number of individual per square meters. 0.2% of outliers of eggs and larvae were removed. Adults abundance are in acoustical index (s_A). Top: Monthly average variability for all regions (Fig. 3.3.1) with box plots showing the median and the 25th and 75th percentiles, the mean (green line) and standard error of the mean (red bars), and the outliers (blue crosses). The numbers indicate the number of cells covered by observations (total number of cells for all regions being 2029). The result of the statistical analysis is illustrated with black and white bands where the black (white) color indicates non-significant (significant) difference between two consecutive (Mann-Withney U test; p>0.05). Bottom: Mean and standard error of the mean of eggs, larvae and adult density by season and region.

	Pairs	TZ	NCZ	CCZ	SCZ
Eggs	(JFM, AMF)	JFM > AMF			
	(JFM, JAS)	JFM ∼ JAS			
	(JFM, OND)	JFM ∼ OND		JFM > OND	
	(AMF, JAS)	AMF < JAS			AMF ∼ JAS
	(AMF, OND)	AMF ∼ OND	AMF < OND		AMF ∼ OND
	(JAS, OND)	JAS > OND	JAS ∼ OND		
Larvae	(JFM, AMF)	JFM > AMF			
	(JFM, JAS)	JFM ∼ JAS			
	(JFM, OND)	JFM > OND	JFM ∼ OND		
	(AMF, JAS)	AMF < JAS			
	(AMF, OND)	AMF < OND			
	(JAS, OND)	JAS ∼ OND			
Adults	(JFM, AMF)	JFM > AMF			
	(JFM, JAS)	JFM > JAS			
	(JFM, OND)	JFM > OND	JFM ∼ OND	JFM > OND	
	(AMF, JAS)	AMF > JAS	AMF ∼ JAS		
	(AMF, OND)	AMF > OND	AMF ∼ OND		
	(JAS, OND)	JAS ∼ OND			JAS < OND

Table 3.2: Mann-Whitney U test results of abundance significance between different seasons for anchovy eggs, larvae and adults.

To complete this description, it is useful to compare the seasonal cycles of eggs and larvae with the change of monthly mean of GSI. Overall, GSI seasonal pattern agrees with observed seasonal cycles of eggs and larvae densities in northern and central coastal zones with a shift of one month. GSI starts increasing in June from mean value below 3 and peaks in September above 5 (Fig. 3.20). Then it decreases slightly but stays at rather high values (above 4) until February, showing the rapid decline in the following months with the lowest values observed in April-May.

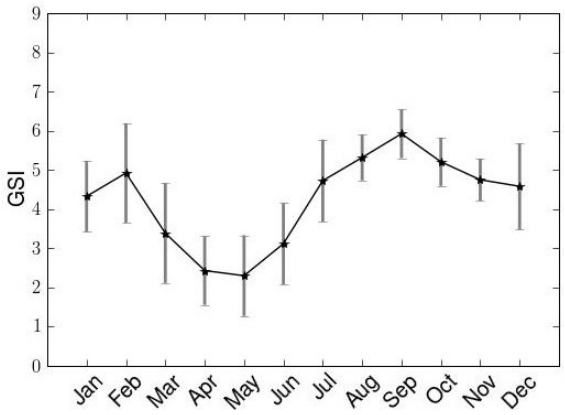

Figure 3.20: Monthly average of Gonado-Somatic Index of anchovy. GSI average and standard error are collected by IMARPE between 1990 and 1999 (Buitrón pers. com.) after removing the period corresponding with the 1997-98 El Niño event.

In summary, GSI with observed density of eggs and larvae show similar and coherent seasonal patterns with a shift of one month. GSI suggested a clear reproductive rest period taking place in April-May, with a maturation of gonads starting in June and high peak in September. The spawning period continues but at lower intensity until February and GSI starts to decrease in March, while eggs and larvae are still abundant.

3.4.2 Sardine

As for anchovy, we studied the monthly seasonality of all data for eggs, larvae and adults (Fig. 3.21, top) as well as the seasonal variability by regions (Fig. 3.21, bottom).

Although anchovy abundance of eggs and larvae show a peak in September followed by a plateau until March, for sardines, one main peak in August-September and a secondary peak in January are clearly separated by a period of lower abundance (Fig. 3.21).

For sardine, highest abundance of eggs occurred along the coast in region NCZ, like for anchovy. For eggs, a clear maximum of abundance appears in January and in August. However, between August and September, the difference was not significant. Also a rest period occurred between October to December and April to July. But from February to July, changes in abundances were not statistically significant (Mann-Whitney U test: p=0.4).

Larvae presented a slightly different pattern with higher abundances from January to May and from August to September and lower abundances from June to July and from October to November.

For adults, the peak of abundance occurred in April-May for all regions. These abundances were not correlated with observed acoustical abundance, however April-May month are the less sampled and therefore affected by higher uncertainties.

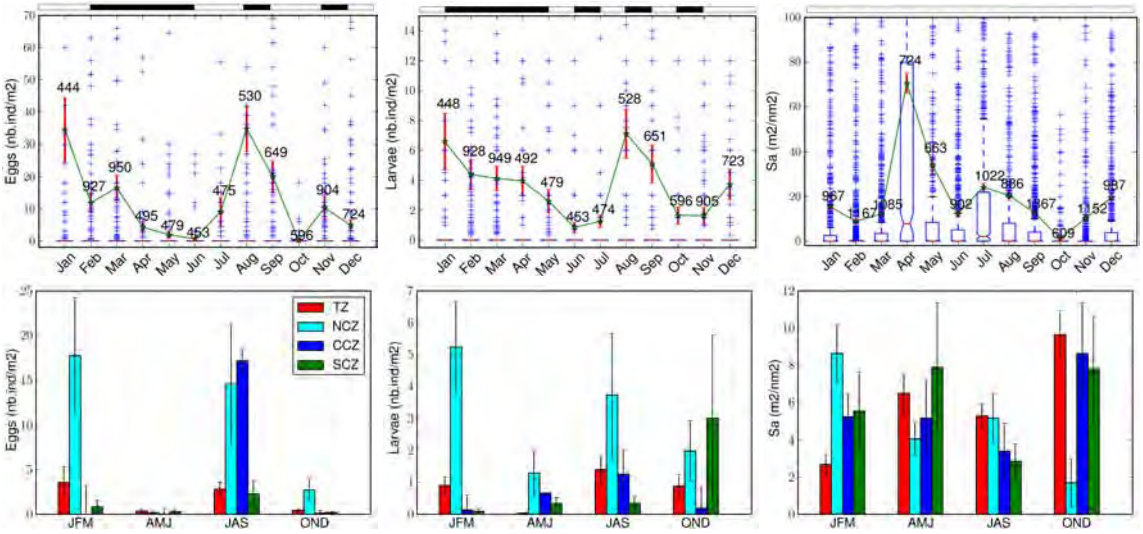

Figure 3.21: Seasonal variability in sardine eggs, larvae and adults density. Eggs and larvae density are given in number of individual per square meters. 0.2% of outliers of eggs and larvae were removed. Adults abundance are in acoustical index (s_A). Top: Monthly average variability for all regions (Fig. 3.3.1) with box plots showing the median and the 25th and 75th percentiles, the mean (green line) and standard error of the mean (red bars), and the outliers (blue crosses). The numbers indicate the number of cells covered by observations (total number of cells for all regions being 2029). The result of the statistical analysis is illustrated with black and white bands where the black (white) color indicates non-significant (significant) difference between two consecutive (Mann-Withney U test; p>0.05). Bottom: Mean and standard error of the mean of eggs, larvae and adults density by season and region.

	Pairs	TZ	NCZ	CCZ	SCZ
Eggs	(JFM, AMF)	JFM > AMF	JFM ∼ AMF	JFM > AMF	Not significant
	(JFM, JAS)	JFM < JAS	JFM ∼ JAS	JFM < JAS	
	(JFM, OND)	JFM ∼ OND	JFM > OND	JFM ∼ OND	
	(AMF, JAS)	AMF < JAS		AMF ∼ JAS	
	(AMF, OND)	AMF < OND		AMF ∼ OND	
	(JAS, OND)	JAS > OND	JAS ∼ OND	JAS > OND	
Larvae	(JFM, AMF)	JFM > AMF		Not significant	Not significant
	(JFM, JAS)	JFM ∼ JAS			
	(JFM, OND)	JFM ∼ OND	JFM > OND		
	(AMF, JAS)	AMF < JAS	AMF ∼ JAS		
	(AMF, OND)	AMF < OND	AMF ∼ OND		
	(JAS, OND)	JAS > OND	JAS ∼ OND		
Adults	(JFM, AMF)	JFM < AMF	JFM ∼ AMF	Not significant	Not significant
	(JFM, JAS)	JFM < JAS			
	(JFM, OND)	JFM < OND	JFM > OND		
	(AMF, JAS)	AMF ∼ JAS			
	(AMF, OND)	AMF ∼ OND	AMF > OND		
	(JAS, OND)	JAS ∼ OND	JAS > OND		

Table 3.3: Mann-Whitney U test results of abundance significance between different seasons for sardine eggs, larvae and adults.

3.5 Interannual Variability: El Niño events

Many studies have analyzed the impact of El Niño on anchovy and sardine populations (e.g., Bertrand et al., 2004; Alheit and Niquen, 2004; Bertrand et al., 2008). Here we present briefly the data used in the model and some important points to be considered.

El Niño event is known to produce a dramatic reduction in anchovy biomass, however each El Niño is different, and stock recovery afterward does not follow the same pattern. "*Recovery was slow after El Niño events in 1982-73, 1977-78 and 1982-83, but rapid after El Niño of 1987 and 1997-98*" (Bertrand et al., 2004). To understand the impact of an El Niño event on the different stocks, we need to evaluate the combination of different factors at different spatiotemporal scales: Inderdecadal regime, strength and duration of the El Niño event, fishing pressure, predation, adaptation of reproductive behaviour, variability of the oxycline and others.

Figure 3.22 shows the spatial distribution of eggs, larvae and adults for anchovy and sardine during all El Niño events over the period 1961-2008 (1983-2009 for adults). Both species were concentrated along the coastline but sardine are still more offshore than anchovy. Adults were highly concentrated along the coast.

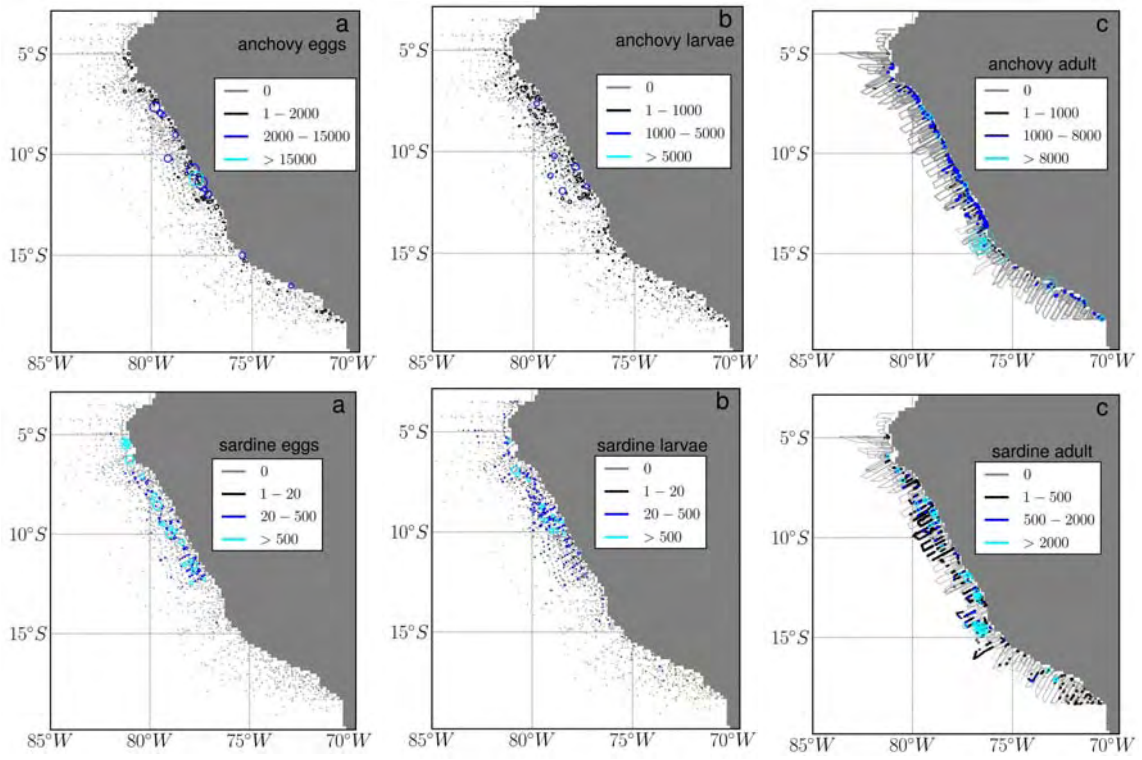

Figure 3.22: Composite distribution maps for eggs (a), larvae (b) collected by the Instituto del Mar del Peru combining all data over the different El Niño periods from 1961 to 2008 (a and b). Circles radius are proportional to density values with the higher biggest circle corresponding to (a) 18 966 (anchovy) and 5136 (sardine) $eggs.m^{-2}$, (b) 3414 (anchovy) and 3510 (sardine) $larvae.m^{-2}$ and (c) s_A=60 780 (anchovy) and s_A=58 057 (sardine) $nm.m^{-2}$. All data were provided by IMARPE.

From 1992 to 1999, the monthly mean gonado-somatic index (GSI) of anchovy (Fig. 3.23) provides information about the interannual variability in the reproduction of the species. Clearly, the seasonal cycle showed a minimum between March and June every year with one exception in 1995. Also, the anomaly of the GSI was minimum from Jul to Oct 1997, strongly increased until late 1997, then decreased to a minimum in Jan 1998 that probably lasted until March 1998 (in accordance to spawning fraction, Buitrón and Perea, 2000).

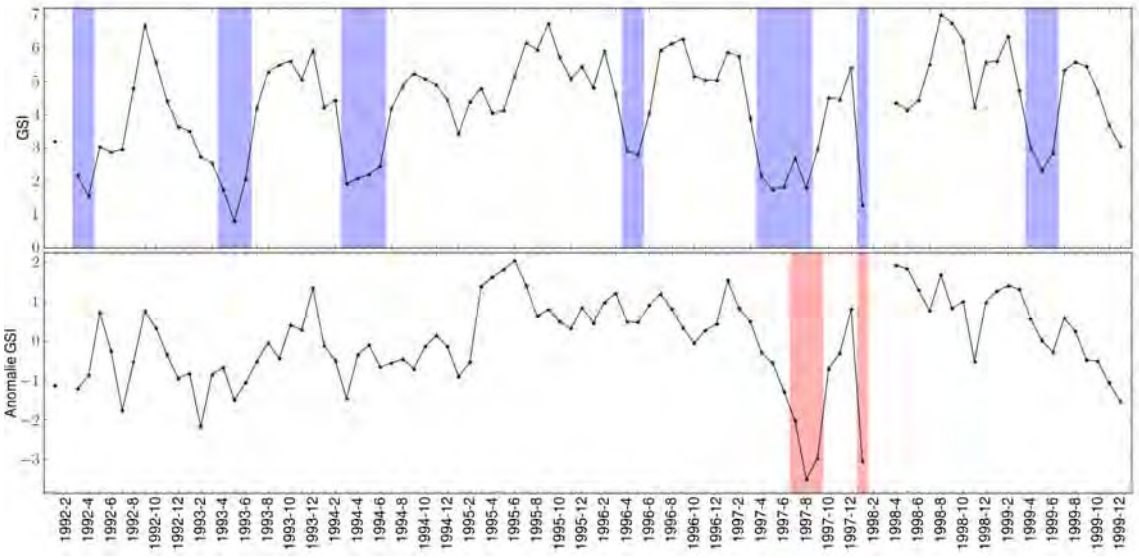

Figure 3.23: Anchovy monthly variability of gonado-Somatic Index between 1992 and 1999. Top: GSI value, Down: GSI's anomaly.

For sardine, we observed the same high anomaly of GSI from 07-1997 to 10-1997. In contrast with anchovy, GSI did not increase in November-December, decreasing until March 1998. A high anomaly was observed from 01-1998 to 03-1998, as observed for anchovy (Fig. 3.23). Sardine did not do better than anchovy during this "El Niño" event (Bertrand et al., 2004).

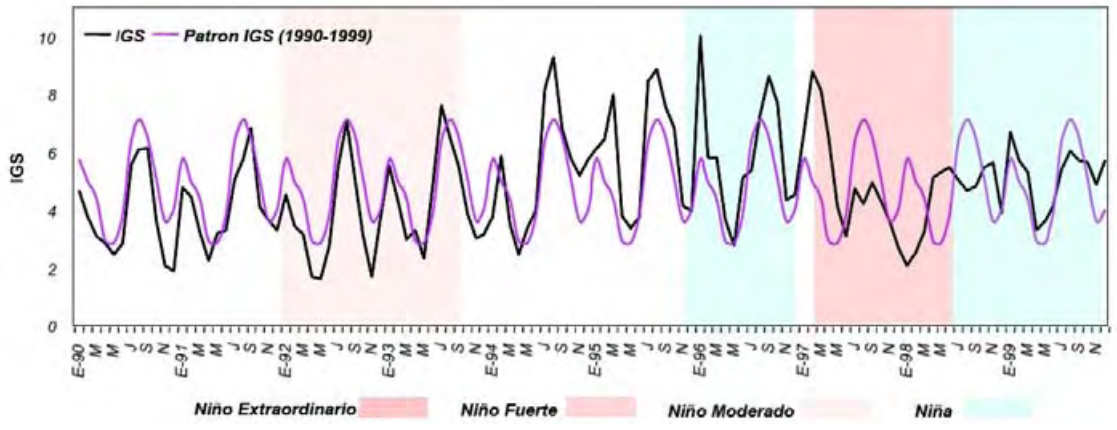

Figure 3.24: Sardine monthly variability of gonado-Somatic Index between 1990 and 1999 (black line) and decadal variability (purple line). From Cardenas (2009).

These results from GSI illustrate the fact that anchovy and sardine were affected in the same way by the 1998 El Niño event, showing bad spawning conditions for both species from 07-1997 to 10-1997 and from 01-1998 to 03-1998.

3.6 Decadal variability

The Pacific Decadal Oscillation (PDO) is a long-term ocean/atmosphere fluctuation of the Pacific Ocean. In the Humboldt Current system, the decadal warm and cold phases distinguish oneself by an anomaly of sea surface temperature of 1°C to 2°C. During colder periods, the intensity of the upwelling is stronger, thermocline is shallower. This conditions increase the productivity and nutrient supply. During warmer periods, opposite situation is observed with a weaker upwelling, a deeper thermocline and therefore a lower productivity.

These large-scales changes in Pacific ecosystem have been associated to periods of dominance of anchovy and sardines. In this system, phases with mainly negative anomalies were associated with anchovy dominance (1950-1970 and late 1990 to present), and phases with mainly positive anomalies were associated with sardines dominance (1970-late 1990) (Chavez et al., 2003). These periods have been described as "regime shifts" (Chavez et al., 2003; Alheit and Niquen, 2004). However, when looking further in the past, decadal fluctuations of anchovy and sardines did not always exist. Indeed, during the 400 years of the Little Ice Age (1400 to 1800s) neither anchovy nor sardine was abundant (Gutiérrez et al., 2009). According to time series from marine sediments, the penultimate period of high anchovy abundance actually began in the early 1900s (Fig. 1.17), but is well documented only from the 1950s, which corresponds to the beginning of industrial fisheries.

Data were processed according to the definition of PDO cold and warm periods. Since 1950, two cold regimes (1950-1971 and 1999-present) and one warm regime (1972-1998) were identified, with a transition period between 1993-1998 when anchovy population begun to increase and conversely sardine abundance started to decrease. Therefore, composite maps of anchovy and sardine observed densities were produced for 4 main periods: 1950-1971, 1972-1992, 1993-1998, 1999-2010 (Figs. 3.25 and 3.26).

The cold regime of 1950-1971 was characterized by high abundances of anchovy with the highest density concentrated along the coast between 6°S and 12°S, and no observation of sardine eggs or larvae. Between 1970 and 1973, marked by overfishing, the end of the cold regime and a rather strong El Niño event (at the end of 1972), the anchovy stock declined from 11 to 3 millions of tons.

In the same time, the sardines appeared and flourished during the warm PDO regime until 1992. Eggs and larvae were observed between latitudes 5°S and 12°S along the coast (Fig. 3.26) whereas adult fish were distributed still further south (5°S to 18°S) and mainly at the border of the continental shelf edge. Unlike anchovies, the areas with

maximum concentration of eggs or adults of sardines did not to coincide, suggesting that sardines would aggregate near the coast for spawning, likely to enhance survival rates of eggs and larvae, while dispersing offshore for feeding. During this warm PDO, eggs, larvae and adults of anchovy were observed only in a very narrow coastal band of a few kilometers (Fig. 3.25).

The warm to cold transition from 1993 to 1998 coincided with the decline of the sardine stock while the stock of anchovy was increasing and their eggs and larvae distribution extended further offshore. This extension was interrupted during the powerful El Niño event of 1997-1998 marked by a strong concentration of adults along the coast, with a particularly dense patch near 15°S (Figs. 3.25 and 3.26).

Since 1999, a new cold phase seems to have established. Anchovy stock is flourishing and its habitat extension has recovered the same area than observed before the anchovy collapse of the 1970s, both offshore and in latitudes south and north (Figs. 3.25 and 3.26). Anchovy eggs and adults were found from 5°S to 15°S (6°S to 12°S before the 70's), larvae were mainly located 6°S to 9°S, suggesting better conditions for larvae survival in this area. Though they did not totally disappear, sardines eggs and larvae became very sparse and were observed mainly offshore, thus possibly with less favorable conditions for survival (less food, higher advection?, more predators?). Investigating the relationship between dissolved oxygen concentration and biomass of anchovy and sardine, Bertrand et al. (2011) proposed to explain this spatial discrimination and the opposed population trajectories of the two species by the shallowing of the oxycline observed in the transition period after the early 1990s, preventing sardines to enter the water layer over the continental shelf (Fig. 1.22).

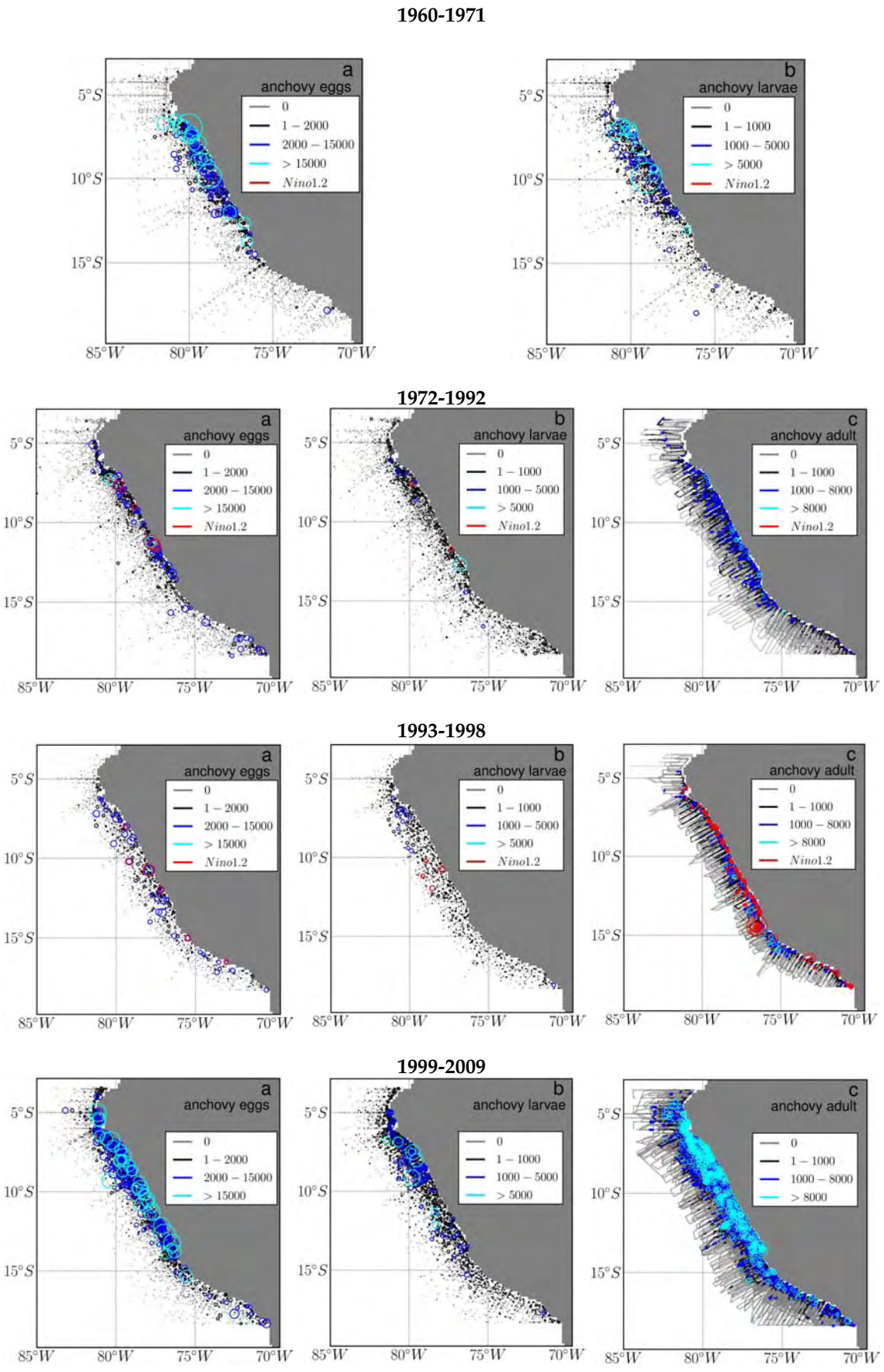

Figure 3.25: Composite distribution maps for eggs (a), larvae (b) and adult(c) of anchovy collected by the Instituto del Mar del Perú combining all data over the different decades. Circles radius are proportional to density values with the higher biggest circle corresponding to (a) 107376 $eggs.m^{-2}$, (b) 84939 $larvae.m^{-2}$ and (c) sA=438333 $nm.m^{-2}$. All data were provided by IMARPE (P. Ayón pers. com. and R. Castillo pers. com.)

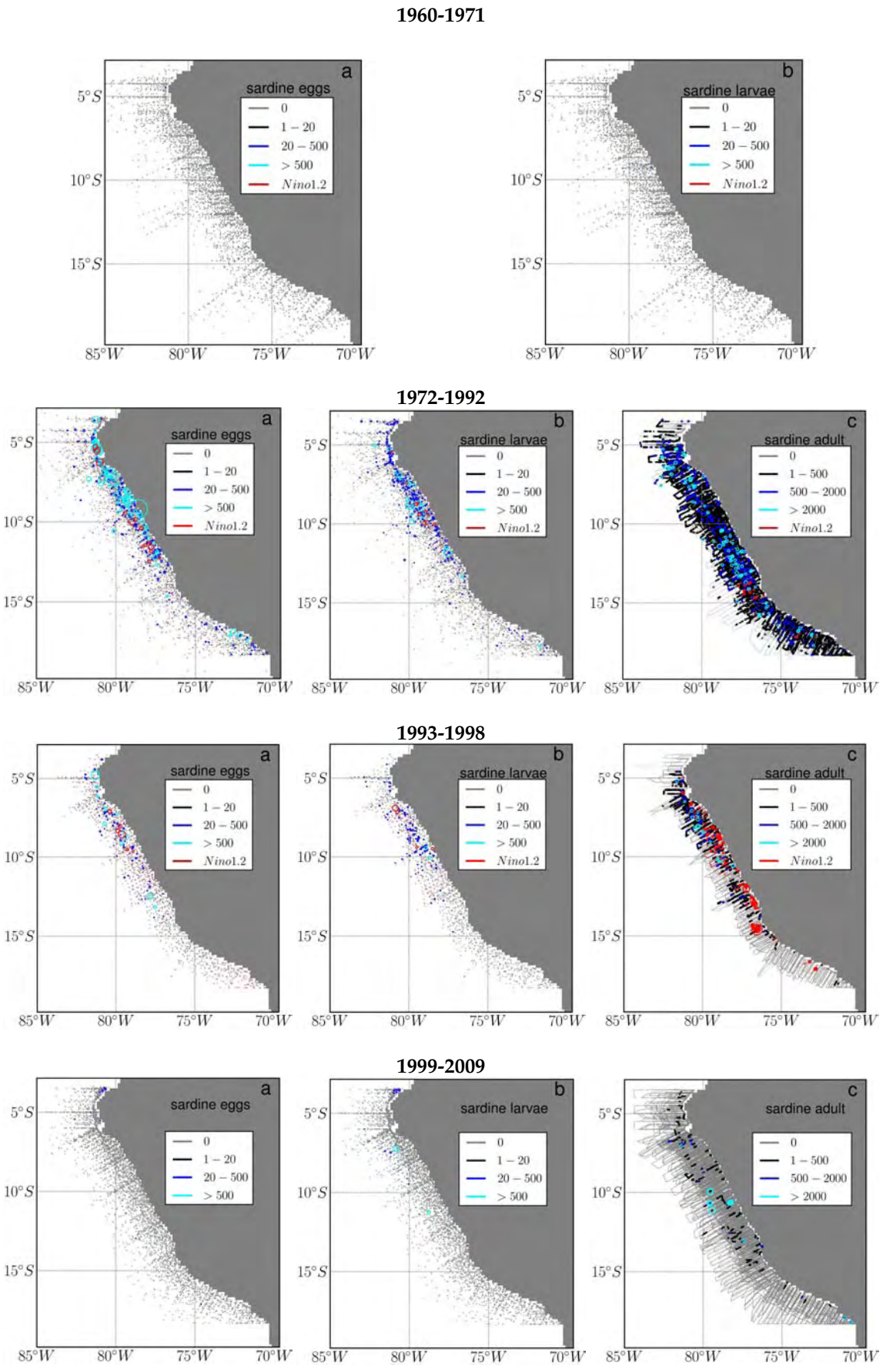

Figure 3.26: Composite distribution maps for eggs (a), larvae (b) and adult(c) of sardines collected by the Instituto del Mar del Perú combining all data over the different decades. Circles radius are proportional to density values with the higher biggest circle corresponding to (a) 63840 $eggs.m^{-2}$, (b) 36120 $larvae.m^{-2}$ and (c)s_A=23304 $nm.m^{-2}$. All data were provided by IMARPE (P. Ayón pers. com. and R. Castillo pers. com.)

3.7 Physical observations for eggs, larvae and adults

3.7.1 Effect of temperature

Many studies have focused on the link between temperature and anchovy distribution. For adults, Gutiérrez et al. (2007) (Fig. 3.27) showed that anchovy and sardine are broadly distributed over a large range of temperature: 14-23°C and 16-26°C respectively.

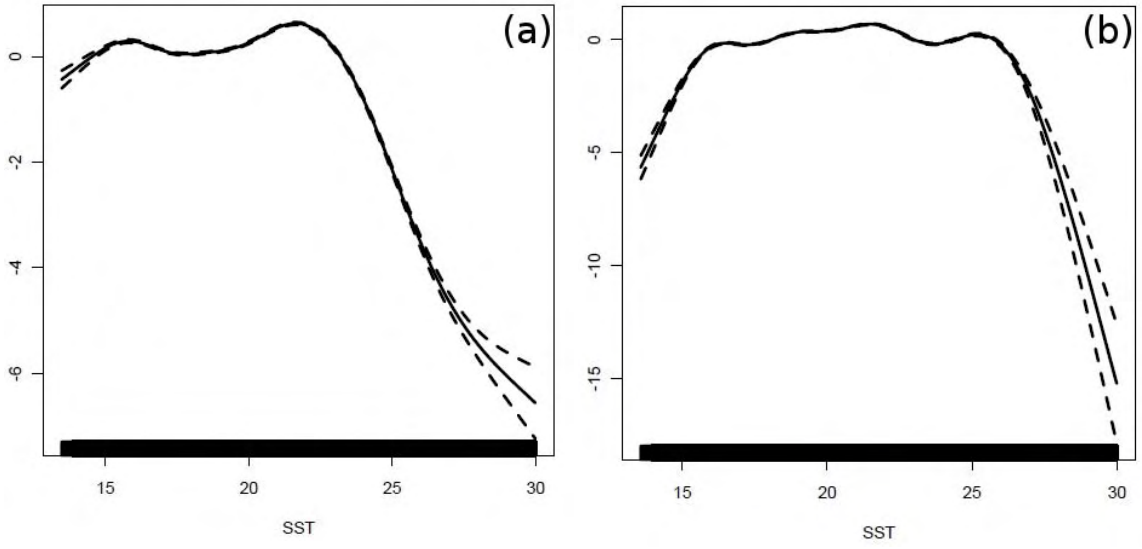

Figure 3.27: Cubic spline smoother fits (black solid lines) of GAMs of anchovy (a) and sardine (b) acoustic local biomasses (150 000 samples) according to sea surface temperature (SST in °C). Data from 1983 to 2005. The y-axes are relative and correspond to the spline smoother that was fitted to the data such that a y-value of zero is the mean effect of the variables on the response. Source: A. Bertrand pers.com.

Here, we focus on the effect of temperature on eggs and larvae. Measures of sea surface temperature (SST) in the HCS waters varied from 14°C to 27°C (from 13890 samples, Fig. 3.28 red curves), with a maximum probability at 18°C. Probability density function of the SST for all values where anchovy was present (black lines) still showed a main peak at 18°C, but also a decrease in the maximum temperature from 27°C to 25°C. Sardine present a high probability of presence at $\sim$ 18°C with a second peak at 22-24°C.

Probability density function for anchovy values higher than 1000 $Ind.m^{-2}$ for larvae and higher than 2000 $Ind.m^{-2}$ for eggs (see Fig. 3.28, blue lines) display a main peak at 18°C for eggs but two temperatures modes at 17°C and 21°C are observed for larvae. Indeed, these two modes correspond to spawning occurring in February-March (highest SST) and in August-September (colder SST) (see Figs. 3.19 and 3.30). The temperature range for anchovy varied therefore from 15°C to 25°C. In the same way, for sardine densities higher than 20 $Ind.m^{-2}$ presented a main peak at 18°C for eggs followed by a secondary peak at 22°C, whereas for larvae the main peak is at 22°C. The temperature

range for sardine therefore varied from 16°C to 27°C (Fig. 3.31).

Anchovy eggs densities higher than 15000 $Ind.m^{-2}$ and anchovy larvae densities higher than 5000 $Ind.m^{-2}$ (green lines), were only observed in September at the end of the winter (Fig. 3.30). Therefore the maximum probability of the SST is 17°C for eggs and 16.2°C, for larvae. However, the number of samples being lower than 60, these results are perhaps not significant. For sardines, probability density function of densities higher than 500 $Ind.m^{-2}$ show a high probability for SST values at 22°C (Fig. 3.28).

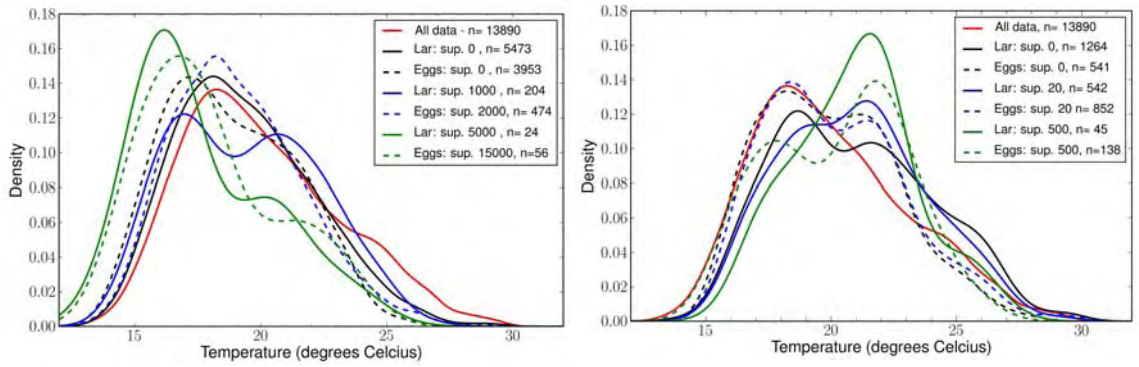

Figure 3.28: Probability density function for sea surface temperature (SST) for anchovy (at left) and sardine (at right) eggs and larvae densities.

To evaluate if a bias in SST measurement exist, we compared the SST intra-annual variability from climatologies made from: (1) SST satellite data, (2) the ROMS model, (3) data obtained with eggs and larvae sampling, and (4) data obtained with adults sampling. We made this average over the coastal regions and offshore regions using only the locations were eggs and larvae were sampled. Interestingly, the main seasonal cycle of these four datasets presents the same pattern. However, in situ data was globally colder by about 1°C all year around and the maximum peak of SST for eggs and larvae datasets occurred in January in contradiction with the 3 other datasets.

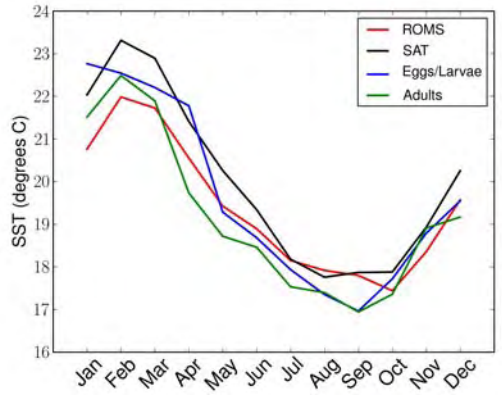

Figure 3.29: Comparison of SST satellite (black), SST from ROMS-PISCES model (red), SST measured from climatology made with eggs (blue) and larvae samples and SST measured from climatology of acoustical sampling (green)

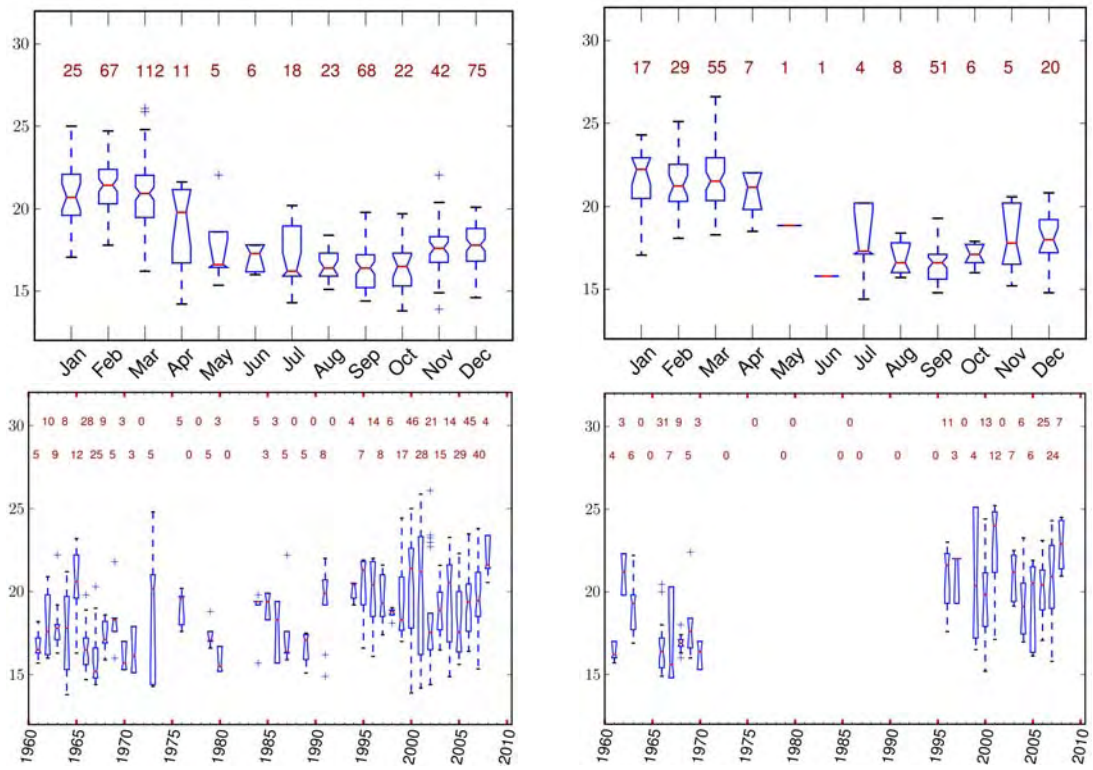

Figure 3.30: Bottom: Monthly mean sea surface temperature (SST) measured at anchovy eggs locations (at left) and anchovy larvae locations (at right) over the period 1961-2008. Annual mean SST measured at measured at eggs locations (at left) for densities higher than 2000 Ind/m^2 and larvae locations (at right) for densities higher than 1000 Ind/m^2 over the period 1961-2008. Numbers in red correspond to the number of observations monthly or annually.

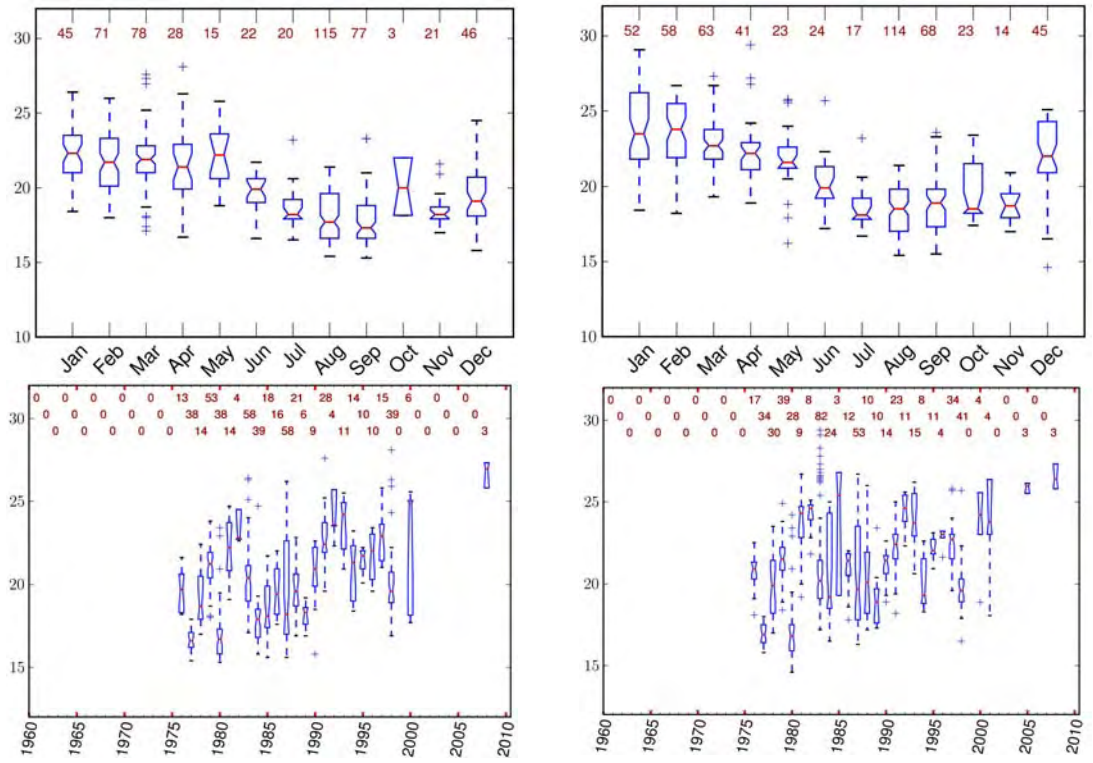

Figure 3.31: Bottom: Monthly mean sea surface temperature (SST) measured at sardine eggs locations (at left) and sardine larvae locations (at right) over the period 1961-2008. Annual mean SST measured at measured at eggs locations (at left) and larvae locations (at right) for densities higher than 20 Ind/m^2 over the period 1961-2010. Numbers in red correspond to the number of observations monthly or annually.

3.7.2 Vertical distribution of eggs and larvae

Vertical distribution for anchovies eggs and larvae have been estimated from 14 multinets profiles performed between 80°W to 81.5°W and 6°s to 8°S, in 2004 (Ayón, 2004) (Fig. 3.32) and 10 multinets profiles at 13°S (P. Ayón, pers. com.) (Fig. 3.33). Eggs were located mostly at the surface in both latitudes. The maximum larvae abundance was located between 40 and 60 m deep at 6-8°S and in the first 30 m at 13°S. As samples were collected in 2004, no sardines were present. The small secondary peak of eggs and larvae abundance deeper than 100 m (Fig 3.32) was assumed to be due to dead sinking eggs and larvae.

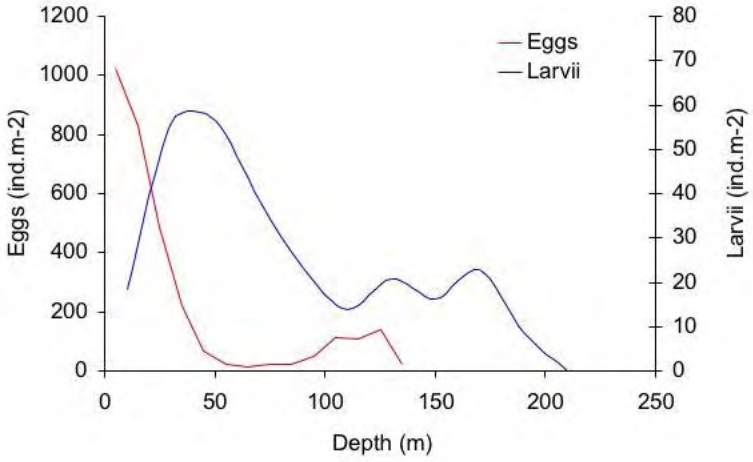

Figure 3.32: Average vertical distribution for anchovies eggs and larvae. From 14 multinets profiles between 80°W to 81.5W and 6°s to 8°S, in 2004 (Ayón, 2004).

Taking into account the currents vertical structure (see section 4.3.3.4), anchovy eggs were distributed within the first layer, where currents push them towards offshore (Fig. 4.13a). Anchovy eggs were not older than two or three days (P. Ayón, pers. com.), therefore, considering a maximum offshore transport of 14 $cm.s^{-1}$ (Fig. 4.14), the eggs should not be pushed further than 36 km off the coast, which is still within the favorable habitat.

The mean age of larvae sampled was estimated to be 5 days, corresponding to a size range between 3 and 6 mm (P. Ayón pers. com.). Larvae swimming speeds typically range around 1 and 2 body length per second (Bradbury et al., 2003), which gives velocities around 0.3 to 1.2 cm.s-1. They can therefore hardly swim against currents and are often considered as passive.

Larvae vertical distribution varied according to the latitude. Indeed, larvae were distributed in the first 30 m (Fig. 3.33) in central Peru but between 40 and 60 in northern Peru (Fig. 3.32).

If larvae is considered as passive as eggs and if larvae is considered to be located in the first 30 m, they will also be be advected offward over a maximum in the surface current. The 5 days old larvae may be advected offward over a maximum of 96 km, and

15 days old larvae, over up to 180 km, where survival conditions will be poor.

If the passive larvae assumption was true, larvae couldn't survive so far off the coast. Which are therefore, the mechanisms that can improve larvae survival conditions or retention?

Several mechanisms can contribute to larvae retention in coastal areas. First, mesoscale features which are not represented in Fig. 4.14 are able to create at the local scale coastward currents. Second, at 6-8°S, the main peak of larvae is observed at 40 to 60 m (Fig 3.32), i.e., at the limit of the nutrient rich current that push them to the coast where they can better survive (Fig. 4.13a). The presence of a larvae peak at this depth may be explained by an increase in the eggs density as eggs get older, which will allow larvae to return to the coast where conditions will improve their survive. Alternatively, larvae vertical swimming would produce the same effect. In both cases, larvae presence in the coastward current should allow them to be retained in coastal areas , where conditions are more favorable. Besides, relative small vertical migrations of organism can reduce the impact of the cross-shore transport of organism (Carr et al., 2008) relative to their passive transport at surface. IBM's simulation taking into account larval vertical swimming behavior have been shown to better match the data (Brochier et al., 2008).

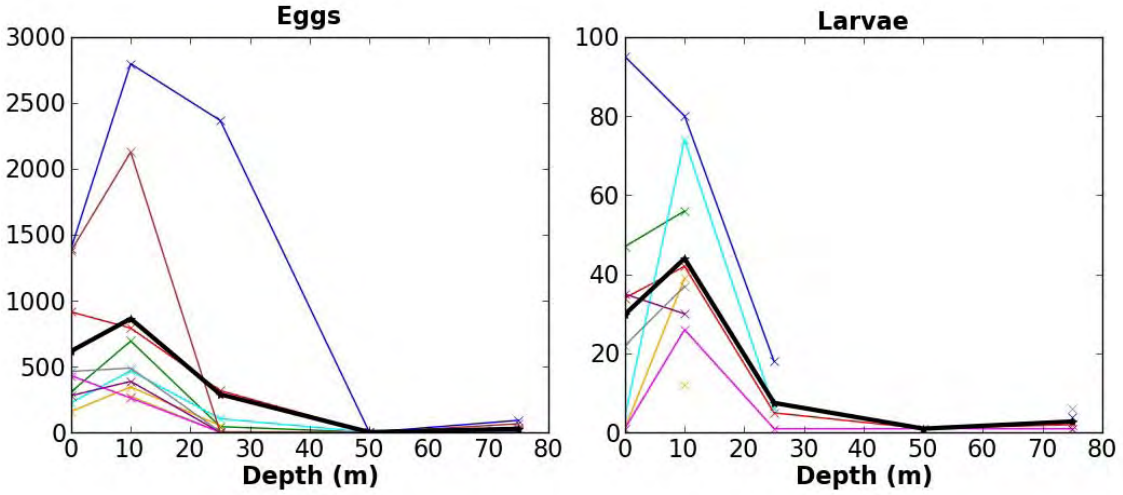

Figure 3.33: Vertical distribution for anchovies eggs and larvae. In color, data from 10 multinet around 12.5°S and 76.9°W, in 2004 (Data from P. Ayón pers. com.). In black, mean average over these 10 multinets.

Bertrand et al. (2011) showed the importance of the oxygen for anchovy and sardine habitat. Anchovy and sardine, as well as eggs and larvae, cannot survive deeper than the oxycline. As observed in Figs. 3.6 and 3.20, at 6-8°S, close to the coast, the oxycline limit is deeper than 40 m. At 13°S, it is much shallower, and no larvae were observed deeper than 30 m (3.33). Therefore in the northern part, the coastward current may be

accessible to larvae, whereas it is not the case in the southern region, where the oxycline is too shallow. This could explain why the northern part appears to be more favorable for larvae (3.15). It should be noted in addition that the oxycline is deeper when moving off the coast. Therefore, the coastward deeper current may become shallower than the oxycline and could therefore be more accessible to larvae off the coast. This provides another chance for larvae to return to more favorables areas.

3.7.3 Spatial co-occurrence of eggs and adults

Figure 3.8 shows that more eggs and larvae were present when abundance adults was higher. Here, we study the spatial co-occurrence of eggs and adults in the HCS.

Densities of anchovy and sardine eggs and adults sampled at the same time have been averaged at 1 degrees of resolution. We calculated for both species the percentage of observations from 1983 to 2008 with, in the same cell: eggs and adults, no eggs with adults, no adults with eggs, and no eggs and no adults (Table 3.4).

	No eggs and No adults	Adults without eggs	Eggs without adults	Both	N_{obs}
Anchovy (1983-2009)	34.8 %	16.2 %	5.3 %	43.7 %	1705
Anchovy (1983-1992)	44.7 %	19.3 %	5.7 %	30.3 %	452
Anchovy (1993-1999)	27.1 %	16.1 %	5.0 %	51.7 %	340
Anchovy (2000-2009)	32.9 %	14.7 %	5.1 %	47.3 %	913
Sardine (1983-1999)	39.1 %	40.7 %	2.6 %	17.5 %	869
Sardine (1983-1992)	19.0 %	53.4 %	2.5 %	25.7 %	470
Sardine (1993-1999)	63.6 %	25.8 %	2.8 %	7.8 %	399

Table 3.4: Repartition of eggs and larvae in observations from 1983 to 2008.

The percentage of cells with eggs but no adults varied from 2 to 5%. This is probably the error corresponding to dispersion of eggs by currents after spawning. The percentage of cells with anchovy adults without eggs was 18% and was quite stable during the study period whereas for sardine, this percentage varies from 26% to 53% (Table 3.4). When focusing on cells where eggs or/and adults were present 25% of the cells contained anchovy adults without eggs and around 70% of sardines adults without eggs (Table 3.5). The percentage of cells without eggs and larvae was higher during the period 1983 to 1992 as already described at the beginning of this chapter. When anchovy was highly abundant (1993-2000), presence of adults with eggs is about 52% of the data whereas on the same conditions for sardines (1983-1992), the presence of adults with eggs is only of 26%. This is most significant when looking the percentages of presence from the data where eggs or adults are present (Table 3.5). We observe that 71% of the data where adults and eggs are present contains both eggs and adults whereas for sardines it reached around 31%. In the same way, the number of adults without eggs reached 22%

for anchovy when for sardines it represented at least 65%.

These results corroborate the idea that for sardines spawning habitat and feeding habitat do not overlap whereas for anchovy it appears to be the case.

	Adults without eggs	Eggs without adults	Both	N_{obs}
Anchovy (1983-2009)	24.8 %	8.1 %	67.1 %	1111
Anchovy (1983-1992)	34.8 %	10.4%	54.8 %	250
Anchovy (1993-1999)	22.2 %	6.8 %	71.0 %	248
Anchovy (2000-2009)	21.9 %	7.7 %	70.5 %	613
Sardine (1983-1999)	66.9 %	4.3 %	28.7 %	529
Sardine (1983-1992)	65.4 %	3.1 %	31.5 %	384
Sardine (1993-1999)	71.0 %	7.8 %	21.4 %	145

Table 3.5: Repartition of eggs and adults within observation from 1983 to 2008 in which either eggs or adults have been observed

The stock - recruitment relationship is defined as the relationship between parental stock size (or spawning biomass) and the subsequent recruitment of young fish to the stock (exploited portion of the population). This approach was developed to integrate the missing processes of early life history in standard stock assessment models. Since SEAPODYM is spatial and includes all stages from eggs to adults, the link between spawning biomass and recruitment has been shifted at the eggs stage, assuming that there is a relationship similar to the Beverton-Holt function applying locally, at the level of the cell of the grid model, between spawning biomass and the level of surviving eggs that are recruited in the first population eggs cohort. In the tuna version of SEAPODYM (Lehodey et al., 2008), this relationship is shifted at the larvae stage, since no eggs cohort exist.

Therefore, here we are interested in looking if such a correlation between the densities of eggs and adults exists. Taking into account only the samples where eggs and adults were observed and averaging them at one degree of resolution, a significant correlation of 0.11 was found for anchovy whereas it was not detectable for sardine. This can be due to a difficulty to sample both eggs and mature sardines together due to the displacement of adult between spawning and feeding habitats.

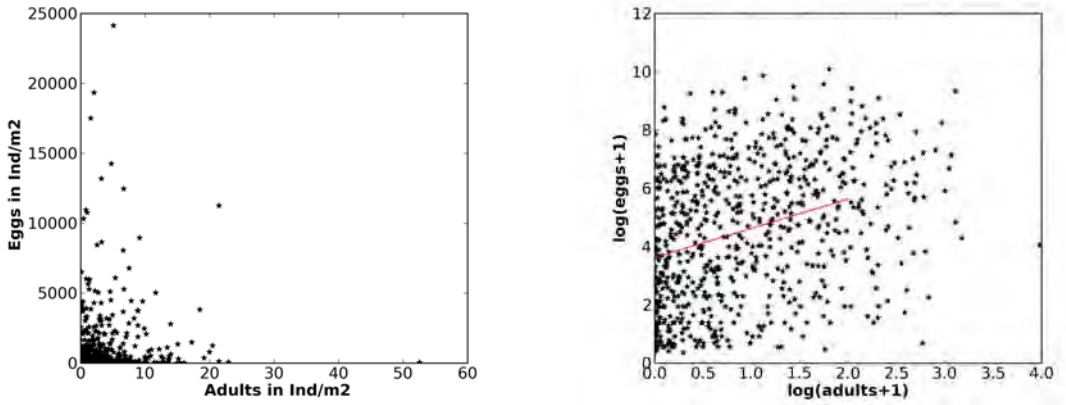

Figure 3.34: Example of the spatial correlation between anchovy eggs and anchovy adults, for all positive data from 1983 to 2008 averaged at one cell degrees. At left, densities in number of individuals; At right, densities in logarithms.

3.8 Predators of anchovy and sardines

As mentioned by Cury et al. (2000), in the four EBUS, natural mortality of small pelagic fish is much higher than fishing mortality. In the Benguela Current system, 55% of small pelagic fish are estimated to be eaten by top predators (Cury et al., 2000). The predation effect appears therefore to have an important role in fish population dynamics and should be considered when modelling the early life stages of anchovy and sardine species.

In addition to anchovy and sardine, others species such as jack mackerel (*Trachurus murphyi*), mackerel (*Scomber japonicus*), jumbo squid (Dosidicus gigas), squat lobster (*Pleuroncodes monodon*), Peruvian Sea catfish (*Galeichthys peruvianus*) and longnose anchovy (*Anchoa nasu*) inhabits the peruvian waters. Jack mackerel and jumbo squid are potentially strong predators for anchovy eggs, larvae and adults (Alheit and Niquen, 2004) in the offshore part of their distribution. In the inshore part, mackerel, squat lobster, anchovy and sardine are others potential predators (Fig. 3.35). Other main predators of pelagic fish in the HCS are sea birds and mammals (Bertrand et al., 2004) but data are scarce (see Jahncke et al., 2004).

Furthermore, offshore, the mesopelagic fish community, consuming mainly zooplankton and fish larvae (Rodolfo and Rolf, 2006), are also potential predators for eggs and larvae of small pelagic fish in the offshore part of their distribution.
This group is dominated by the families phosichtyidae (*Vinciguerria lucetia*), and myctophidae which represent 60.4% and 12.8% of the total mesopelagic scientific catch, respectively (Rodolfo and Rolf, 2006).

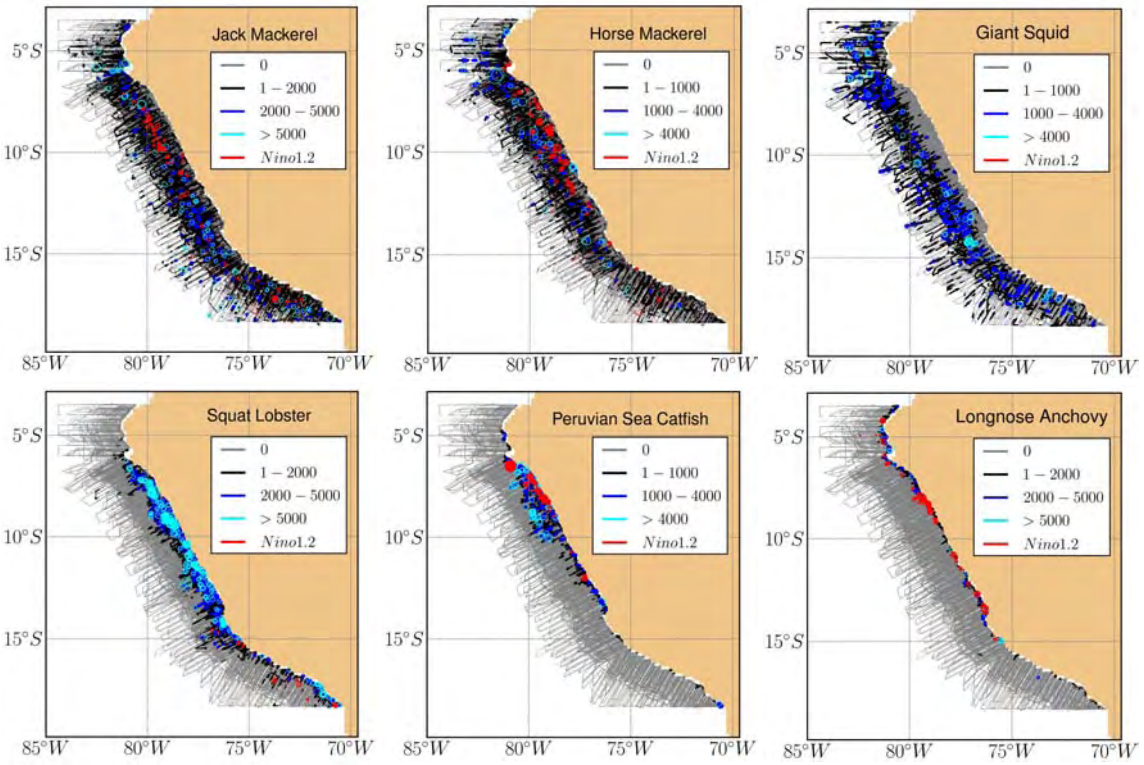

Figure 3.35: Composite distribution maps for jack mackerel (Jurel) (a), mackerel (Caballa) (b), jumbo squid (Pota) (c), squat lobster (Munida) (d), Peruvian Sea catfish (Bagre) (e), longnose anchovy (Samasa) (f) collected by the Instituto del Mar del Peru combining all acoustic data over the period 1983-2009 . Circles radius are proportional to density values with the higher biggest circle corresponding to 53150 $m^2.nm^{-2}$ (a), 53700 $m^2.nm^{-2}$ (b), 15784 $m^2.nm^{-2}$ (c), 83008 $m^2.nm^{-2}$ (d), 30000 $m^2.nm^{-2}$ (e), 43701 $m^2.nm^{-2}$ (f).

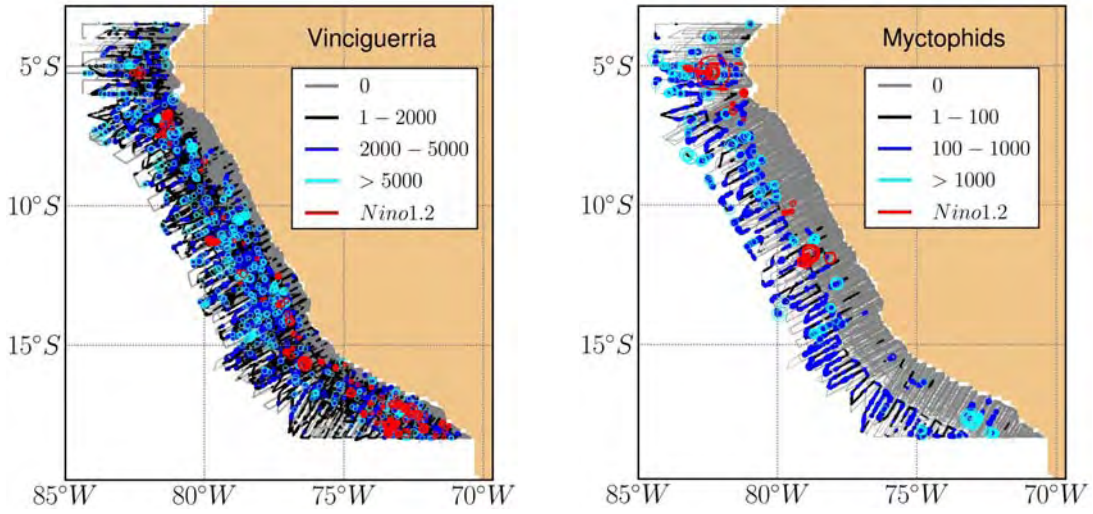

Figure 3.36: Composite distribution maps for mesopelagic fish collected by the Instituto del Mar del Peru over the period 1983-2009. At right for Vinciguerria lucetia, at left for myctophiids. Circles radius are proportional to density values with the higher biggest circle corresponding to 50000 $m^2.nm^{-2}$ and 30000 $m^2.nm^{-2}$ respectively.

Acoustic data shows that mesopelagic fish are mainly distributed offshore, off the continental shelf edge. *V. lucetia* is mainly located at the shelf break while the myctophiids are located more offshore (Fig. 3.36).

Mesopelagic fish densities have been estimated to 20 $g.m^{-2}$ in spring 2006 (Gutiérrez et al., 2009) and to 26 $g.m^{-2}$ in a cruise in 2004 (Gutiérrez et al., 2009). This gives us an order of values of what estimation of mesopelagic fish should be predicted. Since Vinciguerria represents around 60% of the mesopelagic fish, the estimation of biomass of total mesopelagic fish should be higher than the estimated here.

For the same cruises, anchovy densities where estimated to 41 $g.m^{-2}$ and 89 $g.m^{-2}$, respectively.

3.9 Conclusion

Long-term records of eggs, larvae and adults data allowed describing the main seasonal and spatial patterns for anchovy and sardine. To summarize, the main observed spatial patterns are listed below.

Seasonal eggs, larvae and adults variability of anchovy are characterized by:

- **Spatially**, a negative coast-offshore gradient, a pronounced negative north-south gradient for larvae, though less pronounced for eggs and not observed for adults.
- **Temporally**, for eggs and larvae, a seasonal pattern with a clear resting period between April and June, a peak of eggs and larvae abundance around September, and the presence of relatively high densities between December and March (see also Bouchón et al., 2000; Perea et al., 2011). For adults, high abundance is observed in February and March, but as discuss in section 2.4.1, this pattern can be biased, since more aggregated fishes could result in artificially higher observed biomass.

Seasonal eggs, larvae and adults variability of sardines are characterized by:

- **Spatially**, for eggs and larvae, a decreasing coast-offshore density, with high abundances from 6°S to 14°S. For adults, maximum adult densities are situated offshore at the break of the continental shelf edge, except during El Niño events, when sardines are located more along the coast.
- **Temporally**, for eggs and larvae, a seasonal pattern showing lowest abundance in October and November and June-July. Maximum abundance appears in August-September, as for anchovy and in January. For adults, seasonal pattern show a high abundance peak in April and May, perhaps overestimate in May since observational effort is the lowest at that period.

Eggs and larvae from anchovy and sardines appear to be distributed along the coast, as for anchovy adults, whereas adult sardines locate mostly offshore. Therefore, eggs and larvae of anchovy appear to share the same habitat as adults whereas eggs and larvae of sardines appear to share different habitats. Finally, modelling early lifes stages for anchovy should be more straightforward than for sardines. To understand sardines populations dynamics, it appears necessary to model sardines adults dynamics and displacement between spawning habitat and feeding habitat. However, sardine data is perhaps too scarce to can conclude because since 1999 no sardines are present in the HCS water. Furthermore, historically, IMARPE put perhaps less effort in that species.

At interannual scale, anchovy population increased at the end of the 90's and reached high levels of abundances during this last decade. For sardines, we observe a decrease of abundances after 1992 and a collapse of this species at the end of the 90's. Therefore without an interannual physical model that could begin in the 1980's, where high levels of sardines where still observed, it would be difficult to model the characteristics of this species.

For the optimization process of the early life stages parameters, we will use eggs and larvae climatological data to find the parameters that best explain the observed anchovy and sardine seasonal patterns. However, we should always have in mind the different bias that may affect the climatology. For the different interannual runs, raw data will be used. Despite the high heterogeneity of observations, we assume that data contain enough information to constrain our model with the optimization approach developed in chapter 6, and capture the low frequency signal.

Eggs and larvae data appear to follow a Negative Binomial Distribution, whereas for adult distribution, no good fit was obtained with classical distributions. Nevertheless, as will be explained in Chapter 6, the choice of the cost function does not depend only on the distribution which fit best our data, but also of what we want to fit with our model and which error we want to consider in our data.

A more "exhaustive" analysis could be done with these datasets, for instance to properly quantify the uncertainties in the observations. Thus it is possible that the climatologies used in the following chapters include biases. Spatial interpolation methods as "krigging" could be applied to help reducing the high variability observed at the local scale. It would result in smoothed data that may improve the optimization procedure.

Bibliography

Alheit, J. and M. Niquen, 2004: Regime shifts in the humboldt current ecosystem. *Progress in oceanography*, **60**(2-4), 201–222.

Ayón, P., 2004: Informe a.e.p.s. crucero demersales 1994-2004. *Inf. Técnico - Imarpe*.

Bertrand, A., A. Chaigneau, S. Peraltilla, J. Ledesma, M. Graco, F. Monetti, and F. Chavez, 2011: Oxygen, a fundamental property regulating pelagic ecosystem structure in the coastal southeastern tropical pacific. *PlosOne*, **6**(12), e29558.

Bertrand, A., F. Gerlotto, S. Bertrand, M. Gutiérrez, L. Alza, A. Chipollini, E. Diaz, P. Espinoza, J. Ledesma, and R. Quesquon, 2008: Schooling behaviour and environmental forcing in relation to anchoveta distribution: An analysis across multiple spatial scales. *Progress in Oceanography*, **79**(2-4), 264–277.

Bertrand, A., M. Segura, M. Gutiérrez, and L. Vasquez, 2004: From small-scale habitat loopholes to decadal cycles: a habitat-based hypothesis explaining fluctuation in pelagic fish populations off peru. *Fish and Fisheries*, **5**(4), 296–316.

Bertrand, S., E. Diaz, and M. Lengaigne, 2008: Patterns in the spatial distribution of Peruvian anchovy (*Engraulis ringens*) revealed by spatially explicit fishing data. *Progress in Oceanography*, **79**(2-4), 379–389.

Bouchón, M., P. Ayón, J. Mori, C. Peña, P. Espinoza, and L. H. L. et al., 2000: Peruvian anchovy (*Engraulis ringens Jenyns*) Biology. *Bol. Inst. Mar Perú*, **25**.

Bradbury, I. R., P. V. R. Snelgrove, and P. Pepin, 2003: Passive and active behavioural contributions to patchiness and spatial pattern during the early life history of marine fishes. *Marine Ecology-Progress Series*, **257**, 233–245.

Brochier, T., C. Lett, J. Tam, P. Fréon, F. Colas, and P. Ayon, 2008: An individual-based model study of anchovy early life history in the northern humboldt current system. *Progress in Oceanography*, **79**(2-4), 313–325.

Buitrón, B. and A. Perea, 2000: Aspectos reproductivos de la anchoveta peruana durante el periodo 1992-2000. *Bol. Inst. Mar Perú*, **19**, 45–54.

Cardenas, S., 2009: *Analisis de series de tiempo de los indicadores biologicos, pesqueros y poblacionales de la sardina, Sardinops Sagax (Jenys, 1842), en funcion de la variabilidad ambiental y la pesca*. Universidad Nacional Mayor de San Marcos.

Carr, S., X. CAPET, J. Mcwilliams, J. Pennington, and F. Chavez, 2008: The influence of diel vertical migration on zooplankton transport and recruitment in an upwelling region: Estimates from a coupled behavioral-physical model. *Fisheries Oceanography*, **17**(1), 1–15.

Castillo, P. R., S. Peraltilla, A. Aliaga, M. Flores, M. Ballón, J. Calderón, and M. Gutiérrez, 2009: Protocolo técnico para la evaluación acústica de las areas de distribucion y abundancia de recursos pelagicos en el mar peruano, version 2009. *Inf. Inst. Mar Perú*, **36**, 7–27.

Chavez, F. P., J. Ryan, S. E. Lluch-Cota, and M. Niquen C, 2003: From anchovies to sardines and back: multidecadal change in the Pacific Ocean. *Science*, **299**(5604), 217–221.

Cury, P., A. Bakun, R. Crawford, A. Jarre, R. Quiñones, L. Shannon, and H. Verheye, 2000: Small pelagics in upwelling systems: patterns of interaction and structural changes in "wasp-waist" ecosystems. *ICES Journal of Marine Science*, **57**(3), 603–618.

Foote, K. and I. C. f. t. E. o. t. Sea, 1987: *Calibration of acoustic instruments for fish density estimation: a practical guide*. Number 144. International Council for the Exploration of the Sea.

Gerlotto, F., S. Bertrand, N. Bez, and M. Gutiérrez, 2006: Waves of agitation inside anchovy schools observed with multibeam sonar: a way to transmit information in response to predation. *ICES Journal of Marine Science*, **63**(8), 1405.

Gutiérrez, D., A. Sifeddine, D. B. Field, L. Ortlieb, G. Vargas, F. P. Chavez, F. Velazco, V. Ferreira, P. Tapia, R. Salvatteci, et al., 2009: Rapid reorganization in ocean biogeochemistry off Peru towards the end of the Little Ice Age. *Biogeosciences*, **6**(5), 835–848.

Gutiérrez, M., P. Castillo, S. Peraltilla, M. Flores, C. Maldonado, and L. Vasquez, 2009: Distribution and biomass of some peruvian pelagic resources in summer 2004. *Inf. Inst. Mar Perú*, **36**, 145–152.

Gutiérrez, M., S. Peratlilla, and L. Vasquez, 2009: Distribution and abundance of peruvian pelagic and mesopelagic species; spring 2006. *Inf. Inst. Mar Perú*, **35**, 65–80.

Gutiérrez, M., G. Swartzman, A. Bertrand, and S. Bertrand, 2007: Anchovy (*Engraulis ringens*) and sardine (*Sardinops sagax*) spatial dynamics and aggregation patterns in the Humboldt Current ecosystem, Peru, from 1983-2003. *Fisheries Oceanography*, **16**(2), 155–168.

Jahncke, J., D. M. Checkley, and G. L. Hunt, 2004: Trends in carbon flux to seabirds in the Peruvian upwelling system: effects of wind and fisheries on population regulation. *Fisheries Oceanography*, **13**, 208–223.

Lehodey, P., I. Senina, and R. Murtugudde, 2008: A spatial ecosystem and populations dynamics model (SEAPODYM) - Modeling of tuna and tuna-like populations. *Progress in Oceanography*, **78**(4), 304–318.

Lett, C., P. Penven, P. Ayón, and P. Fréon, 2007: Enrichment, concentration and retention processes in relation to anchovy (*Engraulis ringens*) eggs and larvae distributions in the northern Humboldt upwelling ecosystem. *Journal of Marine Systems*, **64**(1-4), 189–200.

Lewin, W., J. Freyhof, V. Huckstorf, T. Mehner, and C. Wolter, 2010: When no catches matter: Coping with zeros in environmental assessments. *Ecological Indicators*, **10**(3), 572–583.

MacKenzie, D., J. Nichols, G. Lachman, S. Droege, J. Andrew Royle, and C. Langtimm, 2002: Estimating site occupancy rates when detection probabilities are less than one. *Ecology*, **83**(8), 2248–2255.

Maclennan, D., P. Fernandes, and J. Dalen, 2002: A consistent approach to definitions and symbols in fisheries acoustics. *ICES Journal of Marine Science: Journal du Conseil*, **59**(2), 365–69.

Mangel, M. and P. E. Smith, 1990: Presence-absence sampling for fisheries management. *Canadian Journal of Fisheries and Aquatic Sciences*, **47**(10), 1875–1887.

Martin, T. G., B. A. Wintle, J. R. Rhodes, and P. M. e. a. Kuhnert, 2005: Zero tolerance ecology: improving ecological inference by modelling the source of zero observations. *Ecology Letters*, **8**(11), 1235–1246.

Marzloff, M., Y. J. Shin, J. Tam, M. Travers, and A. Bertrand, 2009: Trophic structure of the peruvian marine ecosystem in 2000-2006: Insights on the effects of management scenarios for the hake fishery using the IBM trophic model osmose. *Journal of Marine Systems*, **75**(1-2), 290–304.

Minami, M., C. Lennert-Cody, W. Gao, and M. Roman-Verdesoto, 2007: Modeling shark bycatch: the zero-inflated negative binomial regression model with smoothing. *Fisheries Research*, **84**(2), 210–221.

Misund, O., J. Coetzee, P. Fréon, M. Gardener, K. Olsen, I. Svellingen, and I. Hampton, 2003: Schooling behaviour of sardine *Sardinops sagax* in False bay, South Africa. *African Journal of Marine Science*, **25**(1), 185–193.

Parrish, R., C. Nelson, and A. Bakun, 1981: Transport mechanisms and reproductive success of fishes in the california current. *Biolog. Oceanogr.*, **2**, 175–203.

Perea, A., C. Peña, R. Oliveros-Ramos, B. Buitrón, and J. Mori, 2011: Potencial production of eggs, recruitment and reproductive season fishery closure of Anchoveta (*Engraulis ringens Jenyns*, 1842): implications for fisheries management. *Ciencias Marinas*, **37**(42).

Rodolfo, C. and K. Rolf, 2006: Distribution patterns of mesopelagic fishes with special reference to vinciguerria lucetia garman 1899 (Phosichthyidae: pisces) in the humboldt current region off peru. *Marine Biology*, **149**(6), 1519–1537.

Simmonds, E. J., M. Gutiérrez, A. Chipollini, F. Gerlotto, M. Woillez, and A. Bertrand, 2009: Optimizing the design of acoustic surveys of Peruvian anchoveta. *(ICES) J. Mar. Sci.*, **66**(6), 1341–1348.

Simmonds, J. and D. MacLennan, 2005: *Fisheries Acoustics: theory and practice*. Blackwell Science, 7-27.

Swartzman, G., A. Bertrand, M. Gutiérrez, S. Bertrand, and L. Vasquez, 2008: The relationship of anchovy and sardine to water masses in the Peruvian Humboldt Current System from 1983 to 2005. *Progress in Oceanography*, **79**(2-4), 228–237.

Vaudor, L., N. Lamouroux, and J. Olivier, 2011: Comparing distribution models for small samples of overdispersed counts of freshwater fish. *Acta Oecologica*, **37**, 170–178.

Warton, D., 2005: Many zeros does not mean zero inflation: comparing the goodness-of-fit of parametric models to multivariate abundance data. *Environmetrics*, **16**(3), 275–289.

Chapter 4

Environmental forcing of Seapodym Model

Sommaire

4.1 Introduction . . . 104
4.2 Satellite and *in situ* observations . . . 105
4.2.1 Sea Surface Temperature (SST) . . . 105
4.2.2 Chlorophyll and Primary Productivity . . . 105
4.2.3 Oxygen . . . 107
4.2.4 Currents . . . 109
4.3 ROMS-PISCES model outputs . . . 111
4.3.1 The model . . . 111
4.3.2 Vertical habitat definition . . . 112
4.3.3 Climatological Run . . . 113
4.3.3.1 Configuration . . . 113
4.3.3.2 Euphotic depth and mixed layer depth . . . 113
4.3.3.3 Oxygen . . . 114
4.3.3.4 Currents . . . 117
4.3.3.5 Temperature, Primary Productivity . . . 119
4.3.4 Interannual Run 1992-2000 . . . 123
4.3.5 Interannual Run 2000-2006 . . . 126
4.4 Eggs and larvae predators: the micronekton model . . . 129
4.5 Conclusion . . . 133
Bibliography . . . 134

4.1 Introduction

This chapter describes the environmental forcing used for the SEAPODYM application to anchovy and sardine. It includes temperature, currents, dissolved oxygen and primary production fields simulated by the coupled ROMS-PISCES models. These variables are then used to compute the micronekton functional groups from which larvae predator fields are deduced. Three simulations were provided through the ANR PEPS (Peru Ecosystem Projections Scenarios) project: a climatological run and an interannual run covering the period 1992 to 2000 at spatial resolution of 1/6°, and an interannual run for 2000 - 2006 at resolution 1/9°. Despite the large improvement in ocean physical modelling at basin scale during the last decade, the simulations at high resolution and in coastal regions can have biases and missing processes. **Since these simulations outputs are the forcing of the SEAPODYM application we want to develop, it is first critical to evaluate the quality of these predicted physical and biological fields.**

Satellite and *in situ* observations are used for this evaluation of the ROMS-PISCES inputs. Then, we present the ROMS-PISCES model and its predictions used to drive the SEAPODYM model in the Peruvian region, characterized by a very narrow continental shelf. This particular configuration is due to the subduction of the oceanic Nazca Plate under the Southern American continental Plate, that is at the origin of the Andes mountains (altitude > 6 km) and the Peru-Chile trench (depth >7 km). The 200 m isoline used to delineate the continental shelf break highlights the narrowest of the continental shelf edge (Fig. 4.1), reaching a maximum of 100 km in the northern Peruvian area between 7°S and 10°S.

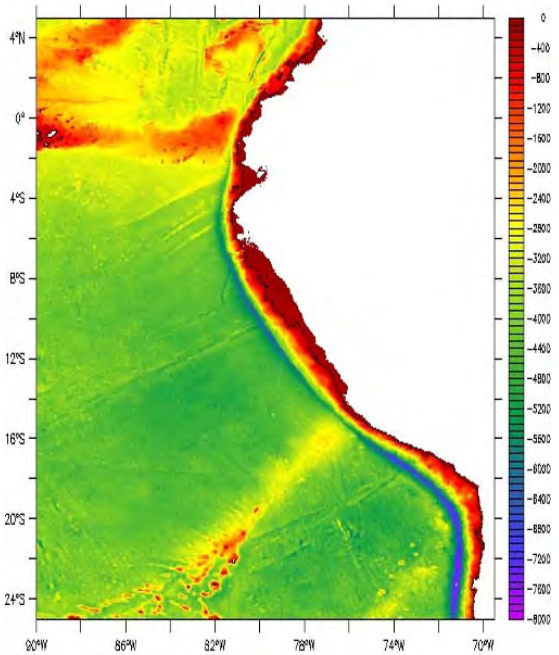

Figure 4.1: Bathymetry of the Peruvian waters. Data from Amante and Eakins (2009). Black line shows the continental shelf limit (200 m bottom depth).

4.2 Satellite and *in situ* observations

4.2.1 Sea Surface Temperature (SST)

Using the NOAA Optimum Interpolation 1/4 Degree Daily SST product (Reynolds et al., 2007), a SST climatology was produced to be compared with the ROMS outputs, with data covering the same period 1992 - 2000 and excluding the 1997-1998 El Niño years.

4.2.2 Chlorophyll and Primary Productivity

For chlorophyll and derived primary production fields, cloud cover and satellite orbital geometry limit the number of observations available in the Peruvian area (Fig. 4.2). The coverage is particularly poor in austral winter. During the period 2000-2006, less than 5 % of the days between June and October were observed. Furthermore, the algorithm used to calculate sea surface chlorophyll tends to systematically underestimate the high concentrations of coastal zones (O'Reilly et al., 1998). Therefore, primary production patterns issued from satellite data in this area need to be considered with caution.

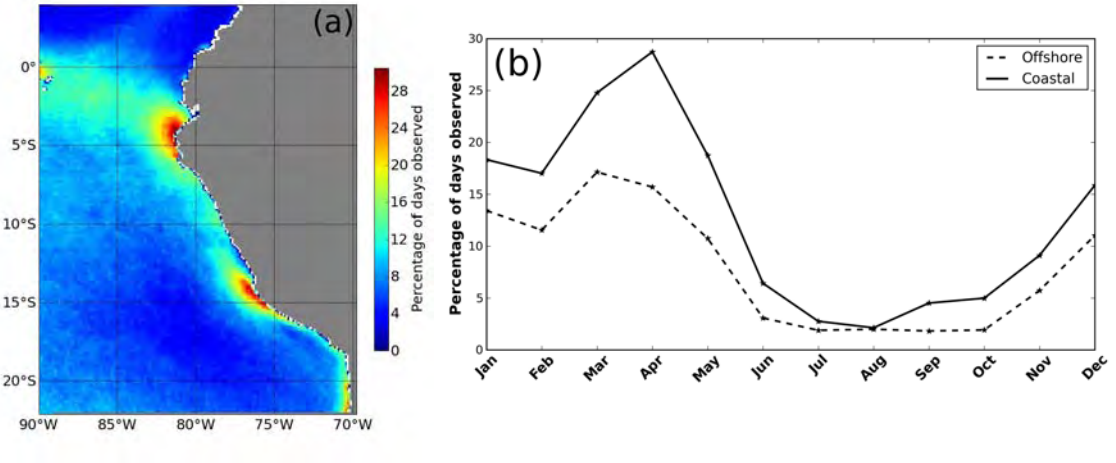

Figure 4.2: (a) Spatial Distribution of the temporal effective coverage of SEAWIFS satellite during the 2000-2006 period. (b): Percentage of observed days by the SEAWIFS satellite averaged over the 2000-2006 period for the 3 coastal regions (NCZ, CCZ, SCZ regions) and the 2 offshore regions (OZ and TZ regions).

Echevin et al. (2008) compared the annual average chlorophyll concentration from satellite and *in situ* data (Fig. 4.3). Both datasets display typical spatial patterns for coastal upwelling zone, with high concentrations of chlorophyll nearshore that decrease gradually offshore. Though these overall spatial and seasonal patterns are consistent with each other (Echevin et al., 2008), here the analysis focus on the smaller region inhabited by anchovy and sardines comprised between 6°S to 15°S. This region includes the richest chlorophyll concentration nearshore, but distribution of maximal values are

shifted between *in situ* data (8°S to 14°S) and satellite data (11°S to 14°S) (Fig. 4.3a,b). Both datasets show the same spatial patterns along the coast north to 6°S, but differ south of 15°S, where *in situ* chlorophyll concentration is higher than satellite chlorophyll.

Seasonality deduced from satellite data shows the highest chlorophyll concentration in austral summer with a decrease during austral winter (Fig. 4.3 and introduction). However, since the seasonal chlorophyll cycle is highly correlated to observational effort (Fig. 4.2b), the low data coverage in austral winter may be a strong source of bias (Fig. 4.2b).

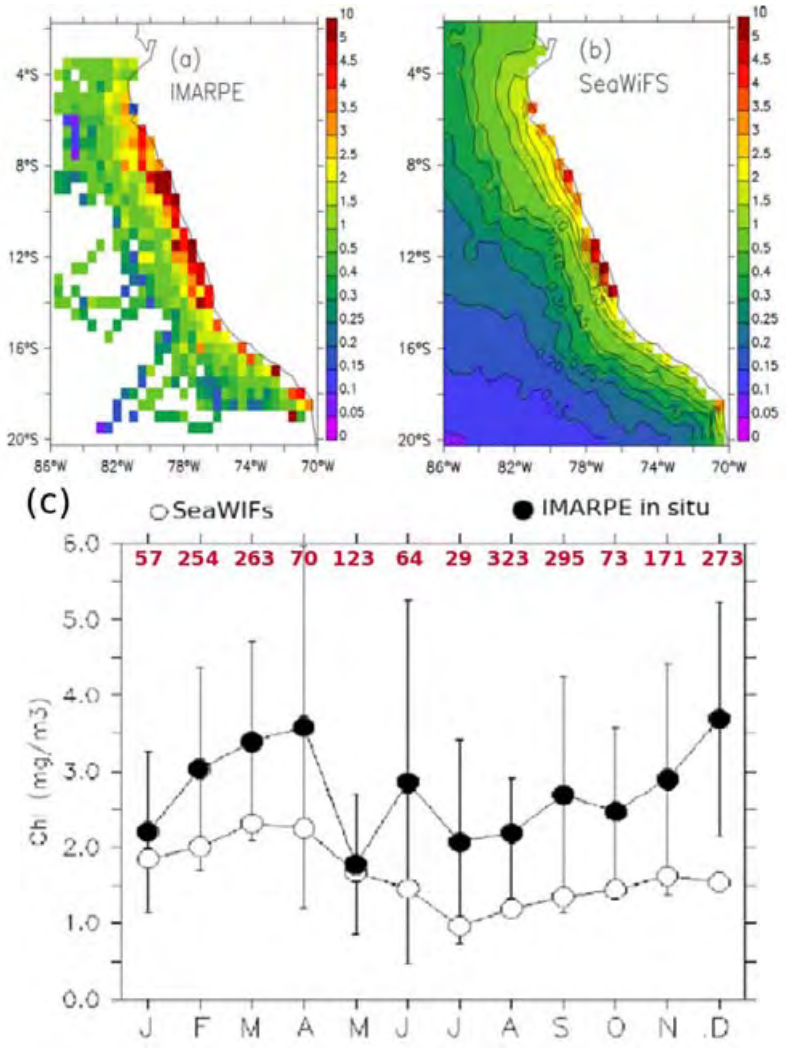

Figure 4.3: (a,b) Annual average surface chlorophyll (in $mgChl.m^{-3}$) at 0.5°x0.5° resolution: (a) IMARPE *in situ* data (1992-2004), (b) SeaWIFs data over the years 1997-2004; (c) Seasonal cycle of the surface chlorophyll concentration (in mgChl/m3) averaged over a coastal box of 250 km zonal width between 4°S and 15°S: SeaWIFs (white circles), IMARPE *in situ* data (black circles). Vertical bars correspond to the interannual variability of the *in situ* data set. In red, number of grid points averaged to compute the monthly mean. Adapted from Echevin et al. (2008).

Overall, IMARPE *in situ* data display higher values than satellite data but with the same high peak during austral summer. However spatial and temporal heterogeneity of the *in situ* sampling is high. February, March, August, September and December are the months with the highest number of observations and therefore the ones with the best spatial representativeness.

Primary productivity (PP) data is calculated from Seawifs/MODIS satellite chlorophyll data, temperature and SeaWiFS photosynthetically active radiation (PAR) (Behrenfeld and Falkowski, 1997). Derived primary production dataset is based on the VGPM algorithm used for the general case of oceanic waters while coastal waters would require a specific parametrization (Behrenfeld and Falkowski, 1997). Furthermore, cloud - filling interpolation techniques performed on satellite chlorophyll data can introduce additional biases. Albert et al. (2010) suggested that these two sources of uncertainty may result in an overestimation of primary productivity from VGPM algorithm. **This has to be considered when comparing these in situ and satellite observation to the ROMS-PISCES predictions.**

4.2.3 Oxygen

A climatology of oxygen concentration was built based on oxygen profiles from 1961 to 2008 obtained from the IMARPE and the World Ocean Database (WOD09, Garcia et al., 2010) at 1/10° resolution ($\sim$ 11 km) (IMARPE, unpublished data, 2010). From now on, we will call this climatology the **Peru climatology**. Following the work of Bertrand et al. (2010), the depth of the upper Oxygen minimum zone (OMZ) is defined by the 0.8 $mL.L^{-1}$ dissolved oxygen concentration.

Surface oxygen concentrations are lowest nearshore and increase gradually offshore (Fig. 4.4). Nearshore, minimum surface oxygen concentration occurs between 6°S and 9°S (Fig. 4.4a). In this area, the oxygen concentration is always higher than the oxycline limit as shown in Fig. 4.4b (in white). South of 9°S, the oxycline is very shallow reaching 20-30 m depth. Oxycline depth estimated from acoustic measurements during a campaign in 2005 (Fig. 4.5) shows similar meridional patterns: a shallow oxycline between 17°S to 8°S and deeper oxycline north of 8°S (Fuenzalida et al., 2009; Bertrand et al., 2010). The latitudinal delimitation of this north-south gradient can also vary in time (A. Bertrand pers. com.)

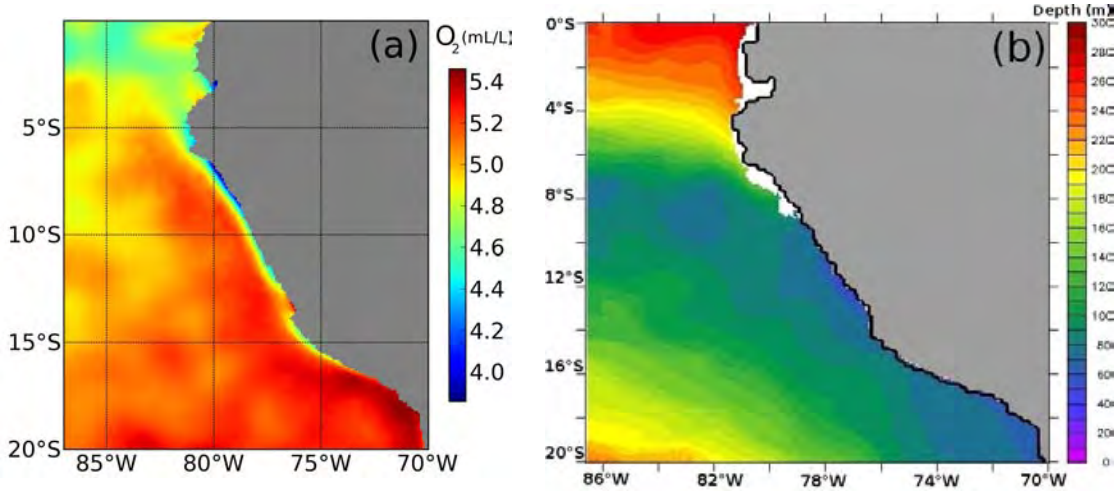

Figure 4.4: (a) Annual averaged spatial distribution of oxygen at surface. (b) Upper OMZ depth estimated from the 0.8 $mL.L^{-1}$ oxygen concentration limit in the Peruvian waters. From Peru climatology (IMARPE, unpublished data, 2010).

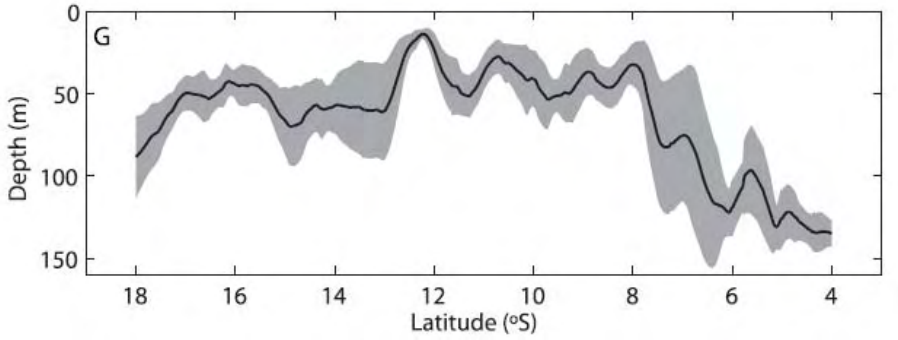

Figure 4.5: Meridional variation of the vertical extension of the epipelagic community (Z_{VEEC}) (equivalent to the oxycline), estimated from acoustic measurements during the 'Pelagic 2005' survey. The bold line is the average of the Z_{VEEC} between the coast and 200 km offshore and the grey shade area corresponds to one standard deviation. From Bertrand et al. (2010).

The seasonality of the oxygen concentration in the coastal regions displays between May and August a minimum in the first 30m, and a maximum between June and September below (Fig. 4.6).

Offshore the seasonal pattern is very different, with maximum concentrations from August to December, and a minimum from March to June. As the depth increases, the oxygen minimum and maximum occur earlier in the year.

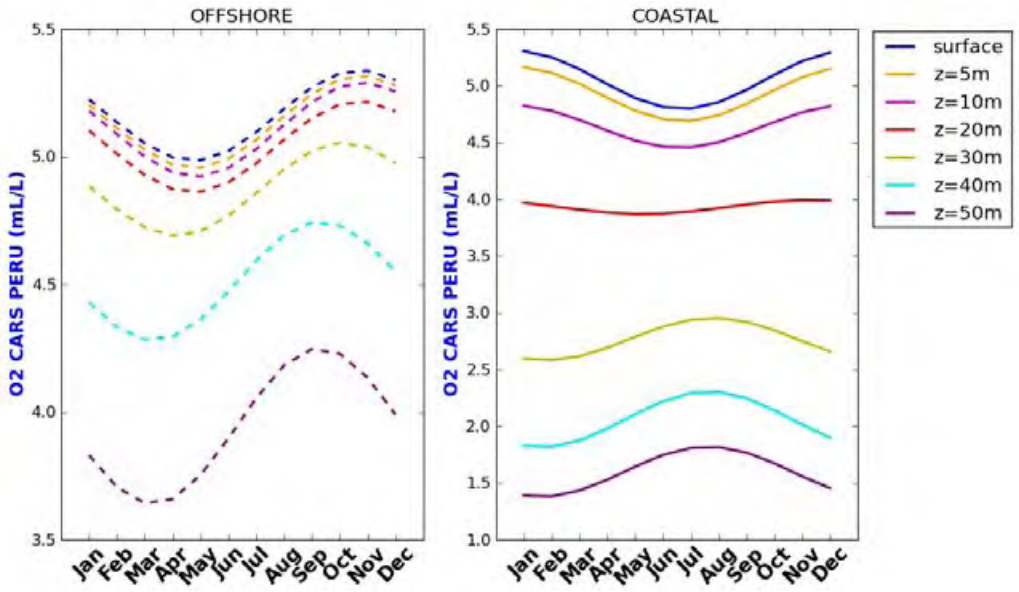

Figure 4.6: Oxygen seasonal variability in the Peruvian coastal regions (solid lines) and offshore regions (dashed lines). Peru climatology from IMARPE, unpublished data, 2010.

4.2.4 Currents

Sea surface currents observations are available from two different datasets: *in situ* drifting buoys and geostrophic currents inferred from altimetry. Fig. 4.7a displays the total number of observations and the average of velocity measurements from satellite-tracked 15m drogue drifting buoys in 18 km boxes (1/6° resolution as ROMS-PISCES climatological model) over the period 1993-2010. There are less than 15 observations per box along the Peruvian coast, meaning that the average value is poorly representative of the mean currents over this period. Drifting buoys where collected in the framework of the international Global Drifter Program and are distributed by the Atlantic Oceanographic and Meteorological Laboratory (AOML) where they have been first quality controlled and krigged (Hansen and Poulain, 1996).

Figure 4.7b displays the average of currents from the SURCOUF satellite product at 1/3° resolution from 2004 to 2010. SURCOUF currents derive from a combination of altimeter geostrophic currents and Ekman currents. The Ekman currents are computed applying a simple 2 parameter model (Rio and Hernandez, 2003, Rio et al., 2011) to wind stress data from the ERA INTERIM reanalysis. This 2 parameter model has a seasonal and latitudinal dependency but does not depend on the longitude. This means that it may not be fully appropriate for our study area. An Ekman model fitted for the study area would certainly results in more accurate surface currents. On the other hand, altimeter data are known to be less accurate in coastal areas, where the ocean dynamics is characterized by smaller spatial scales and where the different corrections applied to the altimeter measurement may suffer from the vicinity of the coast. In particular, at a distance of about

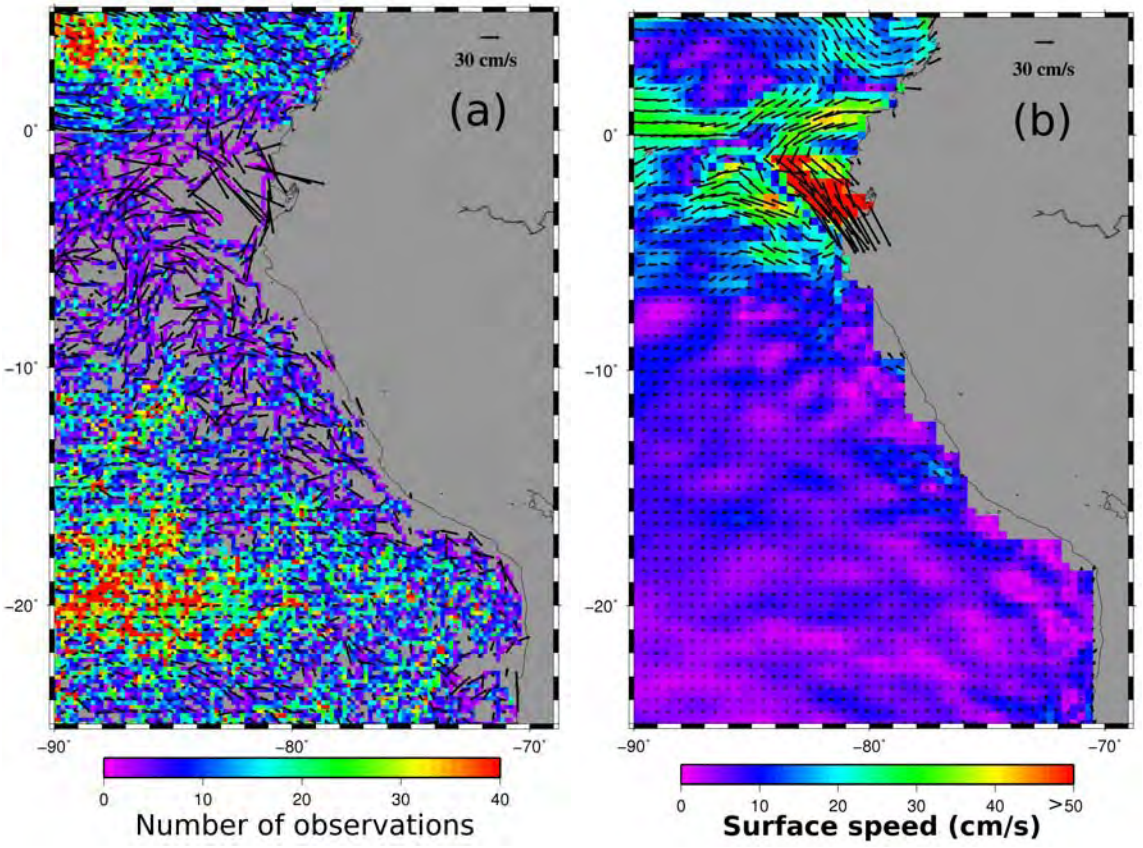

Figure 4.7: (a) Number of drifting buoys used in 18 km boxes; Arrows: Mean currents from drifting buoys dataset between 1993 and 2010. (b) Mean surface currents from SURCOUF analysis at 1/3° resolution from 2004 to 2010. Grey areas correspond to zero data.

50 km from the coast, the radiometer signal used to correct the altimeter range from the propagation changes due to the presence of water vapor in the atmosphere is strongly polluted by the presence of continents. Finally, coastal areas are characterized by a number of dynamical processes other than geostrophic and Ekman's related; in particular the tidal currents (but not important in our region), as well as other non-linear ageostrophic phenomena not resolved by altimetry.

These two datasets from *in situ* and satellite measurements and satellite data illustrate the lack of data and accuracy near the coastal area and particularly in the first 200 km offshore inhabited by anchovy and sardines. Therefore, validation of nearshore circulation of ROMS-PISCES model by comparison with these datasets is impossible and thus will be made in accordance to literature knowledge of circulation in an upwelling system.

4.3 ROMS-PISCES model outputs

4.3.1 The model

Physical and biological forcing fields (temperature, currents, O_2, Primary Production (PP), and euphotic depth) were predicted by the Regional Ocean Modelling System (ROMS, Shchepetkin and McWilliams, 2005) coupled to a biological model (PISCES, Aumont, 2003; Aumont and Bopp, 2006; Fig. 4.8). The model configuration for the Peruvian region has been described in previous publications (Penven et al., 2005; Echevin et al., 2008; Albert et al., 2010).

PISCES (Pelagic Interaction Scheme for Carbon and Ecosystem Studies) is a NPZD[1] type model which simulates the biological productivity and describes the biochemical cycle of the main nutrients (nitrate, ammonium, phosphate, silicate and iron), the carbon cycle and the dissolved oxygen (Fig. 4.8).

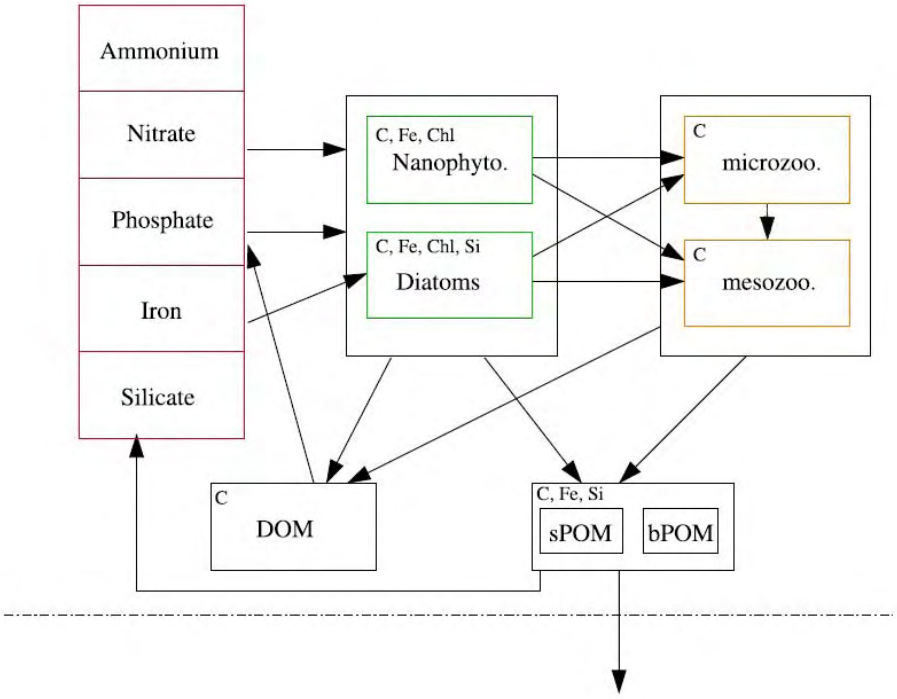

Figure 4.8: Schematic description of the PISCES ecosystem model from Aumont and Bopp (2006). In red, nutrients; in green, phytoplankton; in orange, zooplankton; in black, detritus. The detritus can be in a semilabile dissolved organic matter form (DOM) or small (sPOM) and big sinking particles form (bPOM).

[1]The most simple ecosystem model is a NPZD model, which is divided in 4 boxes: one box of Nutrients, one box of Phytoplankton, one box of Zooplankton and one box of Detritus. The N5P2Z2D2 is therefore composed with 5 classes of nutrients, 2 classes of phytoplankton, 2 classes of zooplankton and 2 classes of detritus.

Phytoplankton growth can be limited by the five nutrients modeled. The model has two phytoplankton size classes/groups (nanophytoplankton and diatoms) and two zooplankton size classes (microzooplankton and mesozooplankton) (Aumont and Bopp, 2006). Diatoms differ from nanophytoplankton by their need in Si, by higher requirements in Fe and by higher half-saturation constants because of their larger mean size (Aumont and Bopp, 2006).

ROMS-PISCES variables are distributed in a three-dimensional Arakawa C-type grid, where active/passive tracers are located at the center of the cell and the horizontal velocity (u and v) are located at the west/east and south/north edges of the cell, respectively. Predicted environmental variables of ROMS-PISCES input are first interpolated by spline method on a regular grid (with all variables located at the center of the grid) to be used with the SEAPODYM model. [2]

4.3.2 Vertical habitat definition

The modelling of eggs, larvae and adults dynamics of anchovy and sardine in SEAPODYM takes place in a single vertical layer. Therefore physical (temperature, currents) and biological (O_2) variables are averaged over this predefined layer that needs to be defined carefully according to physical and biological characteristics.

For the modeling of the micronekton functional groups, the first (epipelagic) layer is defined by the euphotic depth (Zeu), ie., the upper layer receiving sufficient light for photosynthesis. Two other deeper layers are used to represent deeper organisms of the micronekton: the mesopelagic layer between one and three Zeu, and the bathypelagic layer between 3 Zeu and 1000 m. Total primary production is integrated vertically and used as a proxy for larvae's food.

Sardines and anchovies are typically epipelagic species. Their vertical habitat can be thus proposed to be defined by the euphotic layer as in the micronekton model. The bottom of this layer is generally defined as the depth where the sunlight has attenuated to 1 percent of the surface value ("euphotic depth"). This depth varies geographically and seasonally, and can range from a few meters (in highly productive waters such as an upwelling area), to around 200 m in very clear tropical waters. However, in the particular case of the Peru upwelling, the epipelagic layer seems to be controlled by the upper oxygen minimum zone (Bertrand et al., 2010).

[2] Note: At our resolution, the results between the different possible methods (linear, spline interpolation) of interpolation ROMS-PISCES variables into a regular grid will be always smaller than the approximation resulting from the vertical average made on the mixed layer depth or the epipelagic layer.

For eggs and larvae, the vertical habitat could be more limited, e.g., to the mixed-layer, i.e. the surface layer characterized by an homogeneous temperature. This hypothesis would imply that eggs and larvae spatial dynamics are mainly driven by surface current (Lett et al., 2007; Mathiesen, 1989).

The skills of the ROMS-PISCES model to simulate these different vertical layers need to be evaluated. In particular, the vertical averaging over the mixed-layer or the euphotic layer could produce large differences of spatial dynamics that require a detailed comparative analysis to understand the consequences and the results of the simulations.

4.3.3 Climatological Run

4.3.3.1 Configuration

The model domain covers the Humboldt Current system in the region 5N-25°S and 90W-69.5°W, at a spatial resolution of 1/6° (∼ 18 km) over 30 depth layers. The surface layer thickness ranges from a minimum of 5 cm in a 50-m deep nearshore water column, to 4 m in a 4000-m deep water column and the bottom layer thickness ranges from 10 m to 730 m. Five-day average model outputs are recorded.

We used a climatological simulation forced by COADS climatological heat fluxes and a monthly climatology of Quikscat wind stress, as in Albert et al., 2010. At the open boundaries, the model is forced by the dynamical fields and biogeochemical tracers from a monthly climatology of the ORCA2 OGCM simulation at 2° resolution over the period 1992-2000 (Albert et al., 2010). Ten years of spinup were produced to reach a statistical equilibrium. While the eddy-resolving model is forced repeatedly (each year) with the same monthly climatological atmospheric fluxes and boundary conditions, the mesoscale environment is variable for the different years due to intrinsic nonlinear variability.

It should be noted that the model climatology can be partly biased because it was created using an atmospheric forcing for the period 2000-2006, but the boundary forcing corresponds to the period 1992-2000.

4.3.3.2 Euphotic depth and mixed layer depth

Both the mixed layer depth and the euphotic depth are shallowest nearshore and increase gradually offshore to approximatively 60 and 100 meter depth respectively (Figs. 4.9a, 4.9b).

A marked seasonality is observed for the mixed layer depth, which varies throughout the year from 10 m to 40 m in coastal regions and from 20 m to 60 m in offshore

regions (Fig. 4.9c). Conversely, the euphotic depth does not show a significant seasonal variability, with a mean depth of 30 m in coastal areas and 60 m in offshore areas.

Unfortunately, no observations were available to validate these predicted variables.

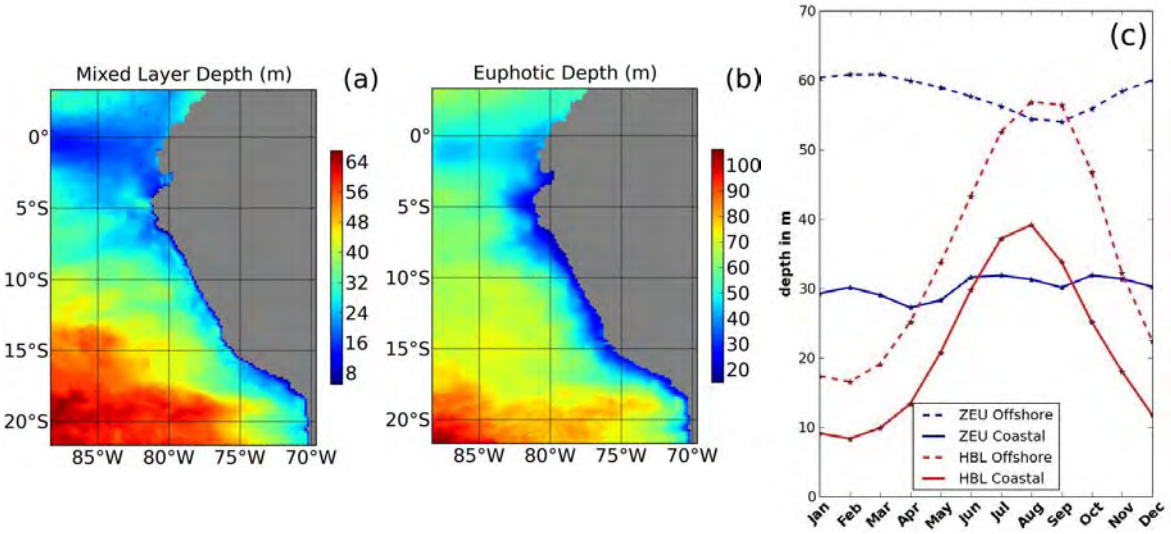

Figure 4.9: Annual mean over the 10 climatological years of (a) mixed layer depth (HBL) and (b) euphotic depth (ZEU) predicted by the ROMS-PISCES model, and (c) predicted seasonal variability of euphotic depth and mixed layer depth in the Peruvian coastal regions and the offshore regions.

4.3.3.3 Oxygen

Based on acoustical data, the epipelagic layer was defined in the Humboldt current according to the depth of the upper oxygen minimum zone, which corresponds to the lower oxycline depth ($\sim$ 0.8 $mL.L^{-1}$ oxygen concentration, Bertrand et al., 2010). Furthermore, oxygen has been proposed to be the "missing link" (Bertrand et al., 2011) which explains spatial dynamics of anchovy and sardines. Oxygen concentrations predicted by the ROMS-PISCES model were analyzed in order to verify if oxygen can be used for spawning habitat definition and/or can be used as the variable defining the epipelagic layer.

The annual mean of surface oxygen concentration, O_2concentration averaged over the mixed layer depth, and O2 concentration averaged over the euphotic layer are compared (Fig. 4.10). The three variables display lower oxygen concentrations nearshore which increase gradually offshore, as observed in the Peru climatology. However, the minimum surface oxygen concentration nearshore between 6°S and 9°S and the deeper oxycline in the same area (Fig. 4.11 versus Fig.4.4) are not observed.

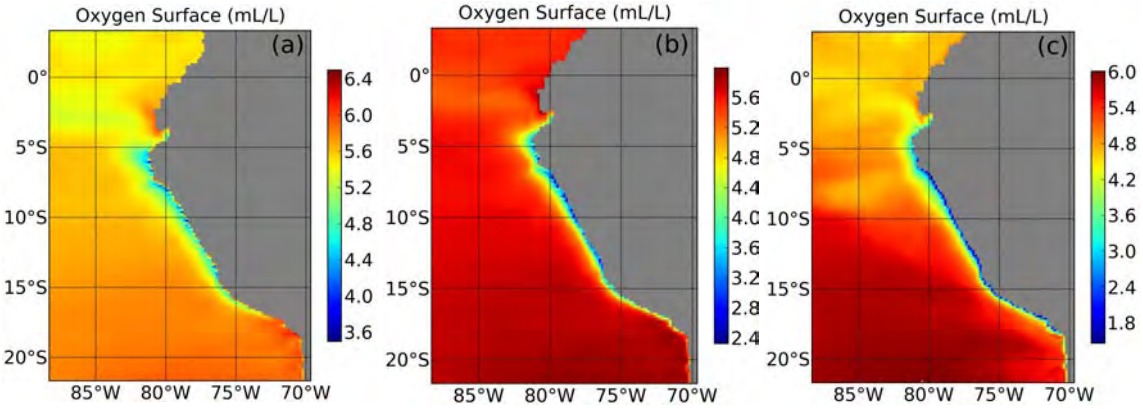

Figure 4.10: Annual mean of oxygen concentration over the 10 climatological years of ROMS-PISCES. a) Surface oxygen concentration; b) O_2 concentration averaged the mixed layer layer; c) O_2 concentration averaged over the euphotic layer.

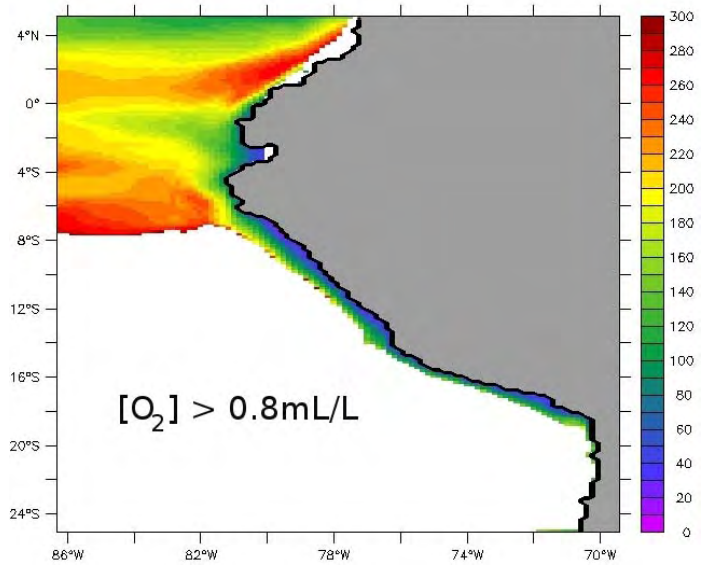

Figure 4.11: Upper OMZ depth estimated from the 0.8 $mL.L^{-1}$ volume of oxygen from ROMS-PISCES model. In white, oxygen concentration higher than 0.8 $mL.L^{-1}$.

The typical meridional pattern observed in the Peru climatology with a shallow oxycline between 17°S and 8-9°S, and a deeper oxycline north of 8-9°S (see section 4.2.3) is not well reproduced by ROMS-PISCES[3]. Furthermore, the 0.8 $mL.L^{-1}$ limit is not represented everywhere, because predicted offshore concentrations are always higher than 0.8 $mL.L^{-1}$ in contrast with the observations.

Though the predicted temporal variations at surface match closely the Peru climatology with a minimum from May to August, ROMS-PISCES oxygen concentrations have a stronger seasonal variability than observed in the Peru climatology. However, when averaging on thicker layers, the minimum is predicted to occur from April to July using the

[3]ROMS-PISCES oxygen concentration are given in μmol.L^{-1}. To convert in $mL.L^{-1}$ in order to compare with oxygen measurements, we use the ideal gas law: $V = nRT/P$, with T the predicted ROMS-PISCES temperature, n the number of moles of oxygen in the volume V (predicted by the model) and R the gas constant. The pressure P is fixed to 1 bar, since measurements are made at the surface.

mixed layer depth, and from April to May using the euphotic depth (Fig. 4.16). The validation of these predictions remains difficult. Nevertheless, the seasonal timing of oxygen minimum computed from the mixed layer depth and the euphotic depth seems realistic since it correlates nicely with the minimum observed for eggs, larvae and gonadosomatic index of anchovy.

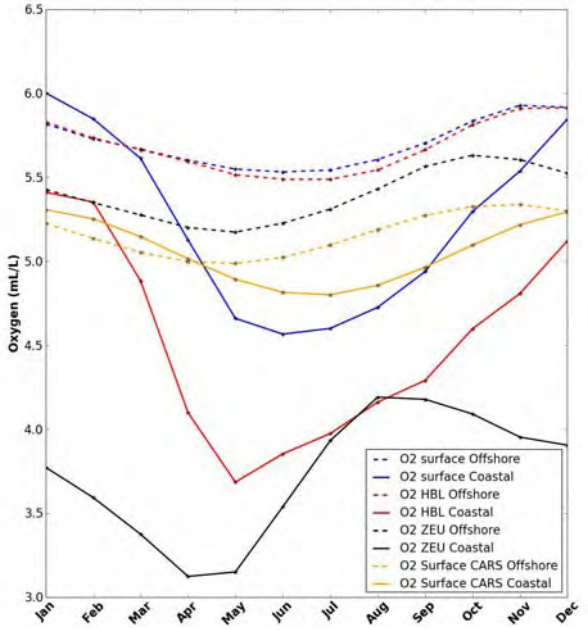

Figure 4.12: Seasonal variability of oxygen averaged over the 3 Peruvian coastal regions (solid lines) and the two offshore regions (dashed lines). Surface oxygen concentration (blue), O_2 concentration over the mixed layer depth (HBL, red), O_2 concentration over the euphotic layer (ZEU, black) for ROMS-PISCES predictions. Surface oxygen concentration (yellow) for the Peru Climatology.

In summary, the ROMS-PISCES climatology does not model quantitatively the oxycline depth to define the epipelagic layer based on oxygen concentration. Nevertheless, since qualitatively the overall prediction of the coast-offshore pattern is reasonable, the sensitivity of the spawning habitat definition to this parameter will be tested independently in optimization experiments (see Chapter 7).

4.3.3.4 Currents

The general scheme of the Peruvian upwelling is presented in Fig. 4.13. Surface predicted annual mean circulation (Fig. 4.13) agrees with this oceanic circulation pattern (Fig. 1.16, Penven et al., 2005; Montes et al., 2010; Montes et al., 2011; Belmadani et al., 2012).

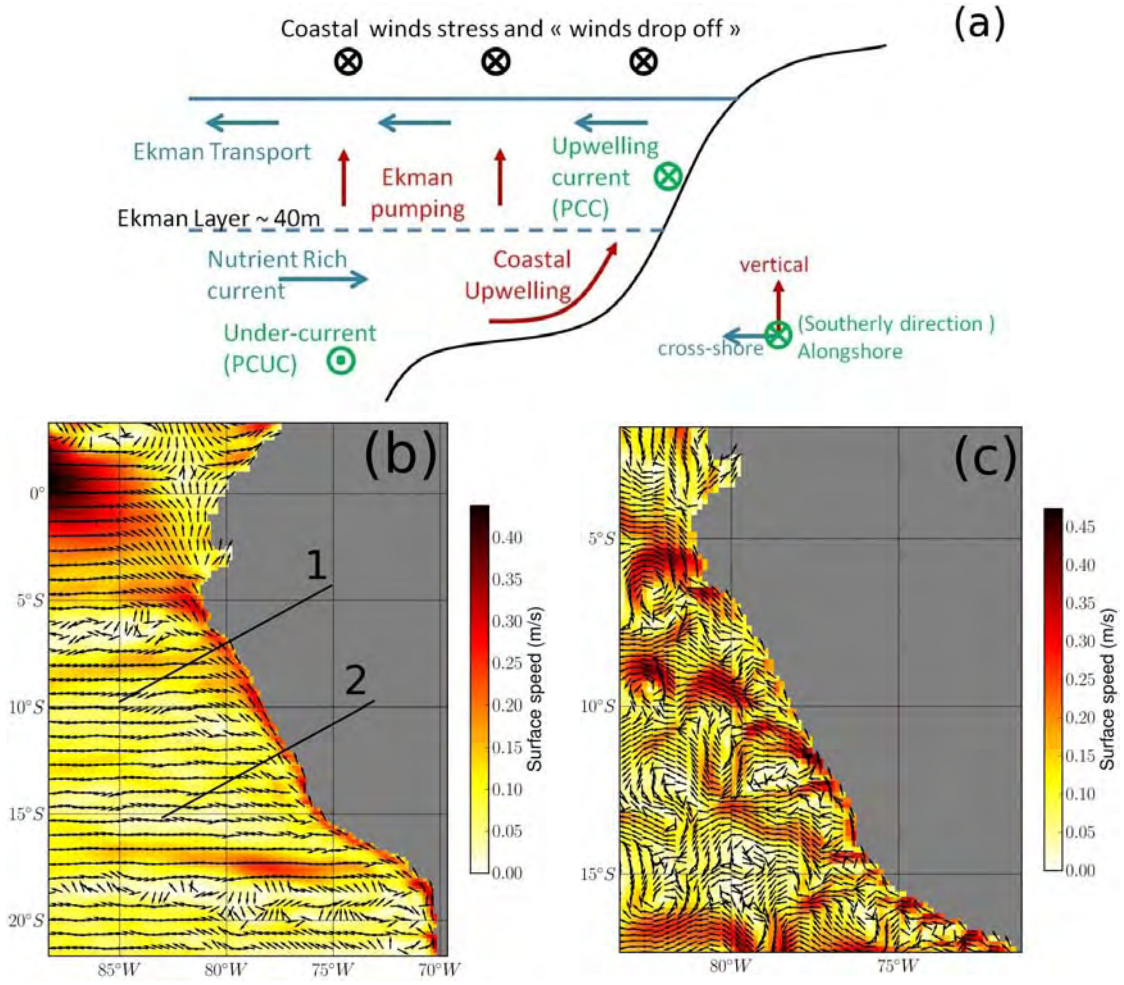

Figure 4.13: (a) Scheme of the Peruvian Upwelling. Adapted from Albert (2011). (b) Annual mean currents in the Peruvian area predicted by the ROMS-PISCES model; (c) Predicted currents for the 16 of February of year 8 of the climatology.

The Peru Coastal Current (PCC), also named Humboldt Current, flows equatorward along the coast. Below a shallow upper layer (around 40 m), the current flows poleward. Mesoscale features are illustrated by displaying the sea surface currents at a given time step (Fig. 4.13b).

Cross-shore currents flow offshore in the Ekman surface layer and flow coastward below. They may strongly influence eggs and larvae retention in the coastal region. The presence and the intensity of these cross-shore currents can be modified locally by the topography, the wind stress and the mesoscale activity. The predicted maximum annual mean offshore transport reaches 14 $cm.s^{-1}$ but with an impact mainly in the first 20 m

(Fig. 4.14). Concerning the simulation of the mesoscale activity and the nearshore circulation, it is difficult to validate them given the lack of high resolution data in this region in the first 100 km (see section 4.2.4).

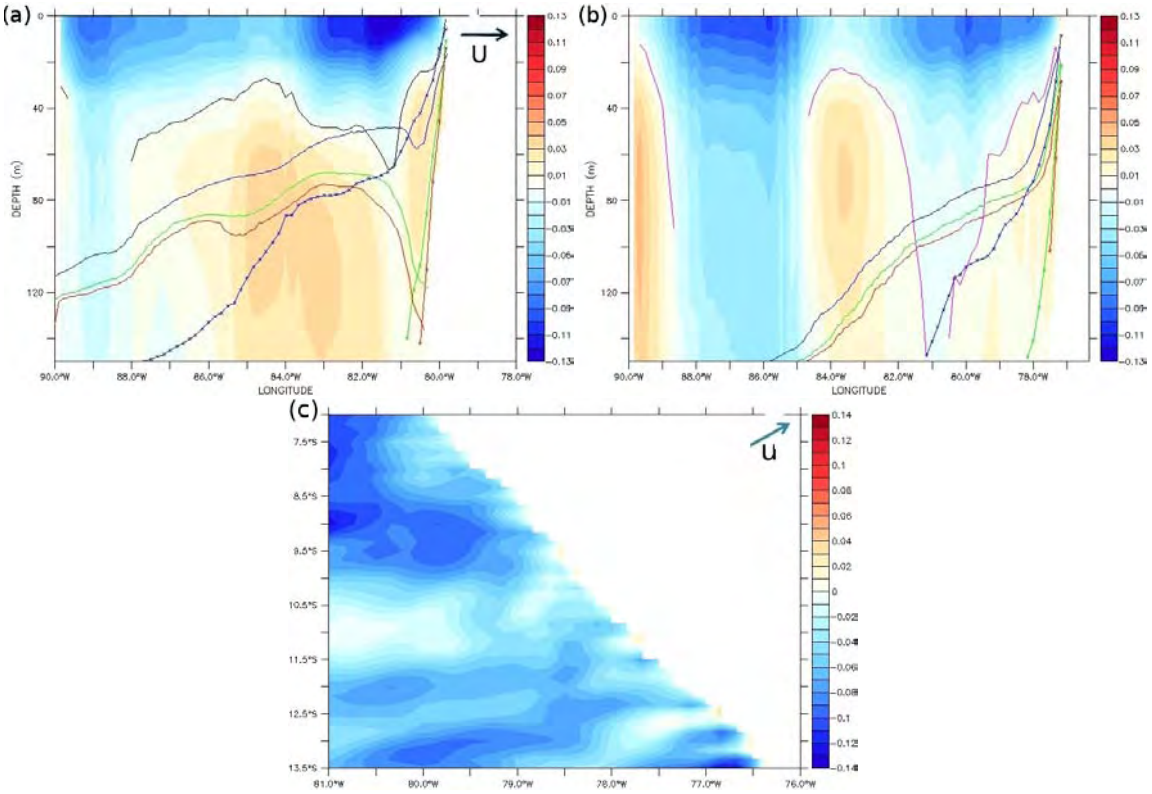

Figure 4.14: (a), (b) Annual mean of predicted cross-shore currents for section 1 and 2 respectively (Fig. 4.13). Black line: limit between offshore and coastward currents; Blue, green and red lines: Depth of 2 $mL.L^{-1}$, $1mL.L^{-1}$, $0.8mL.L^{-1}$ oxygen concentration limit from the Peru oxygen climatology. The lines marked with crosses correspond to the depth of the previous oxygen concentration limits from ROMS-PISCES model. (c) Annual mean of predicted cross-shore currents from 7°S to 13.5°S. Note: Zonal currents are rotated into "cross-shore" directions using an angle of 28.81° (incidence angle of the coast between 6°S and 13.5°S).

Given the vertical structure of the predicted currents described above, their averaging over the mixed-layer to transport eggs and larvae, or over the euphotic layer and related meso- and bathypelagic layers for the micronekton can have strong consequences. Nevertheless, currents averaged over the mixed layer or the euphotic layer show similar patterns but with a lower intensity for the deeper layer (Fig. 4.15a,b), and both vertical definitions integrate correctly the main structure of the currents dynamics. The deeper (mesopelagic) layer also captures fairly well the Peru-Chile UnderCurrent (PCUC), which flows polewards (Fig. 4.15c). Nonetheless, the impact of the choice of vertical definition will be tested on the optimization results in chapter 7.

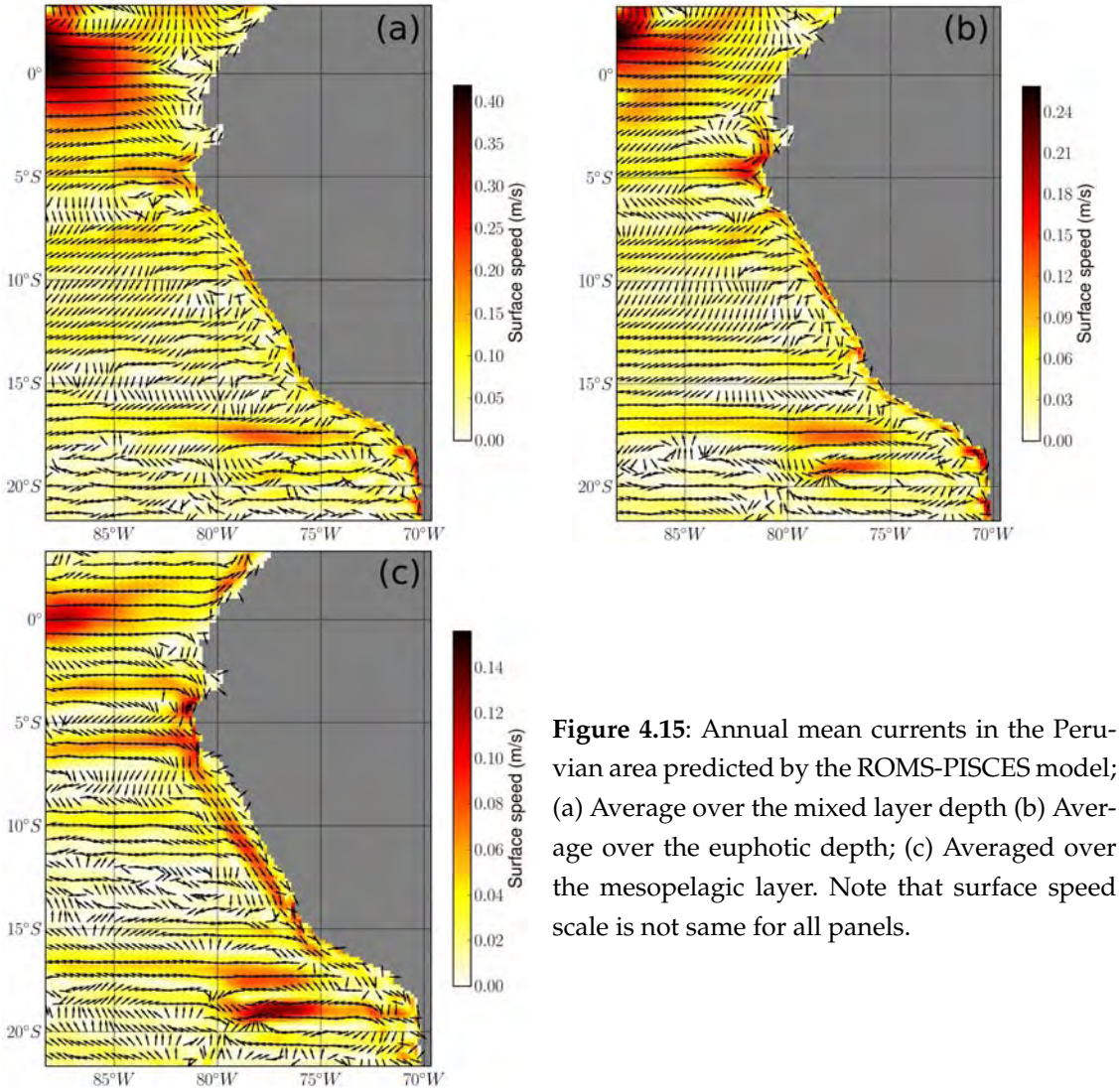

Figure 4.15: Annual mean currents in the Peruvian area predicted by the ROMS-PISCES model; (a) Average over the mixed layer depth (b) Average over the euphotic depth; (c) Averaged over the mesopelagic layer. Note that surface speed scale is not same for all panels.

The SEAPODYM model is limited by its vertical resolution, and vertical approximations are needed. In Chapter VI, we will study how the selected definitions averaging depth may influence the final solution.

4.3.3.5 Temperature, Primary Productivity

The oceanic region off Peru is characterized by a clear seasonal cycle of SST both in the coastal and offshore regions, with a maximum in March, and a minimum in September (Figs. 4.16 and 4.18).

This minimum is linked to stronger winds and maximum upwelling intensity in the austral winter in the nearshore signal, and also to the reduced winter solar flux in the offshore region. The temperature range is lower along the coast (17-21°C) than offshore (18-24°C) due to coastal upwelling. Simulated SST seasonal variability is in good agreement with observations. Modelled temperature values in the coastal and offshore regions

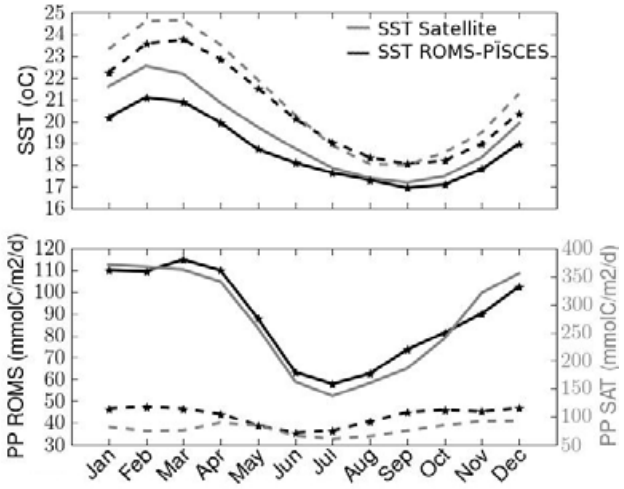

Figure 4.16: Seasonal variation of SST (a) and PP (b) in the Peruvian coastal regions (solid lines) and offshore regions (dashed lines). Black= observations from satellite data; Grey= predictions from ROMS-PISCES.

are underestimated from December to May by 1-2°C. Spatially, northern coastal regions from 8°S to 6°S appear to be overestimated by 1°C (Fig. 4.17).

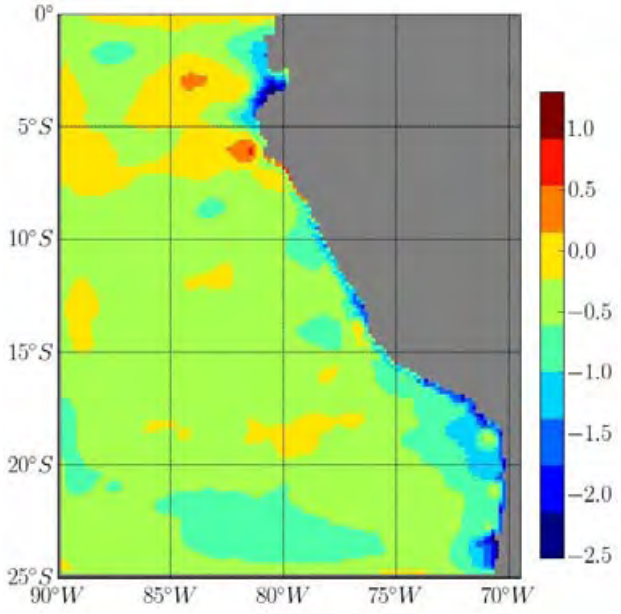

Figure 4.17: Annual Mean of SST from ROMS minus satellite derived SST

The predicted primary production (PP) matches very well the observed seasonality (Fig. 4.16), i.e., increasing after September, peaking in January-March and decreasing after April. However, in comparison with satellite-derived primary production, the model underestimates the primary production by a factor of two (offshore) to three (coast) with a lower amplitude in the range of variability (Figs. 4.16 and 4.18). This discrepancy may be partly due to the fact that computation of primary production is based on the VGPM algorithm used for the general case of oceanic waters, while coastal waters would require a specific parameterization (Behrenfeld and Falkowski, 1997). Furthermore, cloud filling procedure performed on satellite chlorophyll data can also be a source of uncertainties in satellite-based primary production (Behrenfeld and Falkowski, 1997). Another comparison with *in situ* estimates of primary production (Messié et al., 2009) suggests an

underestimation of 20% by the model (Albert et al., 2010). Also, modelled PP seems to be overestimated in the southern coastal zone (SCZ) in austral summer in comparison to VGPM PP (Fig 4.19).

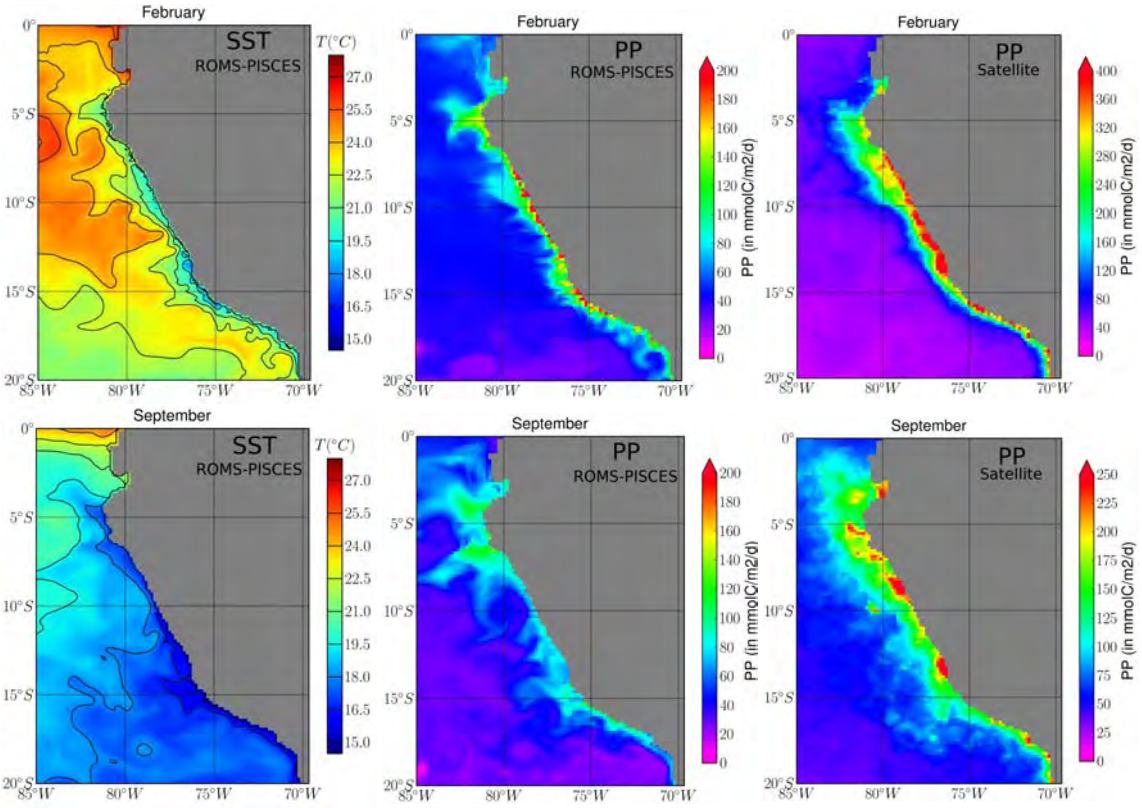

Figure 4.18: Monthly mean spatial distribution in February (top) and September (bottom) of sea surface temperature (left) and primary production (middle) predicted by the ROMS-PISCES model and satellite derived primary production (right).

To investigate the possible reasons of lack of amplitude of primary productivity predictions, we compared the chlorophyll predictions with satellite and *in situ* observations. Indeed, primary production of ROMS-PISCES model is derived from chlorophyll predictions. Interesting, chlorophyll and primary productivity appears to have the same seasonal patterns. Chlorophyll of ROMS is predicted to be higher than satellite ones (Fig. 4.20), however this is in accordance with *in situ* data. For chlorophyll the range of amplitude between ROMS predictions and satellite data are in the same range of variability.

The lack of amplitude of variability of primary productivity should be therefore be investigated by studying the impact of other mechanisms. For example, primary productivity could be limited the lack of realism of some nutrients as nitrate, ammonium, phosphate, silicates and iron predicted by the ROMS-PISCES model.

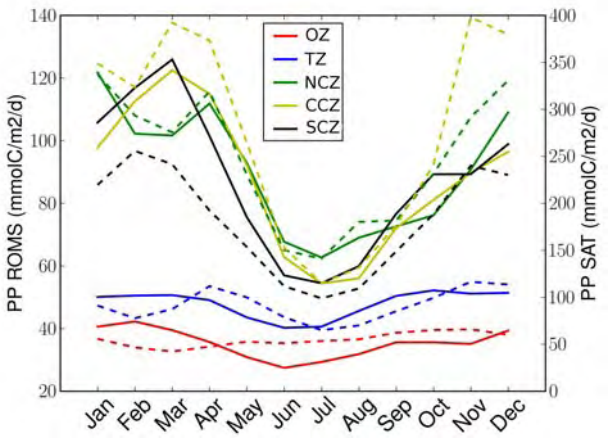

Figure 4.19: Observed and predicted seasonal variability in the Peruvian for the 3 coastal regions (NCZ, CCZ, SCZ regions) and the 2 offshore regions (OZ and TZ regions): predicted primary production (PP) vertically-integrated (solid lines) and observed satellite derived primary production (dashed lines).

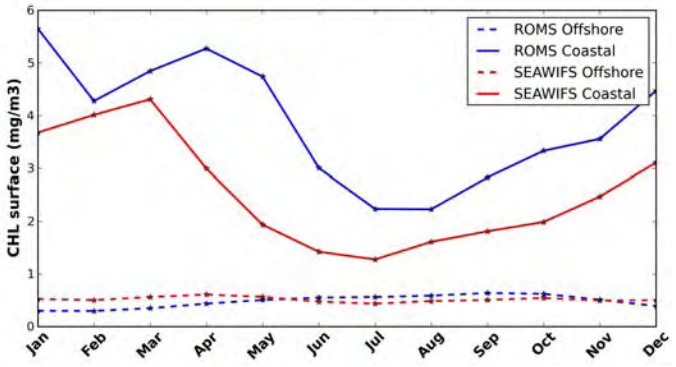

Figure 4.20: Observed and predicted seasonal variability for chlorophyll in the Peruvian coastal regions (solid lines) and offshore regions (dashed lines). Blue line: predicted primary chlorophyll at surface from ROMS-PISCES model; Red line: observed sea surface chlorophyll (SEAWIFS (2000 -2006))

Finally, both temperature and primary production show a strong coastal-offshore gradient, while temperature in the offshore region is also characterized by a decreasing gradient from north to south (Fig. 4.18). Globally, spatial and temporal patterns are well modelled, but the amplitude in the range of PP variability is not well resolved.

4.3.4 Interannual Run 1992-2000

The model domain covers the Humboldt Current system in the region 0°N-45°S and 100°W-70°W, at a spatial resolution of $1/6°$ ($\sim 18\ km$) and 32 depth layers. 5-day-average model outputs are recorded. This run is forced by ERS wind stress over the period 1992-2000, heatfluxes are obtained by the combination of the COADS climatology and ERA40 anomalies (over this period), and CORE precipitation fluxes. At the open boundaries, the regional model is forced by a global simulation from ORCA2-PISCES, which has been forced by ERA40 reanalysis. Monthly dynamical fields and biogeochemical tracers fields are used, except for the oxygen which is obtained from the World Ocean Atlas 2001.

The model is forced by using a "forced flux" formulation (in opposition to "bulk" flux), which means that ROMS SST is restored towards a given observed SST field (Reynolds et al., 2007). The restoring coefficient is a function of observed atmospheric parameters and SST (Barnier et al., 1995). Therefore, it is normal to obtain a perfect accordance between SST of ROMS and SST of observations (Figs. 4.21 and 4.22).

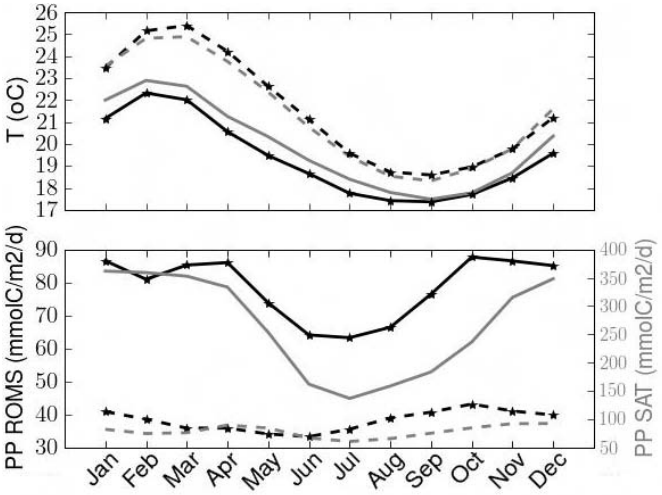

Figure 4.21: Seasonal variation of SST (a) and PP (b) in the Peruvian coastal regions (solid lines) and offshore regions (dashed lines). Black= observations from satellite data (1999-2008); Grey= predictions from ROMS-PISCES (monthly averaged from 1992 to 2000 (years 97/98 from El Niño period were excluded)).

Seasonality pattern of primary productivity display the same lower productivity from June to August than satellite primary productivity. As it occurs for the climatological run, the amplitude of the variability of predicted PP is bad resolved. This run 92-2000 produces also higher productivity than observations from September to October month (Fig. 4.21). We compare the observed and predicted primary productivity during the period 1997-2000 (Fig. 4.23). The main patterns are well reproduced but the low values of primary productivity during the El Niño event are not well predicted.

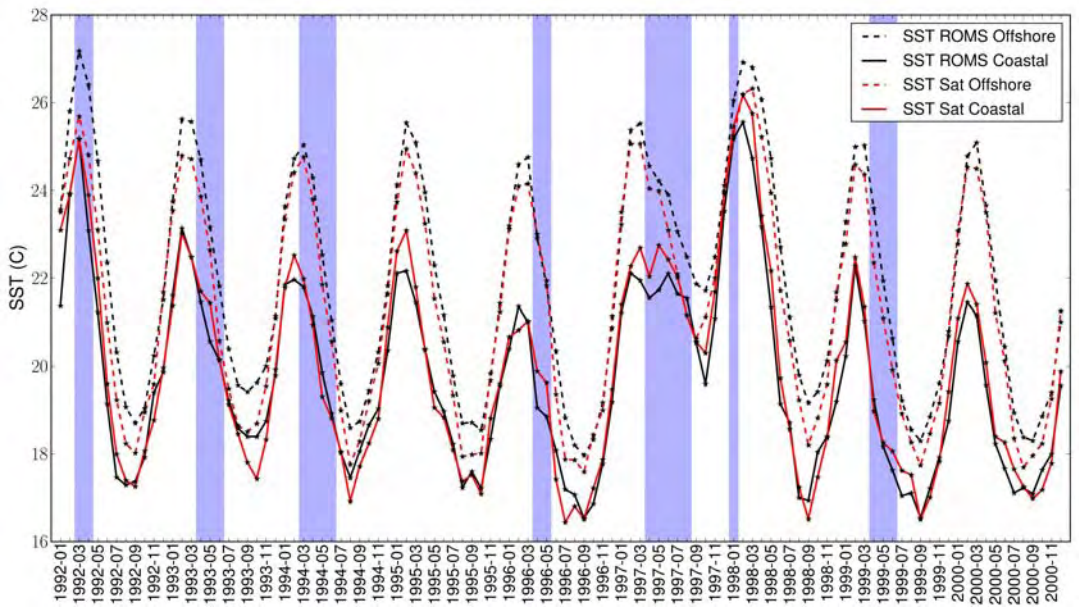

Figure 4.22: SST time series from 1992 to 2000 in coastal regions (solid lines) and offshore regions (dashed lines). In black, SST from ROMS-PISCES model, in red SST from satellite-derived SST climatology (Reynolds et al., 2007). Blue shaded bands indicate the minimal values of gonado-somatic index (Fig. 3.23).

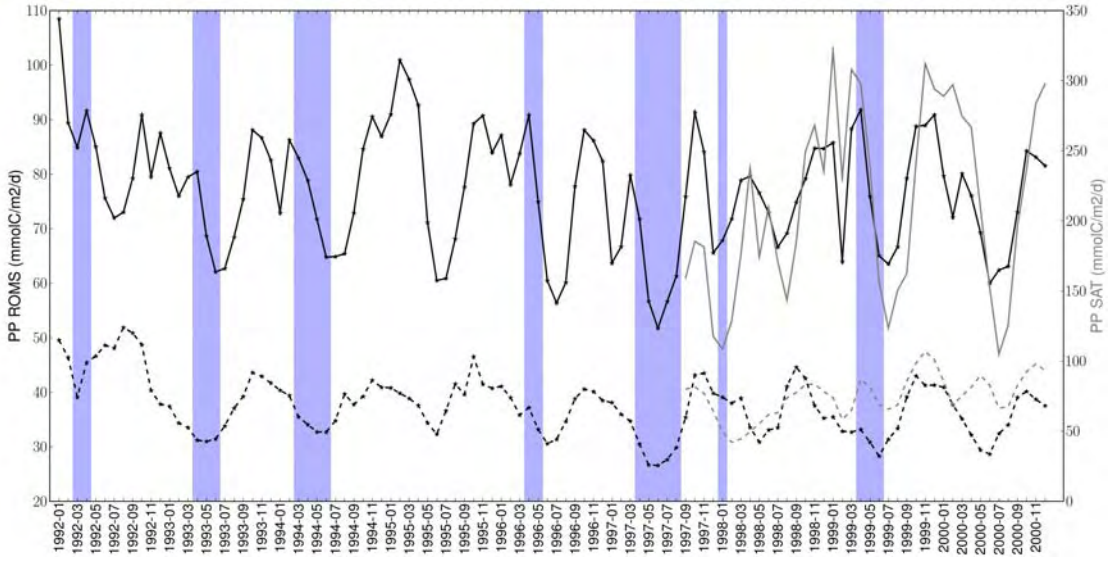

Figure 4.23: PP time series from 1992 to 2000 in coastal regions (solid lines) and offshore regions (dashed lines). In black, PP from ROMS-PISCES model, in grey, observed PP derived from SeaWiFS satellite data, using the VGPM model of Behrenfeld and Falkowski, 1997. Blue shaded bands indicate the minimal values of gonado-somatic index (Fig. 3.23).

We study the impact of the accuracy of the vertical average of the currents. Seasonality of mixed layer depth and euphotic depth show similar patterns of with obtained on the climatological run. The average of euphotic depth in coastal areas is found around 30 m reaching 40 m during the Niño period. Euphotic depth for offshore areas is $\sim$ 60 m. The mixed layer depth show again high range of seasonality varying from 10 to 50 m in coastal areas. During El Niño periods, euphotic depth is modelled to become deeper at the opposite of mixed layer depth will became shallower.

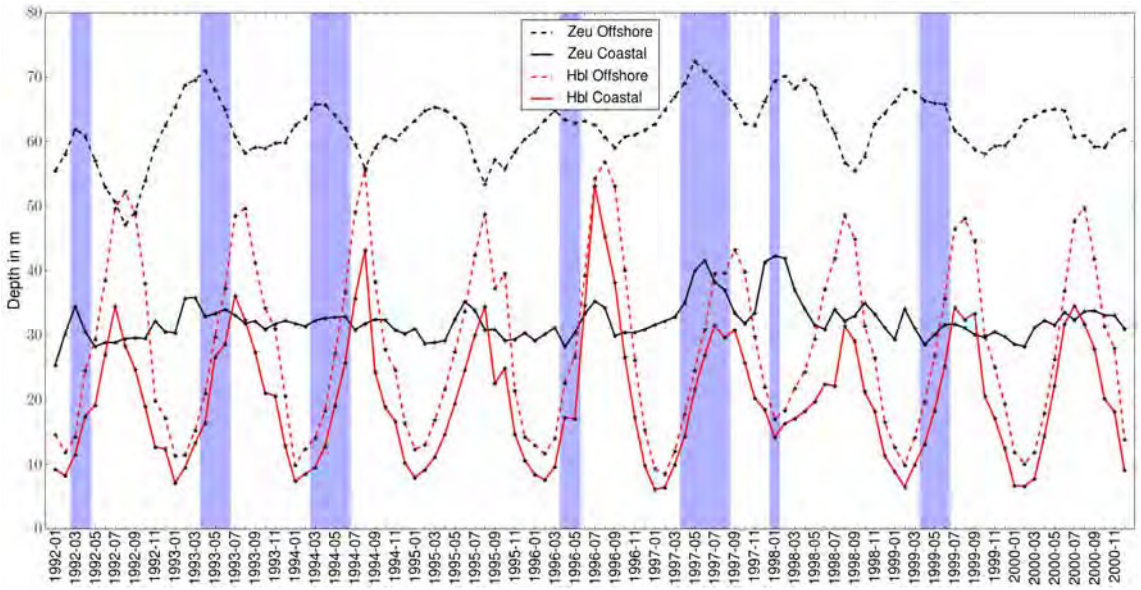

Figure 4.24: Euphotic depth (ZEU) and mixed layer depth (HBL) time series from 1992 to 2000 in coastal regions (solid lines) and offshore regions (dashed lines). Blue shaded bands indicate the minimal values of gonado-somatic index (Fig. 3.23).

For the interannual run, averaging over the mixed layer depth keeps the consistence on the currents with the Peru Coastal Current (PCC) flowing equatorward along the coast (Fig. 4.25). **However, when averaging over the euphotic layer, the averaged currents show high heterogeneity.** From 6°S to 8°S is the undercurrent that flow polewards that is take into account whereas south of 8°S we are considering only the equatorward current. **So, in contrast with the climatological run, the definition of the vertical habitat in this run affects the main known circulation.**

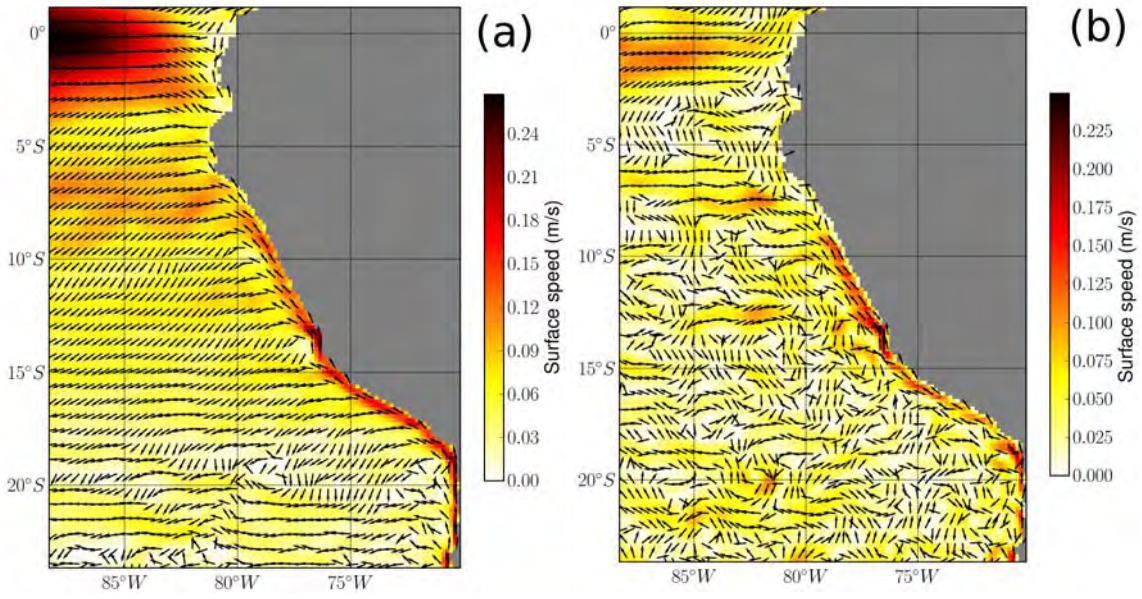

Figure 4.25: Annual mean currents in the Peruvian area predicted by the ROMS-PISCES model; (a) Average over the mixed layer depth (b) Average over the euphotic depth.

4.3.5 Interannual Run 2000-2006

The model domain covers the Humboldt Current system in the region 4°N-22°S and 90°W-70°W, at a spatial resolution of 1/9° ($\sim$ 12.3 km) and 32 depth layers. 5-day-average model outputs are recorded. This run is forced by a 3-days mean output of wind stress, solar and net heat flux from the ORCALIM model, which is forced by daily wind stress and heat fluxes from the IFS ECMWF operational forecast model (http://www.ecmwf.int/research/ifsdocs/). At the open boundaries the model is forced by the dynamical fields (temperature, salinity, velocity and sea level) from a 3 days -mean ORCALIM global ocean general circulation model at 1/4° horizontal resolution (Garric et al., 2008). The biogeochemical initial and boundary conditions are forced by a monthly climatology of the global ORCA2-PISCES bio-dynamic coupled model at 2° resolution over 1992-2000 is used (L. Bopp, pers. com., 2009) for the biological variables and by the World Ocean Atlas 2001 (WOA, (Conkright et al., 2002)) for the oxygen. Dissolved oxygen is specified from the World Ocean Atlas 2001.

As the interannual run 1992-2000, the model is forced by using "forcing flux", therefore again we expect that predicted and observed sea surface temperature (Fig. 4.26) are in good agreement with observation.

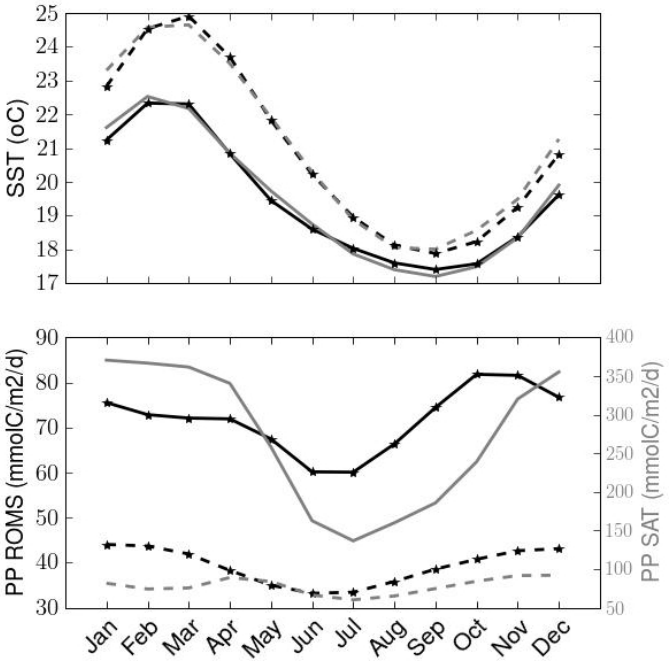

Figure 4.26: Seasonal variation of SST (a) and PP (b) in the Peruvian coastal regions (solid lines) and offshore regions (dashed lines). Black= observations from satellite data (1999-2008); Grey= predictions from ROMS-PISCES (monthly averaged from 2000 to 2006).

Maximum of primary productivity predicted by ROMS-PISCES occurs earlier than observed by satellite. High productivity appears from September to April whereas for satellite data it occurs from December to April (Figs. 4.26 and 4.27). Primary productivity predicted with this run is lower than primary productivity predicted by the runs at 1/6° resolution.

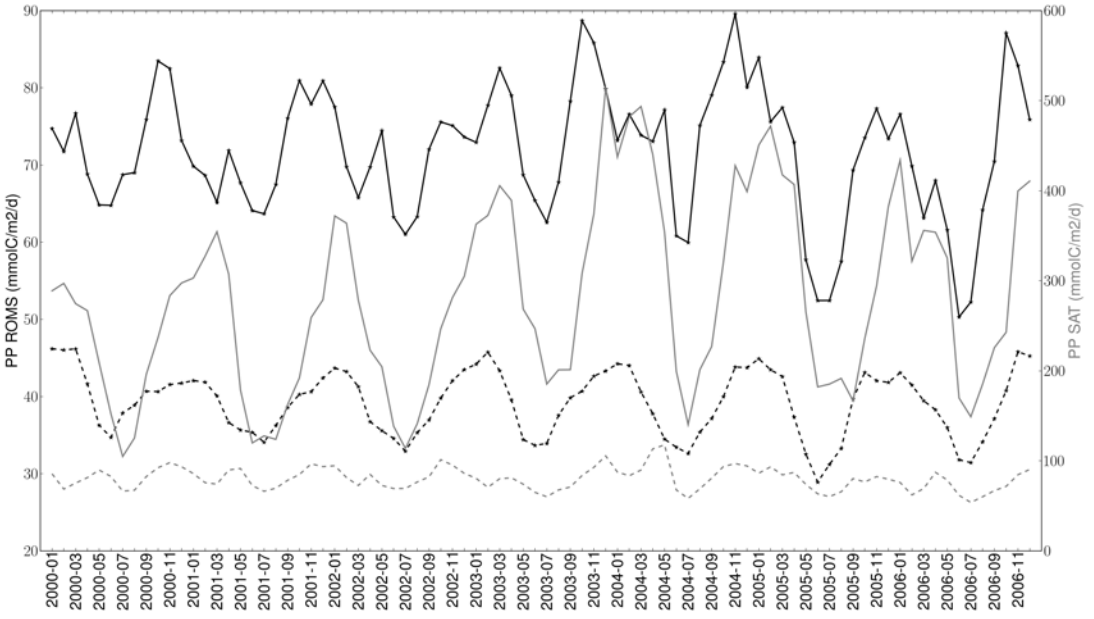

Figure 4.27: PP time series from 2000 to 2006 in coastal regions (solid lines) and offshore regions (dashed lines). In black, PP from ROMS-PISCES model, in grey, observed PP derived from SeaWiFS satellite data, using the VGPM model of Behrenfeld and Falkowski (1997).

Currents averaged over the mixed layer depth or the euphotic depth show similar patterns but their intensity is lower when the averaging depth is bigger. So, both hypotheses appear to keep correctly the main structure of the currents dynamics (Fig 4.28).

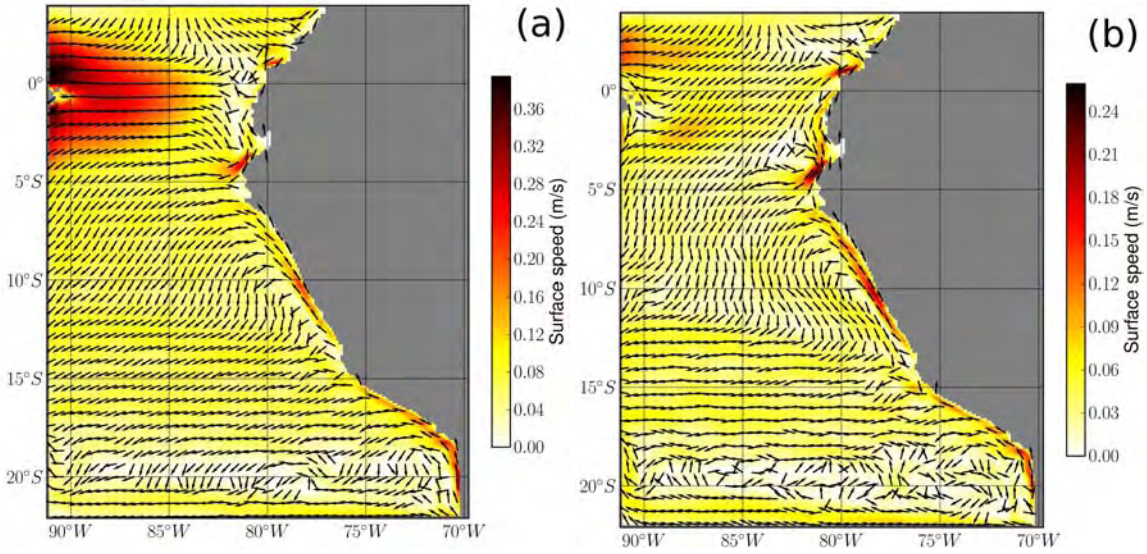

Figure 4.28: Annual mean currents in the Peruvian area predicted by the ROMS-PISCES model; (a) Average over the mixed layer depth (b) Average over the euphotic depth.

4.4 Eggs and larvae predators: the micronekton model

In SEAPODYM, the mid-trophic level is modelled by a micronekton model (Lehodey et al., 2010), which is used as an input for anchovies and sardines predators. The micronekton model is also forced with the ROMS-PISCES physical-biogeochemical environmental input. Though the definition of functional groups of micronekton is based on time of development of organism, the micronekton can be roughly characterized by a size spectrum in the range of 2-20 cm dominated by crustaceans, fish, and cephalopods.
The model includes 6 groups from surface to 1000 m deep with vertically migrant and non-migrant components (Lehodey et al., 2010, Fig. 4.29). Parametrization of the dynamics of these components is based on a temperature-linked time development relationship (Lehodey et al., 2010).

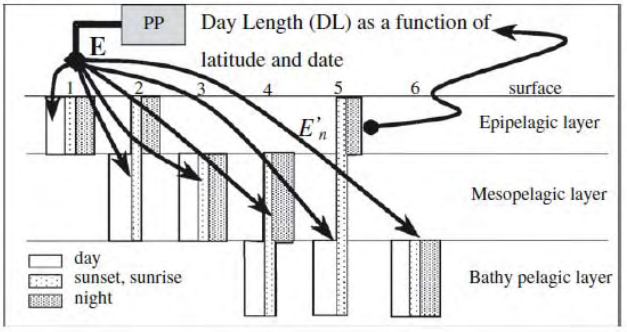

Figure 4.29: Conceptual model of the mid-trophic components in the pelagic ecosystem. Daily vertical distribution patterns of the micronekton in the pelagic ecosystem: (1) epipelagic; (2) migrant mesopelagic; (3) non-migrant mesopelagic; (4) migrant bathy-pelagic; (5) highly migrant bathypelagic; and (6) non-migrant bathypelagic. The part of energy (E) transferred from primary production (PP) to intermediate trophic levels is redistributed (E'_n) through the different components. From Lehodey et al. (2010).

Lehodey et al. (2010) validated the micronekton model at basin scale, but some limitations exist at higher resolution and in a coastal domain. Here, we present the spatial and temporal patterns of the different outputs of the micronekton model for the 3 runs described previously.

The epipelagic group of the micronekton should contain species up to 20 cm long. The micronekton model does not take into account fish swimming speed, it only considers passive advection. Although this approximation may be acceptable at large basin scale, it is less the case at local scale. Indeed, small pelagic fishes are part of this group and are known to actively swim.

Fig. 4.30 presents the annual average of the epipelagic group densities for the 3 different runs. This group contains small pelagic fish which are mainly concentrated in coastal areas, but the model predicts a higher biomass offshore in the 3 cases. In addition, the

absolute biomass should be higher, with anchovy biomass only estimated to be around 80 g/m^2. We note however, that in the anchovy and sardine version of SEAPODYM the absolute biomass is not important.

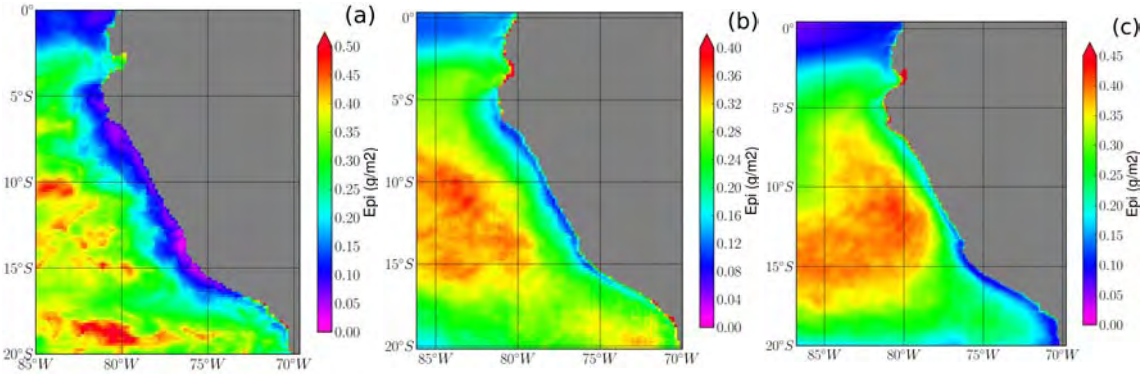

Figure 4.30: Average spatial distribution of the epipelagic group. (a) Mean density over year 8 of the climatological run; (b) Mean density over interannual run 1992-2000 (without El Niño period), (c) Mean density over the interannual run 2000-2006.

Due to the lack of realism of the epipelagic group, this group will not be use as input for eggs and larvae predators. Only mesopelagic and bathypelagic migrants, that come in the epipelagic layer at night, will be considered as possible predators.

Spatial representation of mesopelagic migrants and bathypelagic (Fig. 4.32 are in better agreement with observations which show high biomass offshore the continental shelf edge (Fig. 3.36 in section 3.8).

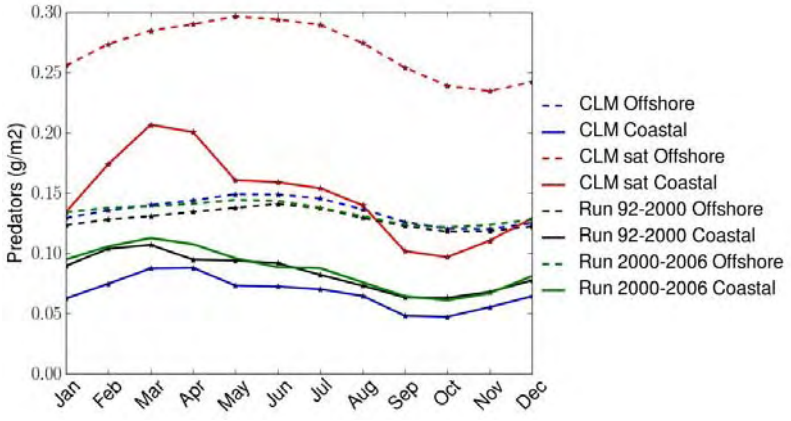

Figure 4.31: Seasonality of mesopelagic and bathypelagic migrants groups for 4 different physical forcing.

Using different physical and biochemical forcing, the 4 runs display similar seasonalities, with high density in February/March and low density in September/October (Fig. 4.31). The main difference is observed by using the satellite primary productivity to run the predators model. Keeping the same seasonality, the amplitude of the variability for this run is higher in coastal and offshore regions. The seasonality of the predators appears

to be mainly controlled by temperature, which peaks in February-March and reaches a minimum in September, as is observed for predators.

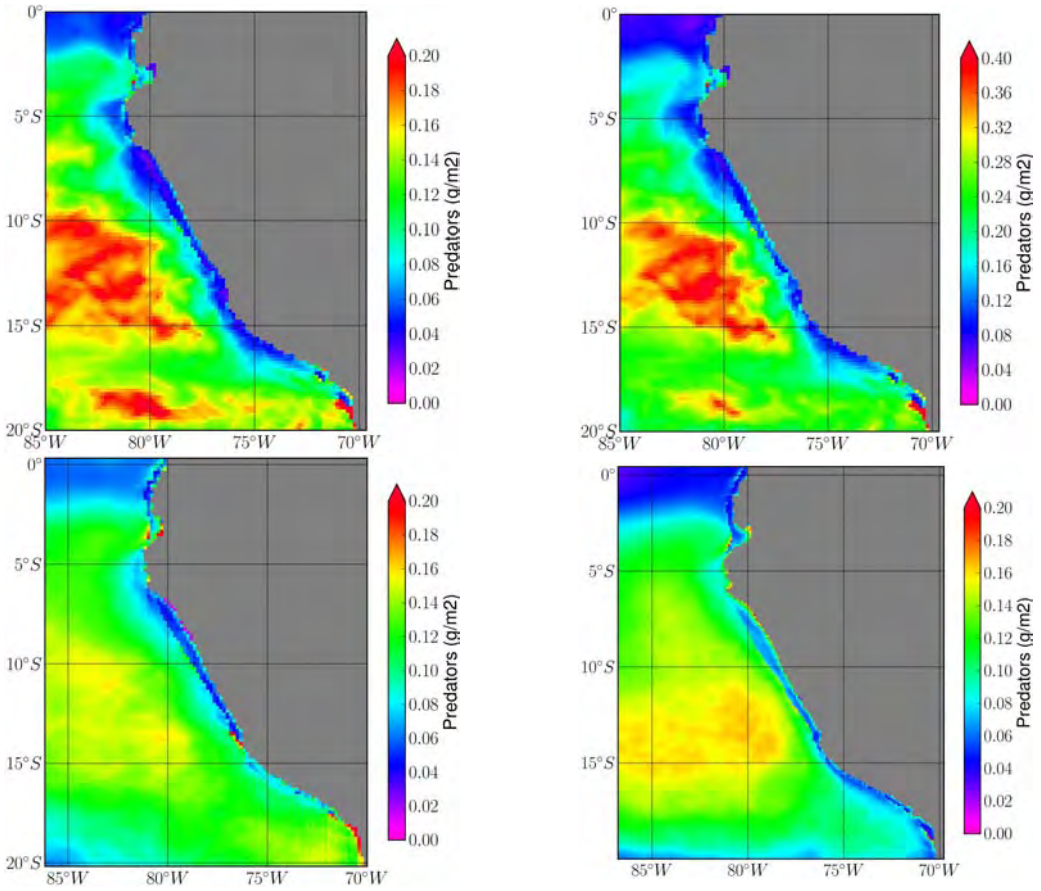

Figure 4.32: Average spatial distribution of the mesopelagic and bathypelagic migrants groups. (a) Mean density over year 8 of the climatological run; (b) Mean density over the year 8 of the climatological run using satellite Primary productivity; (c) Mean over the interannual run 1992-2000 (without Niño period), (d) Mean over the interannual run 2000-2006.

Epipelagic layer has been defined in the micronekton model by using the euphotic depth. However, validation of this predicted variable is difficult to make. We have seen that averaging current in an upwelling system can affect considerably the structure and the corresponding currents. Therefore, we test the sensibility of the micronekton output by averaging the currents over two times the euphotic depth (2ZEU) and over the mixed layer-depth (HBL) (Figs. 4.33 and Fig. 4.34). We should note that averaging currents over 2 times the euphotic depth correspond to consider an average of 60 m for the vertical habitat in coastal areas. In both cases, we do not observe high densities in coastal areas as expected for epipelagic biomass. When averaging over the mixed layer depth, offshore transport is higher and therefore predators biomass are located more offshore. When averaging two times the euphotic depth, we keep more the signal of coastward transport, higher biomass is observed in coastal areas, but still not enough in comparison with offshore regions.

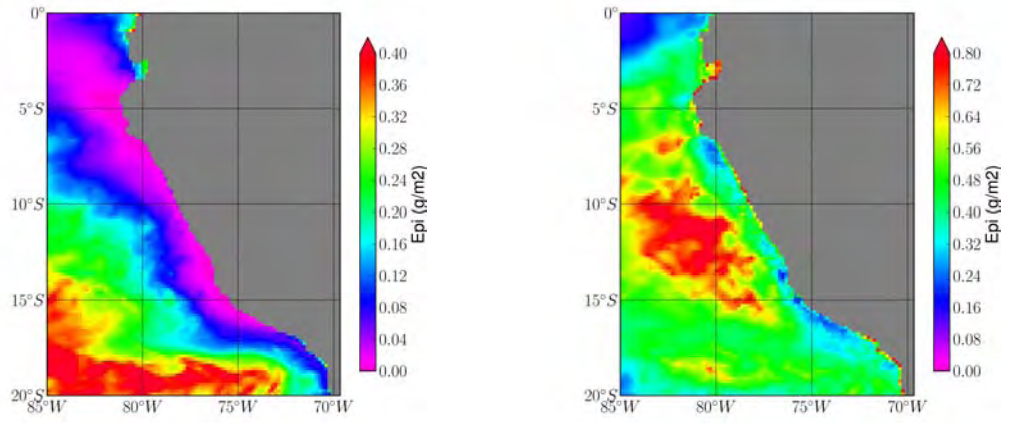

Figure 4.33: Average spatial distribution of the epipelagic group. (a) Average over the mixed layer depth; (b) Average over two times the euphotic depth.

Diffusion parameter was by default fixed to 200 $m^2.s^{-1}$ (S1, Fig. 4.34, for a range of scale phenomenon between 30 km and 120 km (Okubo, 1971), at the same value than the one used on the SEAPODYM model. However, if we use a higher diffusion 400 $m^2.s^{-1}$ (S2, Fig. 4.34) the seasonality displays the same pattern. Averaging the currents on the mixed layer depth affect highly the seasonality of predators' biomass which present a minimum from March to June (Fig. 4.34).

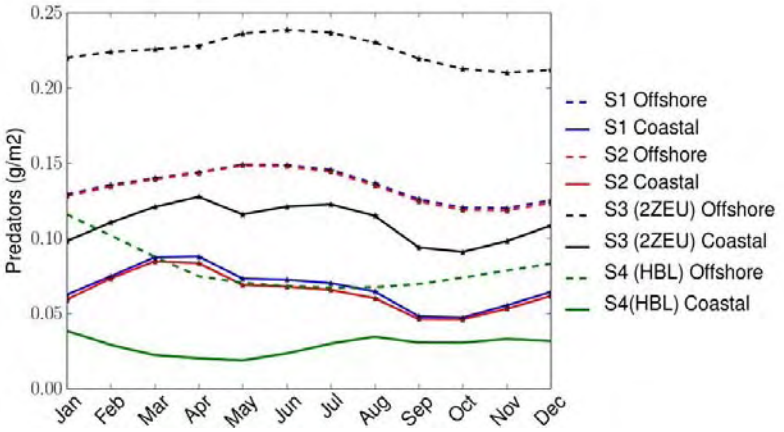

Figure 4.34: Seasonality of the mesopelagic and bathypelagic migrants groups for the year 8 of the climatological run. Sensibility to diffusion parameter and to vertical habitat definition.

Finally, the 3 different ROMS-PISCES runs, and consequently the use of different physical and biogeochemical forcings, predict globally the same seasonal and spatial patterns. The main source of uncertainty which have a high impact on the seasonality and spatial distribution of predators is due to the vertical habitat definition and the low vertical resolution of the micronekton model.

4.5 Conclusion

We have seen in this chapter the three physical and biogeochemical runs of ROMS-PISCES which will be used to force the SEAPODYM model: (i) a climatological run, (ii) an interannual run from 1992 to 2000 at 1/6° of resolution, and (iii) an interannual run from 2000 to 2006 at 1/9° of resolution. These forcing inputs have potential sources of errors. The three runs are in good accordance with the sea surface temperature. Currents are difficult to validate due to the lack of data in coastal areas but appears to be in good accordance with the literature. Predictions of primary productivity fit the observed seasonality in the case of the climatological run and 1992-2000 interannual run, but not with the good amplitude of variability. The 2000-2006 run displays correctly the observed minimum of primary productivity but not the maximum, which occurs in Oct-Nov in contrast with observed cycle.

For eggs and larvae predators, we have shown the high uncertainties of the micronekton model in coastal areas in the Peruvian upwelling. Due to these limitations, we will not consider the epipelagic group, supposing therefore that it should be largely dominated by anchovy population, for which there is no strong evidence of cannibalism. We only consider the mesopelagic and bathypelagic migrants coming into the surface at night (Lehodey et al., 2008) in better agreement with observations.

Finally, definition of the vertical habitat can affect structure and mesoscale features.

By default, to run eggs and larvae dynamics, **we choose to use the mixed layer as vertical habitat definition**. Therefore, we consider that eggs and larvae are mainly driven by surface currents. Currents and oxygen fields will be averaged over this layer. Sea surface temperature will be used to drive the model. Total primary production integrated over the total vertical depth will be used as a proxy for larvae's food.

To run the micronekton model, the euphotic layer will be used to define the vertical habitat definition as it is used on the parametrization of the micronekton model at basin scale. Currents and temperature averaged over the euphotic layer will be use to drive this model.

However, we will test the sensibility of the model results to vertical resolution with the optimization approach, by using both mixed layer and euphotic depth habitat definition (Chapter 7).

Bibliography

Albert, A., 2011: *Etude de la variabilité du phytoplancton dans l'upwelling du Pérou aux échelles saisonnières et intrasaisonnières*. PhD thesis, Thèse de Doctorat - Université Pierre et Marie Curie (Paris VI).

Albert, A., V. Echevin, M. Levy, and O. Aumont, 2010: Impact of nearshore wind stress curl on coastal circulation and primary productivity in the Peru upwelling system. *J. Geophys. Res*, **115**, C12033.

Amante, C. and B. W. Eakins, 2009: Etopo1 1 arc-minute global relief model: Procedures, data sources and analysis. Technical report, 19.

Aumont, O., 2003: An ecosystem model of the global ocean including Fe, Si, P colimitations. *Global Biogeochemical Cycles*, **17**(2), 1060.

Aumont, O. and L. Bopp, 2006: Globalizing results from ocean in situ iron fertilization studies. *Global Biogeochemical Cycles*, **20**(2), GB2017.

Barnier, B., L. Siefridt, and P. Marchesiello, 1995: Thermal forcing for a global ocean circulation model using a three year climatology of ECMWF analyses. *, J. Mar. Syst*, **6**, 363–380.

Behrenfeld, M. J. and P. G. Falkowski, 1997: Photosynthetic rates derived from satellite-based chlorophyll concentration. *Limnology and oceanography* 1–20.

Belmadani, A., V. Echevin, B. Dewitte, and F. Colas, 2012: Equatorially-forced intraseasonal propagations along the Peru-Chile coast and their relation with the nearshore eddy activity in 1992-2000: A modelling study. *Journal of Geophysical Research*.

Bertrand, A., M. Ballón, and A. Chaigneau, 2010: Acoustic observation of living organisms reveals the upper limit of the oxygen minimum zone. *PlosOne*, **5**, e10330.

Bertrand, A., A. Chaigneau, S. Peraltilla, J. Ledesma, M. Graco, F. Monetti, and F. Chavez, 2011: Oxygen, a fundamental property regulating pelagic ecosystem structure in the coastal southeastern tropical pacific. *PlosOne*, **6**(12), e29558.

Conkright, M., R. Locarnini, H. Garcia, T. O'Brien, d. C. S. T. Boyer a, and J. Antononov, 2002: World ocean atlas 2001 : objectives, analyses, data statistics and figures [cd-rom]. *Silver Spring, Md : NOAA Atlas NESDIS 42.*

Echevin, V., O. Aumont, J. Ledesma, and G. Flores, 2008: The seasonal cycle of surface chlorophyll in the peruvian upwelling system: A modelling study. *Progress in Oceanography*, **79**(2-4), 167–176.

Fuenzalida, R., W. Schneider, J. Garcés-Vargas, L. Bravo, and C. Lange, 2009: Vertical and horizontal extension of the oxygen minimum zone in the eastern south pacific ocean. *Deep Sea Res II*, **56**, 992–1003.

Garcia, H., R. Locarnini, T. Boyer, J. Antonov, and O. Baranova, 2010: World Ocean Atlas 2009, Vol. 3: Dissolved Oxygen, Apparent Oxygen Utilization, and Oxygen Saturation. (NOAA Atlas NESDIS 70, U.S. Government Printing Office, Washington, D.C.).

Garric, G., R. Bourdalle-Badie, d. C. B. O. Le Galloudec a, C. Derval, E. Durand, and Y. Drillet, 2008: Description of the interannual experiment ORCA025-T09, 1998-2006, Annex 2 of Mercator Ocean Report n° 235 32.

Gruber, N., Z. Lachkar, H. Frenzel, P. Marchesiello, M. Munnich, J. McWilliams, T. Nagai, and G. Plattner, 2011: Eddy-induced reduction of biological production in eastern boundary upwelling systems. *Nature Geoscience*, **4**(11), 787–792.

Hansen, D. V. and P. M. Poulain, 1996: Quality control and interpolations of WOCE-TOGA drifter data. *Journal of Atmospheric and Oceanic Technology*, **13**(4), 900–909.

Lehodey, P., R. Murtugudde, and I. Senina, 2010: Bridging the gap from ocean models to population dynamics of large marine predators: A model of mid-trophic functional groups. *Progress In Oceanography*, **86**(1-2), 302–315.

Lehodey, P., I. Senina, and R. Murtugudde, 2008: A spatial ecosystem and populations dynamics model (SEAPODYM) - Modeling of tuna and tuna-like populations. *Progress in Oceanography*, **78**(4), 304–318.

Lett, C., P. Penven, P. Ayón, and P. Fréon, 2007: Enrichment, concentration and retention processes in relation to anchovy (*Engraulis ringens*) eggs and larvae distributions in the northern humboldt upwelling ecosystem. *Journal of Marine Systems*, **64**(1-4), 189–200.

Mathiesen, O., 1989: *Adaptation of the anchoveta (Engraulis ringens) to the Peruvian upwelling ecosystem. In: The Peruvian upwelling ecosystem: dynamics and interactions*. The WorldFish Center, Ed. by D. Pauly, P. Muck, J. Mendo, and I. Tsukayama (Eds). ICLARM Conference Proceedings, 220-234.

Messié, M., J. Ledesma, D. Kolber, R. Michisaki, D. Foley, and F. Chavez, 2009: Potential new production estimates in four eastern boundary upwelling ecosystems. *Prog. Oceanogr.*, **83**, 151–158.

Montes, I., F. Colas, X. Capet, and W. Schneider, 2010: On the pathways of the equatorial subsurface currents in the eastern equatorial pacific and their contributions to the Peru-Chile undercurrent. *Journal of Geophysical Research*, **115**, 16 PP.

Montes, I., W. Schneider, F. Colas, B. Blanke, and V. Echevin, 2011: Subsurface connections in the eastern tropical pacific during la nina 1999-2001 and el nino 2002-003. *Journal of Geophysical Research*, **116**, 18 PP.

Okubo, A., 1971: Oceanic diffusion diagrams. *Deep-Sea Res*, **18**, 789–802.

O'Reilly, J. E., S. Maritorena, B. G. Mitchell, D. A. Siegel, K. L. Carder, S. A. Garver, M. Kahru, and C. McClain, 1998: Ocean color chlorophyll algorithms for SeaWiFS. *Journal of Geophysical Research*, **103**(C11), 24937–24953.

Penven, P., V. Echevin, J. Pasapera, F. Colas, and J. Tam, 2005: Average circulation, seasonal cycle, and mesoscale dynamics of the 3 peru current system: A modeling approach. *Journal of Geophysical Research*, **110**, C10021.

Reynolds, R. W., T. M. Smith, C. Liu, D. B. Chelton, K. S. Casey, and M. G. Schlax, 2007: Daily High-Resolution-Blended analyses for sea surface temperature. *Journal of Climate*, **20**(22), 5473–5496.

Rio, M. and F. Hernandez, 2003: High-frequency response of wind-driven currents measured by drifting buoys and altimetry over the world ocean. *Journal of Geophysical Research*, **108**, 19 pp.

Rio, M. H., S. Guinehut, and G. Larnicol, 2011: New CNES-CLS09 global mean dynamic topography computed from the combination of GRACE data, altimetry, and in situ measurements. *Journal of Geophysical Research*, **116**(C7).

Shchepetkin, A. F. and J. C. McWilliams, 2005: The regional oceanic modeling system (ROMS): a split-explicit, free-surface, topography-following-coordinate oceanic model. *Ocean Modelling*, **9**(4), 347–404.

Chapter 5

Modeling early life history

Sommaire

5.1 Introduction . . . 139
5.2 Materials and Methods . . . 141
5.2.1 Anchovy data . . . 141
5.2.2 Bio-physical environment . . . 142
5.2.3 SEAPODYM spawning habitat and larval recruitment . . . 143
5.2.4 Simulations . . . 145
5.3 Results . . . 146
5.3.1 Eggs and larvae distributions . . . 146
5.3.2 Seasonal variability of predicted environment . . . 148
5.3.3 Observed seasonal variability in anchovy reproduction . . . 150
5.3.4 Simulated spatial distributions . . . 153
5.3.5 Simulated seasonality . . . 154
5.4 Discussion . . . 158
A Appendices . . . 161
A.1 Continuous model of larvae spatial dynamics. . . . 161
A.2 Spawning . . . 161
A.3 Mortality . . . 162
Bibliography . . . 163

This chapter is presented under the form of an article to be submitted soon. It presents the first attempt to apply SEAPODYM to early life stages of Peruvian anchovy. As such, it summarizes parts of chapters 2 (theory), 3 (observations) and 4 (environmental forcing description). It shows that the combination of a few simple mechanisms with a limited number of parameters quickly leads to multitude of solutions that only rigorous parameter optimization approaches can help to explore. This work has entirely been performed within my thesis work. The coauthors are my thesis advisors and the partners who provided biological data (from IMARPE) and environmental forcing for the model.

Modeling the spawning habitat and larvae dynamics of Peruvian anchovy within the Eulerian SEAPODYM framework

Olga Hernandez[1], Patrick Lehodey[1], Inna Senina[1], Patricia Ayon[2], Arnaud Bertrand[2,3]
Vincent Echevin[3,4], Ramiro Castillo[2], Betsy Buitron[2], Philippe Gaspar[1]

[1] MEMMS (Marine Ecosystems Modelling and Monitoring by Satellites), CLS, France
Email: ohernandez@cls.fr

[2] Instituto del Mar del Perú, Esquina Gamarra y Gral. Valle sn, Apartado 22, Callao, Lima

[3] IRD, UMR212 EME IFREMER/IRD/UM2, Avenue Jean Monnet, BP 171, 34203 Sète, France

[4] IRD, LOCEAN, Université Pierre et Marie Curie, Boite, 4 Place Jussieu, 75252 Paris, France

Abstract

The Spatial Eulerian Ecosystem and Population Dynamic Model (SEAPODYM) is used to develop an application for small pelagic species. The model domain is the Humboldt Current system (HCS) off Peru. The present work provides a detailed study of the spatial and temporal dynamics of anchovy (Engraulis ringens) eggs and larvae dynamics. The analysis of a large dataset of eggs and larvae samples allowed characterizing the spawning seasonal cycle, which exhibits a clear rest period between April and June. Environmental variables from the ROMS-PISCES coupled physical-biogeochemical model provide offline forcing to simulate the spawning habitat and transport larvae. The habitat is critical since it controls the initial recruitment of larvae in the first cohort and subsequent spatio-temporal variability of natural mortality during their drift with currents described by a system of Eulerian equations. The different processes proposed to control the fish larvae recruitment in the model are presented and the sensitivity of parameterization explored with a climatology and a reasonable range of values found in the literature. These mechanisms include the optimal temperature for larvae, the prey abundance, the trade-off between presence/absence of prey and predators of larvae, and the retention or dispersion by currents. One mechanism alone is not sufficient to explain observed spatial and temporal patterns. A combination of mechanisms, at least temperature and prey availability, is needed to retrieve a good spatial distribution of eggs and larvae, as well as a contrasted seasonal variability. We also show that match-mismatch mechanism is not at play in the HCS for anchovy. Overall, our study shows that the regional coupled model with chosen description of the early life stages of anchovy are not sufficient at this point to fully represent the complexity of the temporal dynamics of eggs and larvae. Though this modelling framework provides a useful tool to test various functional relationships, the tudy demonstrates that the combination of a few simple mechanisms with a limited number of parameters quickly leads to multitude of solutions that only rigorous parameter optimization approaches can help to reach.

5.1 Introduction

The Humboldt Current system (HCS) off Peru and Chile is one of the most productive coastal upwelling system in the world and the most productive system in terms of fish biomass (Bakun and Broad, 2003; Chavez et al., 2008). With less than 0.1 % of the world ocean surface, the annual landings represent about 10% of the world fish catches. This system is characterized by a high variability due to climatic processes interacting on different different time scales: oceanographic variations at secular scale, the Pacific decadal oscillation (PDO) at multi-decadal scale, El Niño Southern Oscillation (ENSO) at the interannual scale, and the seasonal variability at intra-annual scale (Chavez et al., 2008). Peruvian anchovy (*Engraulis ringens*) and sardine (*Sardinops sagax*), with relatively short life span (around 4 and 8 years respectively), a fast growing and early maturity (one and two years respectively) are well adapted to this variability (Bertrand et al., 2004; Gutiérrez et al., 2007; Swartzman et al., 2008; Espinoza and Bertrand, 2008). Anchovy population which dominates the system is exploited by the largest commercial fleet targeting a single species, with mean annual landing of about 6 million tonnes since 1999. Within the last decades fluctuation in anchovy or sardine biomass occurred and the sardine population virtually disappeared from the Peruvian coast in the early 2000s (Gutiérrez et al., 2007, Bertrand et al., 2011).

The strong impact of climate variability on anchovy and sardine stocks and their long term fluctuations have questioned oceanographers and fisheries scientists for decades, leading to a considerable literature and still several debated hypotheses (Chavez et al., 2008; Bertrand et al., 2011). In addition, this strong variability makes the management of the fisheries difficult. In this context, the development of a model that could help us to better understand the impact of the environmental variability on exploited populations would provide a considerable move forward.

Spatial models developed to investigate the relationships of small pelagic species with their (simulated) environment are mostly based on individual-based model (IBM) (i.e., Lagrangian approach), either focusing on larval stages (e.g., Mullon et al., 2002; Lett et al., 2007; Brochier et al., 2008) or growth (Megrey et al., 2007; Rose et al., 2007), or trying to describe the full cycle of the population (Ito et al., 2007; Okunishi et al., 2009; Xi, 2009) or even multi-species interactions (Travers et al., 2009). Comparatively, attempts to use the Eulerian approach for computing fluxes of fish density with advection-diffusion-reaction (ADR) equations are scarce (MacCall, 1990; Magnússon et al., 2004), despite that they can offer an advantageous framework for spatial population dynamics modeling and stock assessment studies due to limited number of parameters and continuous functions allowing the development of inverse models for optimization (Senina et al., 2008).

One such Eulerian modeling approach of population dynamics has been successfully developed for large pelagic species (i.e. tuna and tuna-like species) at basin-scale (Lehodey et al., 2008, 2010). The main features of this Spatial Ecosystem And POpulation DYnamics Model (SEAPODYM) are: (i) forcing by environmental data (temperature, currents, primary production and dissolved oxygen concentration), (ii) prediction of both temporal and spatial distribution of mid-trophic (micronekton) functional groups (Lehodey et al., 2010), (iii) prediction of both temporal and spatial distribution of age-structured predator populations, (iv) prediction of total catch and size frequencies by fleet when fishing effort is available, and (v) parameter optimization based on fishing data assimilation techniques (Senina et al., 2008).

We propose to adapt this model to small pelagic fish but there are several challenging issues. First, the model domain for tuna applications is at basin-scale and can run at relatively low resolution (typically 1 or 2 degree per month) for parameter optimization. Conversely, anchovy and sardine stocks in the HCS are coastal where the main factors driving the variability in eggs and larvae abundances (enrichment, concentration and retention), as identified by Bakun (1996), are likely strongly influenced by mesoscale activity. Thus, the approach requires a regional model with higher resolution and accurate representation of mesoscale physical and biogeochemical patterns. Another key difference is the type of available observation that can be used in the parameter estimation approach. For tuna fisheries, there is a long historical (50 years) fishing period with relatively well detailed (spatially-disaggregated) records of catch, effort and size distribution of catch over all the oceanic basin. For anchovies and sardines, historical catch data is often based on port landings (i.e., no detailed spatial information), and the fishing activity can be huge but short in time, thus providing a very scarce signal on yearly dynamics (Fréon et al., 2008). Nevertheless, important additional information has been collected by the Peruvian marine research institute IMARPE (Instituto del Mar del Peru, http://www.imarpe.pe). Eggs, larvae and fish acoustic biomass have been recorded during regular scientific cruises along the coast of Peru.

Given that spawning and larvae recruitment mechanisms largely determine the dynamics of the population (Beverton and Holt, 1957), and since most of the collected information is relevant for these early life stages, this first modeling effort to develop a SEAPODYM application to small pelagics focuses on the spawning habitat and larvae dynamics of the Peruvian anchovy. The first part of the study evaluates the climatological outputs of the coupled physical-biogeochemical model used to describe the oceanographic environment of the Peruvian upwelling region and to drive SEAPODYM. Then the spawning seasonality of anchovy is analyzed based on the dataset collected by IMARPE. Finally a sensitivity analysis is conducted to test which of the mechanisms proposed to control spawning habitat and larvae dynamics are valid for anchovies in

the HCS. These mechanisms include the response to the optimal temperature for larvae growth, the prey availability, the trade-off between presence/absence of prey and predators of larvae, and the retention or dispersion of larvae by currents.

5.2 Materials and Methods

5.2.1 Anchovy data

Since 1961, IMARPE conduct regular research cruises to sample anchovy eggs and larvae data. Eggs and larvae data were collected along the Peruvian coast between 5°S and 18°S using different nets. Here, only data collected with the Hensen net has been used. This net characterized by 0.33 m^2 mouth area and 300 μm mesh size was towed vertically from 50 meters to the surface.

Monthly climatological series were produced using all available data between 1961-2008 (159 cruises) for eggs and larvae. Corresponding maps were created by averaging eggs and larvae densities at the resolution of the grid model (1/6°, cf. model description below) after removing the data collected during the strongest El Niño events that result in large changes of anchovy distribution and movement (Checkley Jr et al., 2009). These periods were defined using the Niño1.2 sea surface temperature (SST) index provided by the NOAA Climate Prediction Center (http:/www.cpc.ncep.noaa.gov/). Using the 3-month running mean of Niño1.2 SST anomalies, we characterized the periods of strong El Niño events using five consecutive SST anomalies above 1°C. As a result, the following periods were removed: 04/1965-08/1965, 05/1972-01/1973, 06/1976-09/1976, 09/1982-10/1983, 02/1987-07/1987, 02/1992-06/1992, 04/1997-08/1998. For the analysis and comparison between predictions and observations, a regional stratification with several regions (called mask) has been created based on the observed distribution of eggs and larvae densities.

Additional information on the spawning seasonality of anchovy is provided by the gonadosomatic index (GSI), i.e., the ratio of fish gonad weight to body weight. GSI has been monitored by IMARPE regularly (Buitrón and Perea, 2000; Perea et al., 2011) and measured in the main fishing ports of Peru: Paita (5°S), Chimbote (09°S), Huacho (11°S), Callao (12°S), and Pisco (14°S) to produce weekly and monthly average, based on daily sampling of 25-30 adult females (size > 12 cm). The monthly climatology of GSI used in this study is built from a data series covering the period 1990 to 1999 (Buitrón and Perea, 2000, pers. com.). GSI is a good indicator of the beginning of the spawning season since ovaries increase in weight with the maturation process of oocytes and decrease after the release of eggs. However, the phenomena of atresia, i.e., the resorption of non spawned oocytes, can mask the real ending of spawning. Consequently GSI is the most reliable

during periods of minimal values (biological rest) and rapid increase (maturation before spawning).

5.2.2 Bio-physical environment

Physical and biological forcing fields (temperature, currents, dissolved oxygen concentration (O2), Primary Production (PP), and euphotic depth (Zeu) were simulated by the Regional Ocean Modeling System (ROMS; Shchepetkin and McWilliams, 2005) coupled to a biological model (PISCES; Aumont and Bopp, 2006). The model configuration for the Peruvian region has been described in previous publications (Echevin et al., 2008; Albert et al., 2010). The model domain covers the region 5°N-25°S and 90°W-69.5°W, at a spatial resolution of 1/6° with 30 layers. The surface layer thickness ranges from a minimum of 5 cm in a 50-m deep nearshore water column, to 4 m in a 4000-m deep water column, and bottom layer thickness range from 10 m to 730 m. Five-day average model outputs were recorded.

We used a climatological simulation forced by COADS climatological heat fluxes and a monthly climatology of Quikscat wind stress, as in Albert et al. (2010). At the open boundaries the model is forced by the dynamical fields and biogeochemical tracers from a monthly climatology of the ORCA2 OGCM simulation at 2° resolution over 1992-2000 (Albert et al., 2010). Ten years of spinup were produced to reach a statistical equilibrium. While the eddy-resolving model is forced repeatedly (each year) with the same monthly climatological atmospheric fluxes and boundary conditions, the mesoscale environment is slightly variable from year to year due to intrinsic nonlinear variability. One year of ROMS-PISCES simulation output was selected to drive the anchovy model.

Currents fields were averaged over the mixed-layer depth (issue from the ROMS-PISCES outputs) where eggs and larvae inhabit (Lett et al., 2007; Mathiesen, 1989). Larvae forage mainly on zooplankton except maybe on the very first feeding (van der Lingen et al., 2009). Since larvae's stage in the model is only 5 days, total primary production integrated over all the vertical depth is used as a proxy for larvae's food. Each simulation had the same resolution in time and space as the input fields.

To test the mechanisms of predation on larvae by the micronekton, the mid-trophic level model of SEAPODYM (Lehodey et al., 2010) has been used with this mesoscale physical-biogeochemical environmental forcing. Though the definition of functional groups of micronekton is based on time of development of organism, they can be roughly characterized by a size spectrum in the range of 2-20 cm dominated by crustaceans, fish, and cephalopods. This model includes 6 groups from surface to 1000 m with vertically migrant and non-migrant components (Lehodey et al., 2010). Since the coastal epipelagic group is largely dominated by anchovy population for which there is no strong evidence

of cannibalism, we only consider the mesopelagic and bathypelagic migrants coming into the surface at night weighted by an estimated fraction of time when predation is maximal, i.e., one hour at sun-set and one hour at sunrise (Lehodey et al., 2008).

5.2.3 SEAPODYM spawning habitat and larval recruitment

Sea surface temperature, currents and primary production fields of the bio-physical model are used to drive the spawning habitat and larvae movement in SEAPODYM. Several habitat definitions are investigated. In the abundant literature devoted to the problem of fish recruitment (e.g., see reviews in Rothschild (2000) and Govoni (2005)), the main processes that are proposed include the effect of temperature, the temporal and spatial availability of food for the larvae, and the predation of eggs and larvae. Finally, it is also proposed that the redistribution of larvae by the oceanic circulation can create retention of larvae in favorable areas (with lower natural mortality) or conversely move the larvae to unfavorable zones where the natural mortality will be higher Parrish et al., 1981; (Bakun, 1996). Of course, recruitment is also linked to the spawning stock (Beverton and Holt, 1957), but for small pelagic fishes in upwelling regions, the variability in spawning success and larvae survival is thought to be largely driven by environmental conditions (Cury and Roy, 1989; Cury et al., 1995; McFarlane et al., 2002; Brochier et al., 2008). Therefore, in this study, the stock-recruitment relationship has not been included and we simply assume that mature adults are present everywhere.

Different mechanisms have been included in SEAPODYM to simulate the spawning habitat (Hs) and larvae dynamics through functional relationships. The result is a relative index normalized between 0 and 1, thus independent of absolute calibration of primary production and energy transfer to micronekton groups. These mechanisms have been adapted from Lehodey et al. (2008) for anchovy species:

1 - The thermal spawning habitat that controls changes in the spatial extent of the favorable temperature for larval growth was estimated using a Gaussian distribution (Eq. 5.1):

$$f_1(T) = N(T_0^*, \sigma_0) \tag{5.1}$$

with T_0^* the optimal mean temperature and σ_0 the width of tolerance interval (standard deviation) for eggs and larvae stage (Fig. 2.1a). We consider that eggs and larvae temperature tolerance are identical. Sea surface temperature is used to define the thermal spawning habitat of eggs and larvae of anchovy and sardine.

2 - The spatial and temporal availability of the suitable prey. Anchovy larvae certainly need a minimum of accessible food to survive, but the relationship between survival and

food density is likely not linear due to saturation effect at high levels of prey density. These mechanisms were described by Holling (1959) and more particularly in the so-called Holling's type III functional response. However, for practical computational reasons, a different saturating (sigmoid) function is used, with two parameters (a and b). Since the distribution of the small zooplankton is linked to the phytoplankton, primary production is used as a proxy of larvae food.

$$f_2(PP) = \frac{1}{1 + \exp^{a*(PP-b)}} \tag{5.2}$$

3 - To take into account the predators of eggs and larvae, we defined a decreasing sigmoid function (Eq. 5.3) of habitat with increasing biomass of predators (i.e. the biomass of migrant mesopelagics and bathypelagic organisms) on eggs and larvae in the upper layer (Fig. 5.1):

$$f_3(Pred) = \frac{1}{1 + \exp^{c*(Pred-d)}} \tag{5.3}$$

We defined the prey-predator trade-off as the product of the prey function (Eq. 5.2) and the predator function (Eq. 5.3).

4- The redistribution of larvae by currents leading to higher or lower mortality according to the retention in favorable habitat or the drift in unfavorable habitat is included in the treatment of the spatial dynamics using a system of Advection-Diffusion-Reaction equations (see Appendix A, for more details, see Lehodey et al. (2008) and Senina et al. (2008)). The average mortality coefficient for anchovy larvae was set to 0.378 $month^{-1}$, i.e. 5 times the average coefficient used for the adult population by Cubillos et al. (2002).

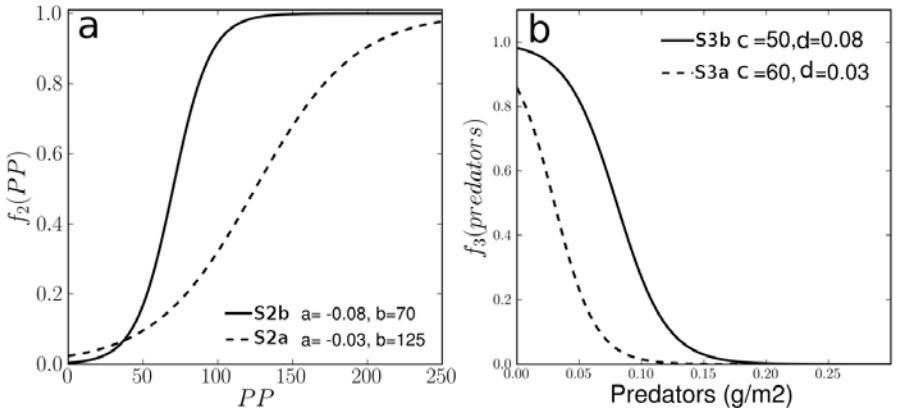

Figure 5.1: Functional relationships used to define the spawning habitat: functions of prey availability (a) and predation (b).

Larvae transport by currents with associated mortality is computed for a time step of five days after spawning, i.e., roughly corresponding to the estimated mean age of larvae collected in a size range between 3 and 6 mm (P. Ayón pers. com.), and based on a length of hatching of 2 mm and a growth function (Marzloff et al., 2009).

5.2.4 Simulations

We performed 10 simulation experiments with different mechanisms as described in Table 5.1. Simulation results are evaluated against observed seasonal and spatial variability of the GSI, eggs and larvae.

The sensitivity of the spawning habitat and larvae dynamics to temperature was tested with a parameterization of the Gaussian function allowing to have contrasted narrow and wide definitions of the thermal range for spawning. The overall thermal habitat of anchovy in the HCS is in the range 8-24.8°C when combining observations from Peruvian and Chilean regions (Bertrand et al., 2008; Gutiérrez et al., 2008; Tarifeño et al., 2008). However when selecting eggs data in the Peruvian region only (Fig. 2), the sea surface temperature range for spawning habitat off Peru is reduced to 15°C - 23°C. Effect of temperature on spawning habitat is explored with two parameterizations. The first experiment (S1A) used an optimum temperature of 18°C and a narrow standard deviation of 1°C. The second experiment (S1B) used the same optimum temperature of 18°C but with a large standard deviation of 4°C, leading to a habitat index higher than 0.5 for temperature between 13.3°C and 22.7°C.

Simulations	Mechanisms	T	σ	a	b	c	d	Function
S1a	Temperature	18	1	-	-	-	-	$H_s = f_1(T)$
S1b		18	4	-	-	-	-	
S2a	Prey availability	-	-	-0.03	125	-	-	$H_s = f_2(PP)$
S2b		-	-	-0.08	70	-	-	
S3a	Prey-predator trade-off	-	-	-0.08	70	60	0.03	$H_s = f_2(PP) * f_3(Pred)$
S3b		-	-	-0.08	70	50	0.08	
S4a	Temperature and Prey availability	18	4	-0.03	125	-	-	$H_s = f_1(T) * f_2(PP)$
S4b		18	4	-0.08	70	-	-	
S5a	Temperature and Prey-predator	18	4	-0.08	70	60	0.03	$H_s = f_1(T) * f_2(PP) * f_3(Pred)$
S5b	trade-off	18	4	-0.08	70	50	0.08	

Table 5.1: Parameters tested in each set of simulation

The larvae feeding ground is characterized by concentration between a maximum of 250 $mmolC.m^{-2}.d^{-1}$ and roughly 70 $mmolC.m^{-2}.d^{-1}$ along the continental shelf break that delineate the upwelling region where anchovy eggs and larvae are the most abundant. Two parameterizations of the prey function were used (Fig. 5.1 a). The first one (S2A, a = -0.03, b=125) has a slow linear increase close to Holling I relationship with a threshold value of PP equal to 125 $mmolC.m^{-2}.d^{-1}$ for Hs=0.5, i.e. the central value of the observed range. The second (S2B) simulates a minimum threshold value in food density (70 $mmolC.m^{-2}.d^{-1}$) before spawning success increases rapidly with food density (S2B: a = -0.08, b= 70). Using the S2B parameterization for the prey function, the prey-predator trade-off mechanism is explored with two parameterizations producing a strongly negative impact as soon as predator densities exceed a low value (S3A: c=60,

d=0.03) or intermediate value (S3B: c=50, d=0.08) of the predators present in the larvae's habitat (Fig. 5.1 b).

Finally, combinations of these different mechanisms were explored. In S4, temperature is combined to the prey availability effect, i.e., the product of temperature (S1B) with two different parameterizations of the prey availability function (S2A, S2B). In S5, temperature is combined to the prey-predators trade-off. i.e., the product of temperature (S1B) with two different parameterizations of the prey-predator trade-off function (S3A, S3B).

5.3 Results

5.3.1 Eggs and larvae distributions

Anchovy eggs and larvae were collected all along the Peruvian coast between 5°S and 18°S, though they were mainly found between 6°S and 14°S. Highest concentrations are observed near the coast, at the edge of the continental shelf (Fig. 5.2).

Though eggs were found all along the coast without particular concentration patterns, larvae appeared frequently in higher density in the northern region from 6°S to 9°S (Fig. 5.2). A possible enrichment by larval drift from nearby regions or better survival rates could explain this favorable region. Lett et al. (2007) already showed with an Individual-Based Model approach that this northern region was an area of accumulation, due to northward surface transport by the Peru Coastal Current.

Based on the observed density of eggs and larvae, a spatial aggregation has been defined with a mask of five coastal and offshore regions (Fig. 5.3). The offshore regions are the offshore zone (OZ) and the transition zone (TZ), and the coastal zones are the Northern coastal zone (NCZ), the Central coastal zone (CCZ), and the Southern coastal zone (SCZ). Northern coastal zone is the region where the highest eggs and larvae densities occurred while offshore zone located offshore is used to characterize the null habitat, with absence of eggs and larvae. Simulation outputs are integrated over these areas and by month to be compared to observation.

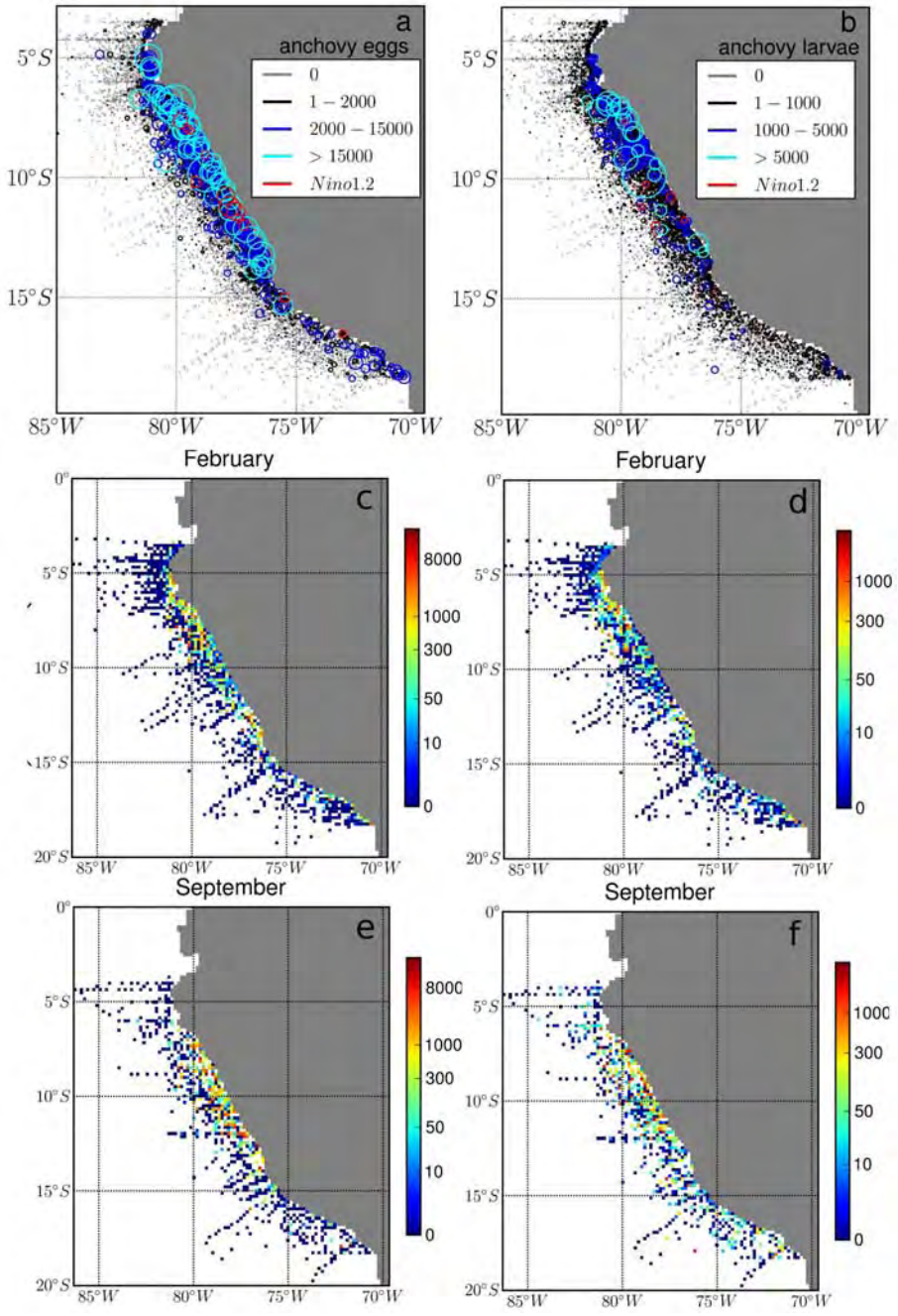

Figure 5.2: Top: Composite distribution maps for eggs (left) and larvae (right) of anchovy collected by the Instituto del Mar del Peru over the period 1961-2008. Raw data with circles radius proportional to density values; the largest circle corresponding to (a) 107 376 $eggs.m^{-2}$, (b) 84 939 $larvae.m^{-2}$. Monthly climatology for February (c,d) and September (e,f).

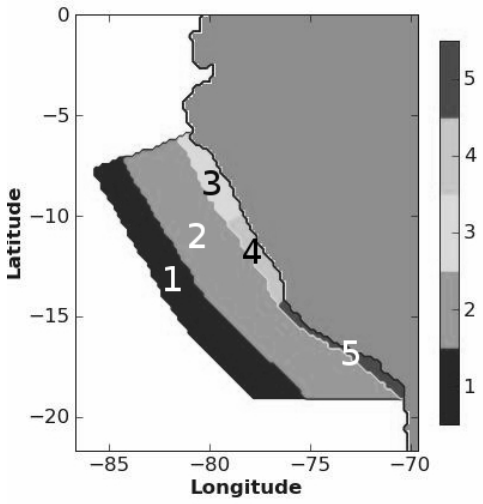

Figure 5.3: Mask used in the analysis with five coastal and offshore regions. OZ=offshore zone (1), TZ= transition zone (2), NCZ= Northern coastal zone (3), CCZ=Central coastal zone (4), SCZ=Southern coastal zone (5).

5.3.2 Seasonal variability of predicted environment

A climatology of observed sea surface temperature was built from satellite data (1992-2000, excluding 1997-1998 El Niño years) from the NOAA Optimum Interpolation 1/4 Degree Daily SST product (Reynolds et al., 2007). The oceanic region off Peru is characterized by a clear seasonal cycle of SST both in the coastal and offshore regions with a maximum in March, and a minimum in September (Fig. 5.4).

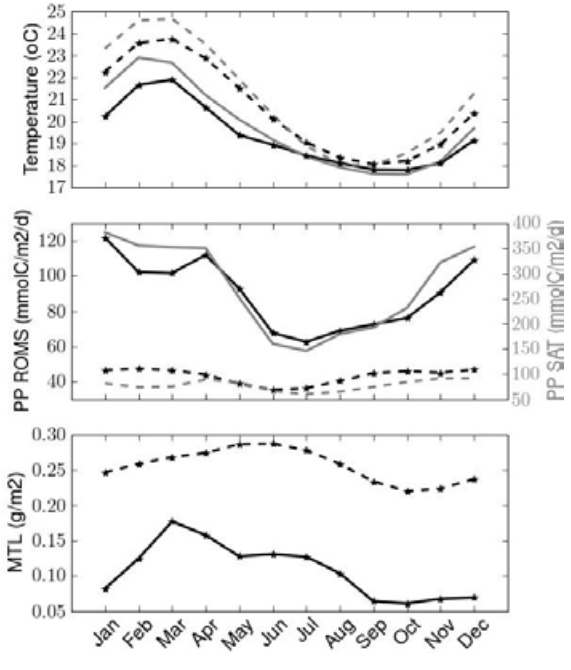

Figure 5.4: Observed and predicted seasonal variability in the Peruvian coastal regions (average over NCZ, CCZ, SCZ regions; solid lines) and offshore regions (average over regions OZ and TZ regions; dashed lines): (top) predicted mean temperature (black line) of the mixed layer is compared to observed (grey line) monthly satellite-derived SST climatology (see text for details); (middle) predicted primary production (PP) vertically-integrated (black line) and observed primary production derived from SeaWiFS satellite data (1999-2008), using the VGPM model of Behrenfeld and Falkowski, 1997; (bottom) predicted larvae predators in g of wet weight per square meter (bottom).

This minimum is linked to stronger winds and maximum upwelling intensity in the austral winter in the nearshore signal, and also to the reduced winter solar flux in the

offshore region. The temperature range is lower along the coast (17-21°C) than offshore (18-24°C) due to coastal upwelling. Simulated sea surface temperature showed a good seasonal cycle in agreement with observation, though winter minimum temperature occurs slightly later (Sept-Oct) than observed (August-Sept). Modelled temperature values in the coastal regions are underestimated from December to May by 1-2°C, but as we used a high temperature range, this had limited impact on our simulation results.

The seasonal cycle of predicted primary production (PP) match very well the observed seasonality (Fig. 5.4), i.e., increasing after September, peaking in January-March and decreasing after April. However, in comparison with satellite-derived primary production, the model underestimated the production by a factor of two (offshore) to three (coast) with a lower amplitude in the range of variability (Fig. 5.4). This discrepancy may be partly due to the computation of primary production based on the VGPM algorithm used for the general case of oceanic waters while coastal waters would require a specific parameterization (Behrenfeld and Falkowski, 1997). Furthermore, cloud filling procedure performed on satellite chlorophyll data can be also a source of overestimation of satellite-based primary production (Behrenfeld and Falkowski, 1997). Another comparison with in situ estimates of primary production (Messié et al., 2009) suggests an underestimation of 20% by the model (Albert et al., 2010). Also, modelled PP seems to be overestimated in the southern coastal zone (SCZ) in comparison to VGPM PP (not shown).

Both temperature and primary production show a strong coastal-offshore gradient, while temperature in the offshore region is also characterized by a decreasing gradient from north to south (Fig. 5.5).

The modelled biomass of larvae predators peaks in March-April in the coastal regions (Fig. 5.4), i.e. when primary production approaches its minimum. Simulated offshore seasonal cycles of primary production and micronekton biomass are in opposite phases as well, the predator biomass peaking in June when primary production is the lowest (Fig. 5.4). This contrasted situation is due to the large biomass of migrant mesopelagic and bathypelagic components predicted to inhabit the deep ocean. Interestingly, this leads to a higher larvae predator biomass offshore than near the coast, especially in September (Figs. 5.4 and 5). In this approach, the absolute biomass values of mesopelagic and bathypelagic migrant micronekton are not critical since we can always choose such parameters which define the change of spawning index between 0 and 1 within any range of micronekton. Predicted distribution of mesopelagic biomass is in accordance with acoustic data (Cornejo and Koppelmann, 2006; Gutiérrez et al., 2009; Gutiérrez et al., 2009). However, it would be essential to confirm the temporal shifts predicted between primary production and biomass of larvae predators that can produce contrasted environment for larvae survival (Figs. 5.4 and 5.5).

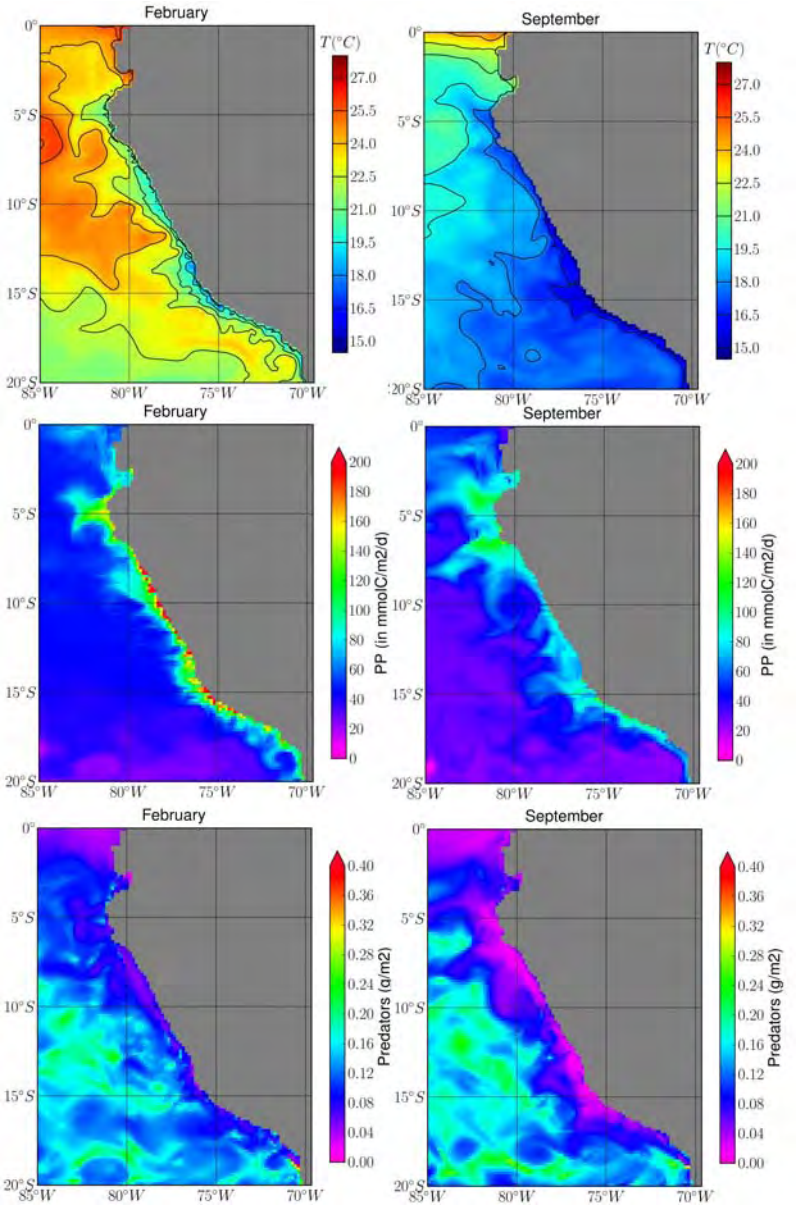

Figure 5.5: Average predicted spatial distribution in March and September of sea surface temperature (top), primary production (middle), and density of larvae predators in g of wet weight per square meter (bottom).

5.3.3 Observed seasonal variability in anchovy reproduction

Densities of eggs and larvae collected at sea are highly variable in time and space with large dispersion and very high proportion of null samples. Due to obvious non-Gaussian distributions of anchovy data, the seasonality was explored using non-parametric statistics. Mann Whitney test (also called Wilcoxon rank-sum test) was used to make a general pairwise comparison. This test allows to assess whether one of two groups that do not follow a normal distribution tends to have larger values than the other. The null hypothesis tested here is that the two groups are significantly different. We respect the null hypothesis if the p-value is lower than 0.05. This analysis allowed to identify the

periods of high and low abundance of eggs and larvae (Fig. 5.6).

We studied the monthly seasonality of all data for eggs, larvae (Fig. 5.6, top) as well as the seasonal variability by regions (Fig. 5.6, bottom).

Highest abundance of eggs occurred along the coast in region NCZ and CCZ, with density starting to increase after July, decreasing in October, and increasing again from December to February to peak in March. Although a major peak occurred in September, egg abundance was not significantly higher in September than in August (Mann-Whitney U test, p=0.4). Egg density decreased when moving towards the south (region SCZ) and was very low in the southern coastal region SCZ. No significant seasonality differences was observed (Mann-Whitney U test, p=0.7) probably because the standard error of the mean was very high. A single small peak occurred (in August) off coast in region TZ (Fig. 5.6), and almost no eggs were sampled far offshore in region OT.

The seasonal pattern for larvae was very similar to the one for eggs, with a high peak of density in September and a secondary peak between December and March (Fig. 5.6). The observed density was maximal in the north (region NCZ) and decreased southward followed by region CCZ (see also Fig. 5.2). In the southern coastal region (region SCZ) density increased in July, earlier than further north. In the transition zone (region TZ), the density of larvae was maximum in September and March.

The seasonality is marked by a clear reproductive rest period during austral autumn, from April to June (Fig. 5.6, bottom). Indeed, Mann-Withney test was always significant between austral summer and austral autumn as well as between austral autumn and austral winter. Furthermore, for most of the regions, differences were not significant between austral summer and austral winter, meaning that both periods can have the same abundances.

To complete this description, it is useful to compare the seasonal cycles of eggs and larvae with the change of monthly mean of GSI. Overall, GSI seasonal pattern agrees with observed seasonal cycles of eggs and larvae densities in northern and central coastal zones with a shift of one month. GSI starts increasing in June from mean value below 3 and peaks in September above 5 (Fig. 5.7). Then it decreases slightly but stays at rather high values (above 4) until February, showing the rapid decline in following months with the lowest values observed in April-May.

In summary, GSI with observed density of eggs and larvae show similar and coherent seasonal patterns with a shift of one month. GSI suggested a clear reproductive rest period taking place in April-May, with a maturation of gonads starting in June and high

peak in September. The spawning period continues but at lower intensity until February and GSI starts to decrease in March, while eggs and larvae are still abundant.

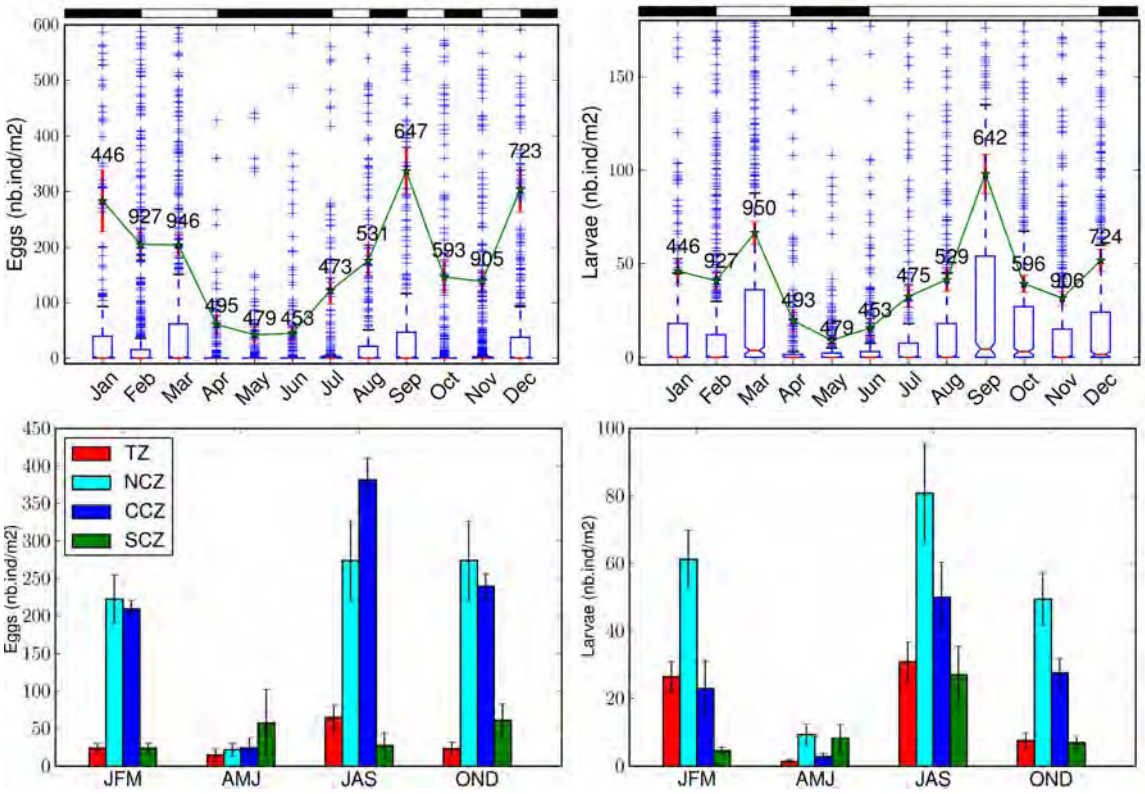

Figure 5.6: Seasonal variability in anchovy eggs, larvae and adult density. Eggs and larvae density are given in number of individual per square meters. 0.2% of outliers of eggs and larvae were removed. Adults abundance are in acoustical index (s_A). Top: Monthly average variability for all regions (Fig. 3.3.1) with box plots showing the median and the 25th and 75th percentiles, the mean (green line) and standard error of the mean (red bars), and the outliers (blue crosses). The numbers indicate the number of cells covered by observations (total number of cells for all regions being 2029). The result of the statistical analysis is illustrated with black and white bands where the black (white) color indicates non-significant (significant) difference between two consecutive (Mann-Withney U test; p>0.05). Bottom: Mean and standard error of the mean of eggs, larvae and adult density by season and region.

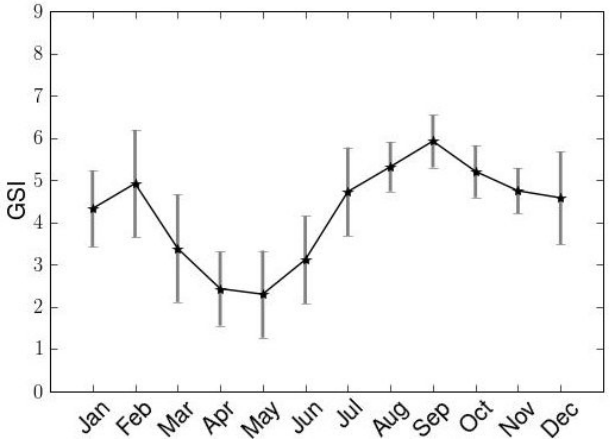

Figure 5.7: Monthly average of Gonado-Somatic Index of anchovy population. GSI average and standard error are computed from data collected by IMARPE between 1990 and 1999 after removing the period corresponding with the 1997-98 El Niño event.

5.3.4 Simulated spatial distributions

Predicted spatial distribution of spawning habitat and larvae densities are compared with observation and particularly the negative gradients from coast to offshore and from north to south.

The modelled spawning habitat and larvae density showed contrasted spatial distributions under different modelling assumptions (Fig. 5.8). Spatial distributions were compared for each month but we show here the results for September which displayed the most prominent features in terms of eggs and larvae density (Fig. 5.8). Using temperature only and considering a small thermal range (S1A) for eggs and larval survival, a negative gradient in eggs and larvae densities appears from the coast to offshore north of 7°S, but an opposite pattern arises south of 7°S. Along the coast, densities decrease from north to south which is in accordance with observations (Fig. 5.2). Note that, if we choose a cooler (warmer) temperature, the simulation (not shown) results in a worst (better) representation of the north-south gradient on the coast but a better (worst) representation of the coast-offshore gradient. This general feature persists also in the second parameterization (S1B), where the favorable habitat simply extends further offshore due the wider thermal range. As a result, the model predicts a high density of larvae offshore with more or less extended offshore distribution (Fig. 5.8). Clearly, temperature alone does not allow to obtain a distribution of the spawning habitat and larvae distribution that is consistent with the observed one.

The prey function (Eq. 5.2) alone leads to better results, as both observed gradients from coast to offshore and from north to south are correctly predicted (Fig. 5.8). Nevertheless, the parameterization giving a slower habitat increase and no saturation with prey density (S2A) leads to a reduced habitat in September conflicting with observations, which show the peak of spawning in this month (Fig. 5.6). With strongly nonlinear response of habitat to the food concentration (S2B) the coast-offshore gradient and north-south gradient are closer to observed patterns. However, favourable spawning index predicted in the south (SCZ) in September (Fig. 5.8) and during all other months (not shown) does not seem realistic with regards to observations (Figs. 5.2 and 5.6). Predicted density north of 5°N also appears overestimated (Figs. 5.2 and 5.8). Note that larvae densities follow the same spatial patterns as eggs densities (S2C).

When combining prey and predator functions (S3) both coast-offshore and north-south gradients are reinforced with a spawning ground slightly more contracted to the coast, which is in better agreement with observations. The use of rapid predation effect even at low predator density (S3A) allows representing rather well the habitat for Austral winter, but suppresses all eggs and larvae in February-March, in contrast with the data. The use of predation effect starting from a larger predator densities (S3B) provides the

best compromise for all months (Fig. 5.8, S3B).

Finally, combining temperature effect with the prey availability (S4) or the prey-predator trade-off (S5) effects did not change significantly the general spatial structure described above with S2 and S3 respectively. Nevertheless, in both cases adding temperature effect improved the favourability of the northern regions compared with southern regions, in agreement with observed patterns, especially the decrease in eggs and larvae density south of 10°S (Fig. 5.2). The impact of temperature was particularly strong in February with less eggs and larvae (not shown).

Since larvae are only 5 days old, the impact of retention is not highly significant. Nevertheless, small differences appears between eggs and larvae patterns. Since eggs and early larvae are supposed here to inhabit the mixed layer, which correspond to the surface Ekman layer at the coast, they are generally advected offshore by currents. This is visible for all simulations on Fig 5.8.

5.3.5 Simulated seasonality

The modelled spawning habitat, assumed to be close to the eggs distribution, and larvae densities were averaged monthly in coastal (NCZ, CCZ, SCZ) and offshore (OZ, TZ) regions (Fig. 5.1) to show predicted seasonal cycles (Fig. 5.9). These predictions were evaluated against the main criteria identified from data above, and that are: i) a low abundance of eggs and larvae peaking in May-June, ii) a rapid increase in July peaking in August-September, prolonged by a period of relatively high abundance between October and February, then decreasing in April and iii) null or very low densities offshore in contrast to coastal densities. With temperature only, the favourable Jan.-Feb and unfavourable May-June seasons cannot be reproduced, and only one long favourable period can be predicted after April (coast) or May (offshore). The parameterization with a larger range of favourable temperature (S1B) allows obtaining a non null habitat from January to March (Fig. 5.9), but in this context the coast-offshore gradient is not well reproduced (Fig. 5.8 - S1B). Therefore, temperature alone does not allow simulating a plateau from October to February consistent with observations.

The prey availability alone cannot produce a high abundance peak in August-September, since the peak in primary production, the proxy used as larvae's food, is observed from December to April (Fig. 5.4). With near-linear habitat response to prey density (parameterization S2A), too low larvae density is predicted in September when the highest abundance is observed (Fig. 5.6). When habitat quickly becomes favorable at lower prey densities, (S2B and S2C), the predicted abundance of eggs and larvae starts to increase earlier but still too late with respect to observations and the minimum of abundance is predicted in June-August (Fig. 5.9), i.e. about one month later than

observed (Fig. 5.6).

In simulations S3A, prey-predator trade-off mechanism is introduced to obtain a rapid effect of predation even at low density of predators. In that case, highest spawning habitat values are predicted in September-December, i.e., starting one month later and finishing too early given the observed high density of eggs and larvae in January - March. With a moderate predation effect (S3B), the minimum is reached in June through August, which is similar to simulation S2B with the prey function alone and less shifted with respect to observed minimum(May-June) than S3A. In distinction from S2B however, the favourable spawning habitat last longer in S3B (April-May) than in observations, which show instead a rapid decline from March to April. The same conclusions apply for larvae. Without this secondary peak the general seasonal pattern would have been fairly close to the observed one.

By combining temperature with prey availability (S4) or prey-predator trade-off (S5), the main difference from corresponding S2 and S3 simulations is a decrease in habitat values and larvae densities in Feb-March, which does not suppress however the secondary peak in April-May (Fig. 5.9). From the five mechanisms analyzed, simulations combining temperature and either prey availability only (S4) and prey-predator (S5) effects gave the closest results to the observed temporal variability of eggs and larvae densities; the first one being better in reproducing the observed cycle in the beginning of the year, the second - in the second half. None however fit correctly the entire annual cycle. For spatial variability, there is a small advantage in the simulations S5 as they successfully predict the maximum of densities between September (Fig. 5.8) and January while keeping low density offshore. Nevertheless, they did not reproduce the observed minimum of densities in April - May.

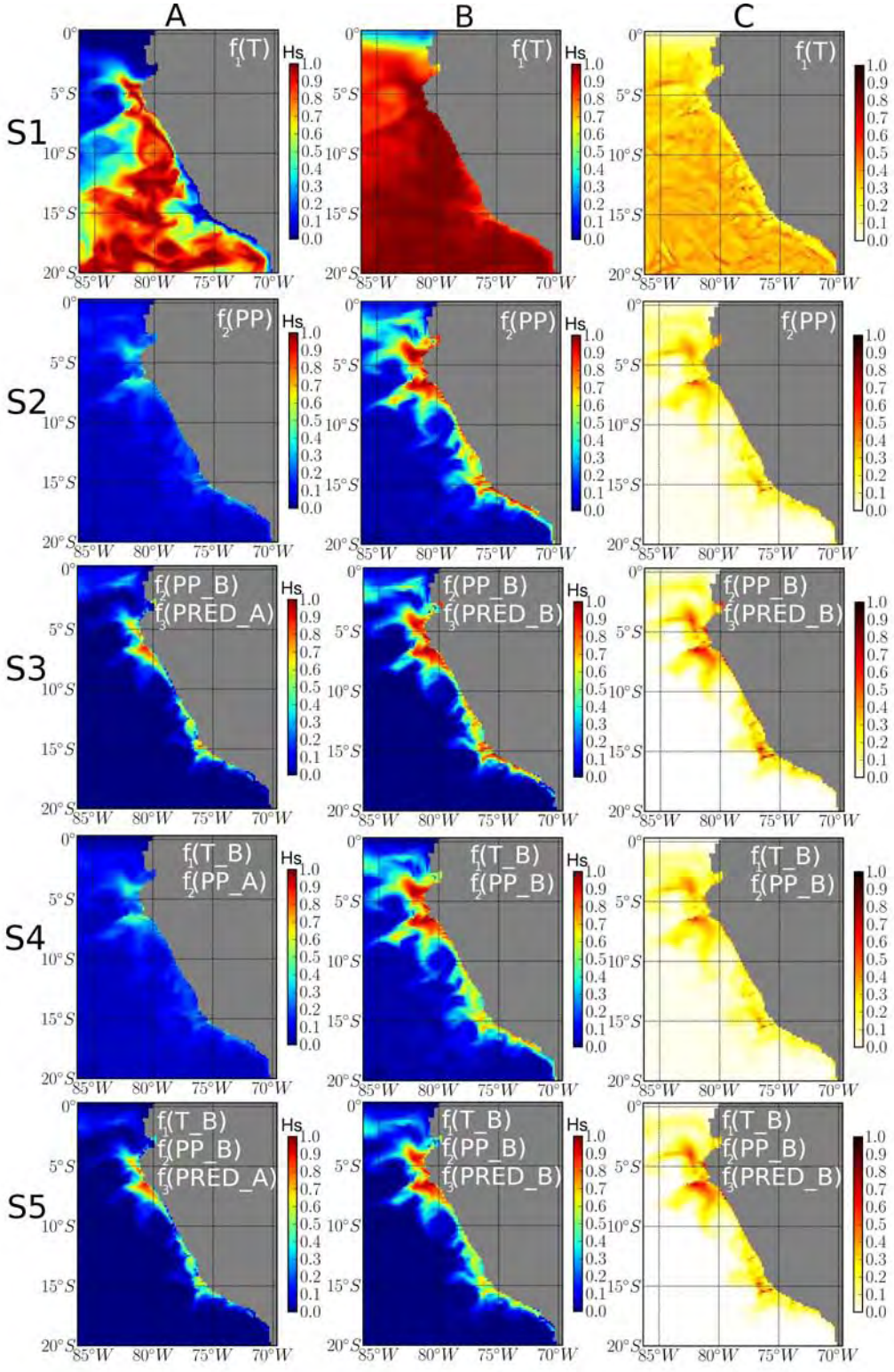

Figure 5.8: Spawning habitat (A, B) and larvae density (C) in September simulated with different mechanisms (Eqs. 1-3) and parameters (Table 1). S1: temperature function only (Eq. 5.1) with narrow thermal range (A) and wide thermal range (B); S2: prey function only (Eq. 5.2), with either (A) near-linear increase or (B) abrupt switch to habitat maximal value at low densities of preys; S3: predator-prey trade-off (Eq. 5.3) with strong(A) and weak (B) predation effect; S4: combination of S1B with S2A (A) or S2B (B); S5: combination of S1B with S3A (A) or S3B (B).

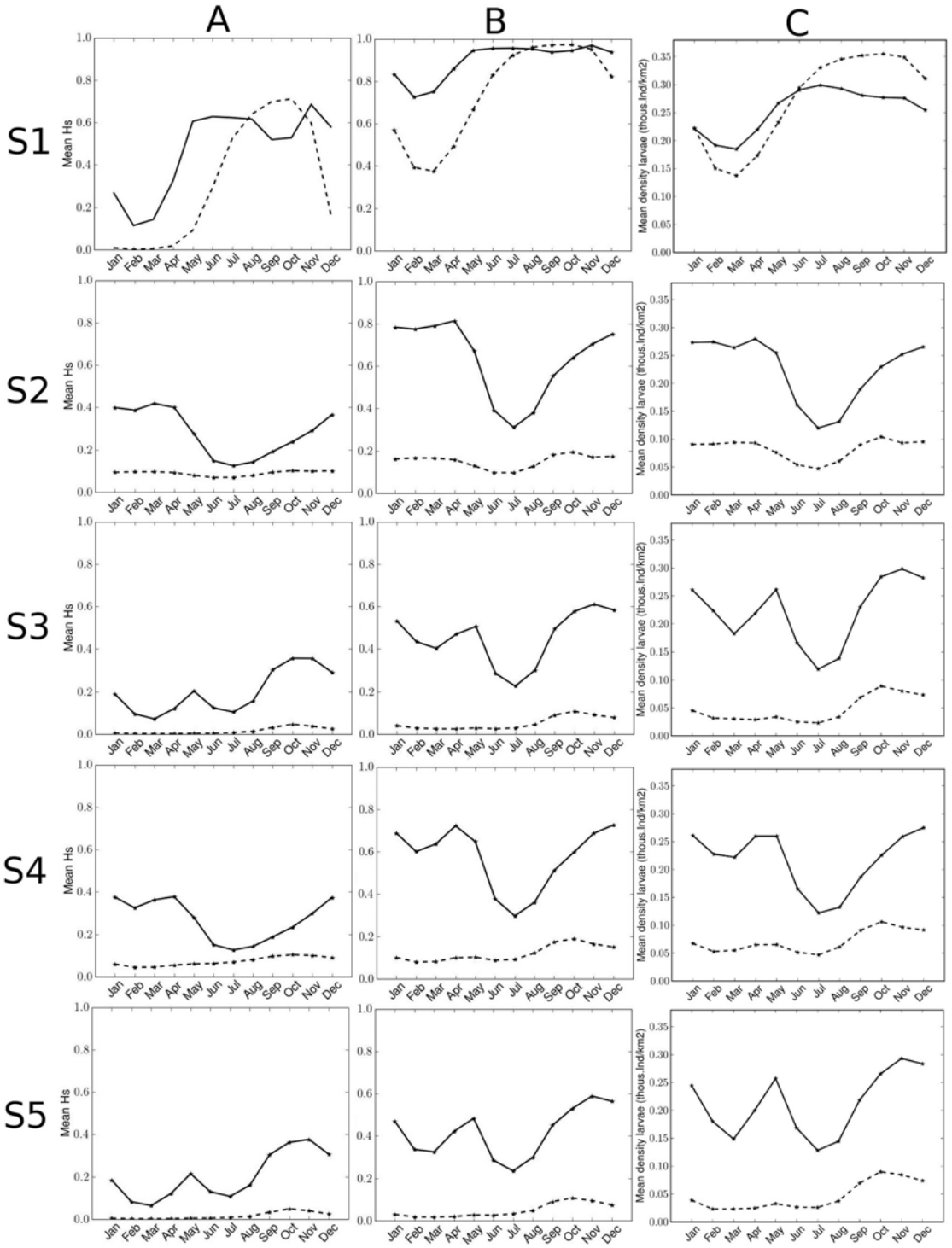

Figure 5.9: Seasonal variability of the simulated spawning habitat (A, B) and larvae density (C) in coastal regions (NCZ, CCZ, SCZ; continuous lines) and offshore regions (OZ and TZ; dotted lines). S1: temperature function only (Eq. 5.1) with narrow thermal range (A) and wide thermal range (B); S2: prey function only (Eq. 5.2), with either (A) near-linear increase or (B) abrupt switch to habitat maximal value at low densities of preys; S3: predator-prey trade-off (Eq. 5.3) with strong(A) and weak (B) predation effect; S4: combination of S1B with S2A (A) or S2B (B); S5: combination of S1B with S3A (A) or S3B (B).

5.4 Discussion

Spawning and larvae recruitment mechanisms largely determine the dynamics of the populations of small pelagic fish. In upwelling regions, the variability in the spawning success of these species is strongly related to environmental conditions (Bakun, 1996; Cury and Roy, 1989; Cole, 1999), likely dominating density dependence effects and particularly the stock-recruitment relationship, excepted at very low levels of spawning biomass (Myers et al., 1999; Beverton and Holt, 1957).

In this study, we have considered the role of environmental conditions only. In addition, the analysis focused on seasonal climatological variability. Based on observations, it is characterized by the following prominent features in eggs and larvae density:

- Spatially, a negative coast-offshore gradient, a pronounced negative north-south gradient for larvae, though less pronounced for eggs.
- Temporally, a seasonal pattern with a clear rest period between April and June, a peak of eggs and larvae abundance around September, and the presence of relatively high densities between December and March (see also Bouchon et al., 2000; Perea et al., 2011).

As far as the spatial structures are concerned, the simulations combining the temperature effect with either prey spatial and temporal availability or the predator-prey trade-off mechanisms gave the best results. The latter seems better to some extent as it predicted very unfavourable habitat offshore and more concentrated coastal spawning grounds coinciding with known spawning areas both in the north and south regions. The simulations showed less satisfactory agreement with the seasonal cycles of egg and larvae densities, all of them having difficulties to simulate a rapid decrease in April and minimum values in May-June. Though many studies considered temperature as a direct key factor driving anchovy population and recruitment (e.g., Muck, 1989), this study suggests that temperature alone cannot explain the spatial and temporal patterns of spawning habitat (eggs) and larvae of anchovy in the HCS. This result is not surprising since anchovy, which has to cope with a variable environment, is highly plastic (Espinoza and Bertrand, 2008; Gutiérrez et al., 2008) and distributes and spawns over a very large range of temperature. Nevertheless, temperature has likely an impact on the larval recruitment in combination with other factors.

Our results also show that in the HCS there is no seasonal coincidence between the peaks of primary production and larvae. Since anchovy larvae feed on small zooplankton developing almost immediately after phytoplankton, the primary production can be seen as a good proxy for prey of larvae. Thus, the match-mismatch mechanism as defined by Cushing (1975, 1990) is not at play in the HCS since if the match existed, maximum peak

of larvae should occur in March-April in contradiction to observations which present a maximum peak in September and very little spawning in April. This can be due to the fact that this system is characterized by an all-year-long upwelling and that prey are probably never limiting excepted during strong El Niño events or during centennial periods of low productivity, e.g. the little ice age (Gutiérrez et al., 2009). Nevertheless, the availability of prey, combined with other processes, is needed to approach temporal and more particularly spatial patterns.

Larvae predators were assumed to be mainly migrant mesopelagic and bathypelagic organisms. We did not consider eggs and larvae predation by epipelagic organisms which in the coastal domain would consist mainly of anchovy. A recent study (Espinoza and Bertrand, 2008) suggested that there is no evidence of strong cannibalism in Peruvian anchovy. However other species of fish or invertebrates, such as the squat lobster, Pleuroncodes monodon (see Gutiérrez et al., 2008), may exert strong predation on anchovy first stages. It would be important to confirm the level of coastal predation since it may be an interesting key mechanism explaining the abundance of this species and its domination in the system.

Alternatively, there are potential sources of errors in the observed and modelled climatologies. The model climatology was created using an atmospheric forcing for the recent period (2000-2006) but boundary forcing corresponding to the period 1992-2000, thus potentially leading to a bias. All available eggs and larvae data for the period 1961 to 2008, after removing exceptional El Niño events, were used to construct climatological series to maximize data coverage. Over this 50-year period, seasonal characteristics in the HCS may have changed, especially in a regional climate change context. Finally, the climatology made with data from such a long period, with cruises carried out with different aims, different coverage and different research teams may have introduced additional errors. Furthermore, we should consider that the more eggs and larvae collected in a net sample, the lower the relative error due to double or missing counts. Since anchovies are more dispersed in winter than in summer (Lett et al., 2007), when the intensity of the upwelling is weaker, data could be partially biased.

In addition to underestimated variability noted above, the physical modeling framework has other sources of uncertainty. The ROMS-PISCES model configuration providing environmental conditions at spatial resolution of 1/6° simulated fairly well the average seasonal cycles of temperature and primary production, but there is a critical lack of observations to validate the predicted currents circulation and the overall mesoscale structure very close to the shore. Preliminary results testing larval drift using a higher resolution configuration (1/12°) of the model suggested that higher concentrations of larvae can be achieved in the northern coastal region (T. Brochier. pers. com.).

The predicted environmental variables are then averaged in the mixed-layer to be used with SEAPODYM, thus potentially affecting the spatial structures. Therefore, it would be useful to compare our results produced in an Eulerian framework with those from IBM simulations using the same predicted currents, but with a full vertical resolution and including the effect of eggs floatability (Brochier et al., 2008).

Despite all the possible sources of errors, bias and uncertainties listed above, the present study demonstrates that the SEAPODYM Eulerian modeling approach can be adapted to small pelagic species in a regional domain. This framework allows investigating the mechanisms controlling the success of larval recruitment, showing that the combination of a few simple concepts could reasonably explain the spatial repartition of eggs and larvae density, but still are not sufficient to explain observed seasonality. Therefore, the presence of adults in a full life cycle model could help to improve the predicted seasonality. Indeed, spatial and seasonal abundance estimates from acoustic sampling suggest that anchovy adult density is higher from February to March along the coast, either because habitat contraction and fish aggregation or due to the peak in recruitment. In both cases, higher concentration of adults would amplify the spawning success in the coastal habitat.

This model provides a useful tool to test various functional relationships as the ones we have tested in this study. The other key advantage is that it provides the formalism required for developing parameter optimization approach. This will be all the more straightforward since the adjoint code of the model has been already developed (Senina et al., 2008) and will require only minor changes. Indeed, this study showed that the combination of a few simple mechanisms with a limited number of parameters quickly leads to multitude of solutions that only rigorous parameter optimization approaches can help to reach.

Future studies will thus develop the data assimilation scheme needed in the model to use eggs and larvae data for parameter optimization. The effect of stock-recruitment relationship, possibly with the seasonal movement of adult fish may need to be included to increase the realism of simulations. Similar simulation experiments will be conducted for sardines and the comparison between both species, as well as the use of real time series including the interannual ENSO variability should greatly help in the evaluation of the model outputs. Finally it would be important to test for the effect of oxygen, which seems to be an important driver in fish population in the NHCS (Bertrand et al., 2011).

A Appendices

A.1 Continuous model of larvae spatial dynamics.

Let us denote N the larvae density determined at point (x,y) and time t (here after we will omit the notations of space and time). For brevity, we use divergence operator of a vector field $div(v^*) = \delta_x u + \delta_y v$ and $\Delta = divgrad$ for Laplacian of scalar field of population density. v^* denotes average of oceanic currents over the mixed layer depth, u the corresponding zonal current and v the corresponding meridional current.

The ADR equation used to describe dynamics of the anchovy larvae is:

$$dtN = -div(Nv^*) + d(\Delta N) - mN + S_0 \tag{5.4}$$

where d is a constant diffusion coefficient, fixed at $198\ m^2/s$, m the mortality rate, S_0 the source/sink (S/P) term. The larvae stage is divided in different cohorts of 5 days (time resolution of ROMS-PISCES input). At each time step, surviving eggs move in the first larvae cohort, while a new recruitment of eggs occurs in the egg cohort. At each 5 days time step (time resolution of physical input model), surviving larvae are transferred in the following oldest cohort. In the main part of this study, larvae transport by currents with associated mortality is computed for a time step of five days only after spawning, i.e., roughly corresponding to the estimated mean age of larvae collected in a size range between 3 and 6 mm (P. Ayón pers. com.), and based on a length of hatching of 2 mm and a growth function (Marzloff et al., 2009).

The equation 5.4 is complemented with Neumann zero-flux boundary conditions, discretized (see Sibert et al., 1999) and solved on a 10 nmi squared cell-centered grid and 5-days time step with help of alternate direction implicit method.

A.2 Spawning

We define the number of eggs in each cell of the grid at a given time as the product between H_s and a number R (Eq. 5.5). Theoretically, R, is linked to the number of mature fish present in the cell. However, in this study, as no adult are modelled, as a first hypothesis we considered that adult biomass is the same everywhere. R is therefore considered as a constant value which need to be parameterized. It allows in other terms to fit the absolute values of eggs density observed.

$$N_0 = R * H_S \tag{5.5}$$

In its more complex form, spawning habitat index (H_S) is the product of a temperature function ($f_1(T)$, Eq. 5.1), a prey function ($f_2(PP)$, Eq. 5.2) and a predator function ($f_3(Pred)$, Eq. 5.3):

$$H_S = f_1(T) * f_2(PP) * f_3(Pred) \tag{5.6}$$

A.3 Mortality

The average mortality coefficient for anchovy larvae μ was set to 0.378 mo^{-1}, i.e. 5 times the average coefficient used for the exploited adult population by Cubillos et al. (2002). However, to account for the effects of environmental variability, the local mortality rate varies with habitat index (Eq. 5.7).

$$m = \mu(1+\epsilon)^{1-2H_S} \tag{5.7}$$

where parameter ϵ=1.3, which corresponds to 130% increase of μ if Hs<0.5 and decrease otherwise. Thus, during bad conditions (no food, unfavourable temperature, or high predation) mortality rate can reach 0.87 mo^{-1}.

Bibliography

Albert, A., V. Echevin, M. Levy, and O. Aumont, 2010: Impact of nearshore wind stress curl on coastal circulation and primary productivity in the Peru upwelling system. *J. Geophys. Res*, **115**, C12033.

Aumont, O. and L. Bopp, 2006: Globalizing results from ocean in situ iron fertilization studies. *Global Biogeochemical Cycles*, **20**(2), GB2017.

Bakun, A., 1996: Patterns in the ocean: Ocean processes and marine population dynamics. *University of California Sea Grant, San Diego, California, USA, in cooperation with Centro de Investigaciones Biológicas de Noroeste, La Paz, Baja California Sur, Mexico*.

Bakun, A. and K. Broad, 2003: Environmental 'loopholes' and fish population dynamics: comparative pattern recognition with focus on el niño effects in the pacific. *Fisheries Oceanography*, **12**(4-5), 458–473.

Behrenfeld, M. J. and P. G. Falkowski, 1997: Photosynthetic rates derived from satellite-based chlorophyll concentration. *Limnology and oceanography* 1–20.

Bertrand, A., A. Chaigneau, S. Peraltilla, J. Ledesma, M. Graco, F. Monetti, and F. Chavez, 2011: Oxygen, a fundamental property regulating pelagic ecosystem structure in the coastal Southeastern tropical Pacific. *PlosOne*, **6**(12), e29558.

Bertrand, A., F. Gerlotto, S. Bertrand, M. Gutiérrez, L. Alza, A. Chipollini, E. Diaz, P. Espinoza, J. Ledesma, and R. Quesquon, 2008: Schooling behaviour and environmental forcing in relation to anchoveta distribution: An analysis across multiple spatial scales. *Progress in Oceanography*, **79**(2-4), 264–277.

Bertrand, A., M. Segura, M. Gutiérrez, and L. Vasquez, 2004: From small-scale habitat loopholes to decadal cycles: a habitat-based hypothesis explaining fluctuation in pelagic fish populations off peru. *Fish and Fisheries*, **5**(4), 296–316.

Beverton, R. and S. Holt, 1957: On the dynamics of exploited fish populations. *Fish. Invest. Ser. II, XIX HMSO, London* 1–533.

Bouchon, M., P. Ayón, J. Mori, C. Peña, P. Espinoza, and L. H. L. et al., 2000: Peruvian anchovy (*Engraulis ringens Jenyns*) Biology. *Bol. Inst. Mar Perú*, **25**.

Brochier, T., C. Lett, J. Tam, P. Fréon, F. Colas, and P. Ayon, 2008: An individual-based model study of anchovy early life history in the northern Humboldt Current system. *Progress in Oceanography*, **79**(2-4), 313–325.

Buitrón, B. and A. Perea, 2000: Aspectos reproductivos de la anchoveta peruana durante el periodo 1992-2000. *Bol. Inst. Mar Perú*, **19**, 45–54.

Chavez, F. P., A. Bertrand, R. Guevara-Carrasco, P. Soler, and J. Csirke, 2008: The northern humboldt current system: Brief history, present status and a view towards the future. *Progress in Oceanography*, **79**(2-4), 95–105.

Checkley Jr, D., P. Ayón, T. Baumgartner, M. Bernal, J. Coetzee, R. Emmett, R. Guevara-Carrasco, L. Hutching, L. Ibaibarriaga, H. Nakata, et al., 2009: *Habitats. In: D.M. Checkley Jr., J. Alheit, Y. Oozeki and C. Roy, Editors, Climate Change and Small Pelagic Fish, Cambridge University Press, Cambridge, pp. 12-44 (Chapter 3).*

Cole, J., 1999: Environmental conditions, satellite imagery, and clupeoid recruitement in the northern Benguela upwelling system. *Fish. Oceanogr.*, **8**(1), 25–38.

Cornejo, R. and R. Koppelmann, 2006: Distribution patterns of mesopelagic fishes with special reference to *Vinciguerria lucetia* Garman 1899 (Phosichthyidae: Pisces) in the Humboldt Current Region off Peru. *Mar. Biol.*, **149**, 1519–1537.

Cubillos, L., D. Bucarey, and M. Canales, 2002: Monthly abundance estimation for common sardine *Strangomera bentincki* and anchovy *Engraulis ringens* in the central southern area off Chile (34-40S). *Fish Res*, **57**, 117–130.

Cury, P. and C. Roy, 1989: Optimal environmental windows and fish recruitment success in upwelling areas. *Can. J. Fish. Aquat. Sci.*, **44**, 670–680.

Cury, P., C. Roy, R. Mendelssohn, A. Bakun, D. Usby, and R. Parrish, 1995: Moderate is better: exploring nonlinear climatic effects on the Californian anchovy (*Engraulis mordax*). *Can. J. Fish. Aquat. Sci.*, **121**, 417–424.

Cushing, D., 1975: Marine ecology and fisheries. *Cambridge Univ. Press, Cambridge, England* 278.

Cushing, D., 1990: Plankton production and year-class strength in fish populations: an update of the match/mismatch hypothesis. *Adv. Mar. Biol.*, **9**, 295–354.

Echevin, V., O. Aumont, J. Ledesma, and G. Flores, 2008: The seasonal cycle of surface chlorophyll in the peruvian upwelling system: A modelling study. *Progress in Oceanography*, **79**(2-4), 167–176.

Espinoza, P. and A. Bertrand, 2008: Revisiting Peruvian anchovy (*Engraulis ringens*) trophodynamics provides a new vision of the Humboldt Current system. *Progress in Oceanography*, **79**(2-4), 215–227.

Fréon, P., M. Bouchon, C. Mullon, C. Garcia, and M. Ñiquen, 2008: Interdecadal variability of anchoveta abundance and overcapacity of the fishery in Peru. *Progress in Oceanography*, **79**(2-4), 401–412.

Govoni, J., 2005: Fisheries oceanography and the ecology of early life histories of fishes: a perspective over fifty years. *Sci. Mar.*, **69 (Suppl. 1)**, 125–137.

Gutiérrez, D., A. Sifeddine, D. B. Field, L. Ortlieb, G. Vargas, F. P. ChÃ¡vez, F. Velazco, V. Ferreira, P. Tapia, R. Salvatteci, et al., 2009: Rapid reorganization in ocean biogeochemistry off Peru towards the end of the Little Ice Age. *Biogeosciences*, **6**(5), 835–848.

Gutiérrez, M., P. Castillo, S. Peratlilla, M. Flores, C. Maldonado, and L. Vasquez, 2009: Distribution and biomass of some Peruvian pelagic resources in summer 2004. *Inf. Inst. Mar Perú*, **36**, 145–152.

Gutiérrez, M., S. Peratlilla, and L. Vasquez, 2009: Distribution and abundance of Peruvian pelagic and mesopelagic species; Spring 2006. *Inf. Inst. Mar Perú*, **36**, 65–80.

Gutiérrez, M., A. Ramirez, S. Bertrand, O. Morón, and A. Bertrand, 2008: Ecological niches and areas of overlap of the squat lobster "munida" (*Pleuroncodes monodo*n) and anchoveta (*Engraulis ringens*) off Peru. *Progress in Oceanography*, **79**(2-4), 256–263.

Gutiérrez, M., G. Swartzman, A. Bertrand, and S. Bertrand, 2007: Anchovy (*Engraulis ringens*) and sardine (*Sardinops sagax*) spatial dynamics and aggregation patterns in the Humboldt Current ecosystem, Peru, from 1983-2003. *Fisheries Oceanography*, **16**(2), 155–168.

Holling, C., 1959: Some characteristics of simple type of predation and parasitism. *Can. Entomol.*, **91**, 385–398.

Ito, S., B. Megrey, M. Kishi, D. Mukai, Y. Kurita, Y. Ueno, and Y. Yamanaka, 2007: On the inter-annual variability of the growth of Pacific saury (*Cololabis saira*): a simple 3-box model using NEMURO. FISH. *Ecological modelling*, **202**(1-2), 174–183.

Lehodey, P., R. Murtugudde, and I. Senina, 2010: Bridging the gap from ocean models to population dynamics of large marine predators: A model of mid-trophic functional groups. *Progress In Oceanography*, **86**(1-2), 302–315.

Lehodey, P., I. Senina, and R. Murtugudde, 2008: A spatial ecosystem and populations dynamics model (SEAPODYM) - Modeling of tuna and tuna-like populations. *Progress in Oceanography*, **78**(4), 304–318.

Lehodey, P., I. Senina, J. Sibert, L. Bopp, B. Calmettes, J. Hampton, and R. Murtugudde, 2010: Preliminary forecasts of pacific bigeye tuna population trends under the a2 IPCC scenario. *Progress In Oceanography*, **86**(1-2), 302–315.

Lett, C., P. Penven, P. Ayon, and P. Fréon, 2007: Enrichment, concentration and retention processes in relation to anchovy (*Engraulis ringens*) eggs and larvae distributions in the northern Humboldt upwelling ecosystem. *Journal of Marine Systems*, **64**(1-4), 189–200.

MacCall, A., 1990: *Dynamic geography of marine fish populations*. Washington Sea Grant Program Seattle, WA.

Magnússon, K., S. Sigurdsson, P. Babak, S. Gudmundsson, and E. Dereksdottir, 2004: A continuous density Kolmogorov type model for a migrating fish stock. *Discrete and Continuous Dynamical Systems Series B*, **4**, 695–704.

Marzloff, M., Y. J. Shin, J. Tam, M. Travers, and A. Bertrand, 2009: Trophic structure of the peruvian marine ecosystem in 2000-2006: Insights on the effects of management scenarios for the hake fishery using the IBM trophic model osmose. *Journal of Marine Systems*, **75**(1-2), 290–304.

Mathiesen, O., 1989: *Adaptation of the anchoveta (Engraulis ringens) to the Peruvian upwelling ecosystem. In: The Peruvian upwelling ecosystem: dynamics and interactions*. The WorldFish Center, Ed. by D. Pauly, P. Muck, J. Mendo, and I. Tsukayama (Eds). ICLARM Conference Proceedings, 220-234.

McFarlane, G., P. Smith, T. Baumgartner, and J. Hunter, 2002: Climate variability and Pacific sardine population and fisheries. *Am. Fisheries Soc. Symp.*, **32**, 195–214.

Megrey, B., K. Rose, R. Klumb, D. Hay, F. Werner, D. Eslinger, and S. Smith, 2007: A bioenergetics-based population dynamics model of Pacific herring (*Clupea harengus pallasi*) coupled to a lower trophic level nutrient-phytoplankton-zooplankton model: description, calibration, and sensitivity analysis. *Ecological Modelling*, **202**(1-2), 144–164.

Messié, M., J. Ledesma, D. Kolber, R. Michisaki, D. Foley, and F. Chavez, 2009: Potential new production estimates in four eastern boundary upwelling ecosystems. *Prog. Oceanogr.*, **83**, 151–158.

Muck, P., 1989: *Relationships between anchoveta spawning strategies and the spatial variability of sea surface temperature off Peru.* In: Pauly, D., Muck, P., Mendo, J., Tsukayama, I. (Eds.), The Peruvian Upwelling Ecosystem: Dynamics and Interactions. ICLARM Conference Proceedings, 168-1973.

Mullon, C., P. Cury, and P. Penven, 2002: Evolutionary individual-based model for the recruitment of anchovy (*Engraulis capensis*) in the southern Benguela. *Canadian Journal of Fisheries and Aquatic Sciences*, **59**(5), 910–922.

Myers, R., K. Bowen, and N. Barrowman, 1999: Maximum reproductive rate of fish at low population sizes. *Canadian Journal of Fisheries and Aquatic Sciences*, **56**, 2404–2419.

Okunishi, T., Y. Yamanaka, and S. Ito, 2009: A simulation model for Japanese sardine (*Sardinops melanostictus*) migrations in the western North Pacific. *Ecological Modelling*, **220**(4), 462–479.

Parrish, R., C. Nelson, and A. Bakun, 1981: Transport mechanisms and reproductive success of fishes in the california current. *Biolog. Oceanogr.*, **2**, 175–203.

Perea, A., C. Peña, R. Oliveros-Ramos, B. Buitrón, and J. Mori, 2011: Potencial production of eggs, recruitment and reproductive season fishery closure of Anchoveta (*Engraulis ringens Jenyns*, 1842): implications for fisheries management. *Ciencias Marinas*, **37**(42).

Reynolds, R. W., T. M. Smith, C. Liu, D. B. Chelton, K. S. Casey, and M. G. Schlax, 2007: Daily (high-resolution-blended) analyses for sea surface temperature. *Journal of Climate*, **20**(22), 5473–5496.

Rose, K., F. Werner, B. Megrey, M. Aita, Y. Yamanaka, D. Hay, J. Schweigert, and M. Foster, 2007: Simulated herring growth responses in the Northeastern Pacific to historic temperature and zooplankton conditions generated by the 3-dimensional NEMURO nutrient-phytoplankton-zooplankton model. *Ecological modelling*, **202**(1-2), 184–195.

Rothschild, B., 2000: Fish stocks and recruitment?: the past thirty years. *ICES J. Mar. Sci.*, **57**, 191–201.

Senina, I., J. Sibert, and P. Lehodey, 2008: Parameter estimation for basin-scale ecosystem-linked population models of large pelagic predators: Application to skipjack tuna. *Progress in Oceanography*, **78**(4), 319–335.

Shchepetkin, A. F. and J. C. McWilliams, 2005: The regional oceanic modeling system (ROMS): a split-explicit, free-surface, topography-following-coordinate oceanic model. *Ocean Modelling*, **9**(4), 347–404.

Sibert, J. R., J. Hampton, D. A. Fournier, and P. J. Bills, 1999: An advection-diffusion-reaction model for the estimation of fish movement parameters from tagging data, with application to skipjack tuna (*Katsuwonus pelamis*). *Canadian Journal of Fisheries and Aquatic Sciences*, **56**(6), 925–938.

Swartzman, G., A. Bertrand, M. Gutiérrez, S. Bertrand, and L. Vasquez, 2008: The relationship of anchovy and sardine to water masses in the peruvian humboldt current system from 1983 to 2005. *Progress in Oceanography*, **79**(2-4), 228–237.

Tarifeño, E., M. Carmona, A. Llanos-Rivera, and L. Castro, 2008: Temperature effects on the anchoveta *Engraulis ringens* egg development: do latitudinal differences occur ? *Environ Biol. Fish*, **81**, 387–395.

Travers, M., Y. Shin, S. Jennings, E. Machu, J. Huggett, J. Field, and P. Cury, 2009: Two-way coupling versus one-way forcing of plankton and fish models to predict ecosystem changes in the Benguela. *Ecological Modelling*, **220**(21), 3089–3099.

van der Lingen, C., A.Bertrand, A. Bode, R. Brodeur, L. Cubillos, P. Espinoza, K. Friedland, S. Garrido, X. Irigoien, T. Miller, C. Möllmann, R. Rodriguez-Sanchez, H. Tanaka, and J. A. Y. O. C. R. A. Temming. Trophic dynamics, D.M. Checkley, 2009: *Trophic dynamics in "Climate Change and Small Pelagic Fish"*, volume 79. Cambridge University Press, Cambridge, UK.

Xi, Y., 2009: *Modeling the seasonal and interannual variability of Peruvian anchovy (Engraulis ringens) population dynamics: linking environmental conditions with fish.* PhD thesis, The University of Maine.

Chapter 6

Parameter estimation in data assimilation framework

Sommaire

6.1 Introduction to data assimilation concepts 170
6.2 Bayesian Approach 171
6.3 Global optimization approach: Grid search method 174
6.4 Local optimization approach: SEAPODYM framework 175
6.4.1 Adjoint methodology 177
6.4.2 Parameter scaling and bounds 179
6.4.3 Analysis of the quality of the estimate: Hessian Information 181
6.5 Implementation of a new likelihood function on SEAPODYM 182
6.5.1 Observations and predictions 182
6.5.2 Cost function 183
6.5.3 Initial conditions 184
6.5.4 Sensitivity analysis 185
6.5.4.1 Sensitivity of model to parameters 185
6.5.4.2 Sensitivity of objective function to parameters 185
6.6 Selecting the distribution of the likelihood function 186
Bibliography 192

6.1 Introduction to data assimilation concepts

The optimization method in SEAPODYM has been implemented by Senina et al. (2008) using fishing data. The authors used the maximum likelihood estimator (MLE) to estimate the set of parameters that minimizes the difference between predictions and observations. Here, we need to incorporate additional information on fish abundance and distribution from acoustic data and eggs and larvae campaigns.

This chapter presents a brief introduction to data assimilation methods used in marine sciences, which are needed to understand the implementation of the new cost function applied for early life stages of anchovy and sardine in SEAPODYM as well as the different validation procedures.

Data assimilation (DA) can be defined as the ensemble of techniques which combines in an optimal way (which will be defined later) the information about the system comprising:

- the physical information (observations)
- the mathematical information (model)
- statistical errors: errors in the model and errors in observations

This optimal way depends on the problem being studied. Indeed, the aims of data assimilation are multiple. For instance, it can be used to improve hindcast, nowcast or forecast from numerical modeling (more often used in oceanography and meteorology) (Ghil, 1989; Brasseur et al., 2005), to identify systematic errors in models (Griffith and Nichols, 2000), to estimate unobserved variables (Tarantola, 2005; Dunn and Hernandez, 2009), to estimate the parameters used in physical/biological laws (more often used in biology and geophysics) (Lawson et al., 1995; Sibert et al., 1999, Vallino, 2000; Senina et al., 2008), to interpolate/extrapolate observations (optimal interpolation or krigging (in statistical terms) (Hansen and Poulain, 1996; Menemenlis et al., 1997).

In fact, assimilation problems can be dealt with in different ways: Bayesian methods, Optimal Interpolation, 3D- and 4D-variational methods, and Kalman filter. These different approaches differ mostly in their numerical cost, their optimality and their suitability for real-time data assimilation (Bouttier and Courtier, 1999).

The Bayesian approach however offers a unifying theoretical framework from which most existing data assimilation techniques can be derived (Wikle and Berliner, 2007; Jamet and Loisel, 2009).

6.2 Bayesian Approach

Lorenc (1986), Tarantola (1987) were among the first to write about Bayesian approach for DA (Wikle and Berliner, 2007). Indeed, *"Lorenc (1986) showed that one of the advantages of thinking about data assimilation from the Bayesian perspective is that it provides a common methodology that links many of the seemingly disparate approaches to the subject"* (Jamet and Loisel, 2009). Recent reviews on data assimilation explaining the different approaches and the linkages between them can be found in Wikle and Berliner (2007) and Jamet and Loisel (2009). A short summary is provided below.

In data assimilation, we are looking for the conditional probability of the model state (x) given the observations (y_{obs}) $P(x|y_{obs})$, called the "posterior distribution". $P(x|y_{obs})$ is a function of the product of a background probability density function (pdf) "$P(x)$", and an observation probability density function "$P(y_{obs}|x)$" (equation 6.1). If y_{obs} is not dependent of the state, we consider $P(y_{obs}) \approx 1$ (Bayes' theorem).

$$\boxed{P(x|y_{obs}) = \frac{P(y_{obs}|x) \times P(x)}{P(y_{obs})} \;\; \alpha \;\; P(y_{obs}|x) \times P(x) \;\; (Bayes' \; Theorem)} \tag{6.1}$$

where:

- y_{obs} is the vector of observations
- $P(x)$ is the background pdf, also called the **"a priori pdf" or prior distribution** of the model state before the observations are considered. This distribution quantifies our a priori understanding of the variables of interest.
- $P(y_{obs}) = \int (p(y_{obs}|x)p(x)dx)$ is the **probability density function of the observations** or "prior predictive probability" (or "marginal distribution").
- $P(y_{obs}|x)$ is the observation pdf, also called **the observation's likelihood, or data distribution,** and measures how likely the observations are described by the model state.
- $P(x|y_{obs})$, the posterior *"is the update of our prior knowledge about x as summarized in P(x) given the actual observations y_{obs}. In this sense, the Bayesian approach is inherently "scientific" in that it is analogous to the scientific method: one has prior belief (information), collects data, and then updates that belief given the new data (information)"* (Wikle and Berliner, 2007).

Let's assume that both model and data have gaussian distribution. Therefore, equation 6.1 can be simplified as described in equation 6.2 and 6.3.

$$P(x) \;\; \alpha \;\; \exp^{-\frac{1}{2}(x-x_b)^T B^{-1}(x-x_b)} \tag{6.2}$$

$$P(y_{obs}|x) \;\; \alpha \;\; \exp^{-\frac{1}{2}(y_{obs}-H(x))^T R^{-1}(y_{obs}-H(x))} \tag{6.3}$$

with:

- H the observation operator that makes the link between observations and predictions
- x^b the "background model state", the a priori information of the model state vector
- x^a the analysis model state. It represents the optimal solution for the observations given the "a priori information".
- R the covariance matrix of observations errors $(y - H(x))$. The observations errors contain errors in the observation process, errors in the design of the operator H and errors of discretization (Bouttier and Courtier, 1999)
- B the covariance matrix of model errors $(x_b - x)$. The background errors (or prior information) are the estimation errors of the background state, ie the difference between the background state vector and its true value.

From equation 6.2 and 6.3, taking the log of the "posterior probability" (log likelihood function):

$$L_a(x) = -ln(P(x|y_{obs})) \approx -ln(P(x) - ln(P(y_{obs}|x)) \tag{6.4}$$

$$\approx \underbrace{\frac{1}{2}(x - x_b)^T B^{-1}(x - x_b)}_{L_B(x)} + \underbrace{\frac{1}{2}(y_{obs} - H(x))^T R^{-1}(y_{obs} - H(x))}_{L_O(x)} \tag{6.5}$$

with L_B corresponding to the background penalty cost function or "a priori information" and L_O the observational cost function. In others words, the term L_B allows not to go too far away from the "a priori information", and the term L_O tries to get a solution closer to observations (see Fig. 6.1a). Eq. 6.5 is the usual formulation of the variational analysis (3D, 4D var, Bouttier and Courtier, 1999) which derives explicitly from the Bayesian principles.

Note: In biology, we note L_B and L_0 for the likelihood, whereas in physical oceanography and meteorology J_B and J_0 are used.

The best estimation of x (x_a) given the observation y_{obs} is *"the value that best reflects what a realization of x can really be in regard to y_{obs}"*. ***This definition is subjective, so several criteria can be proposed to define optimality"*** Cosme, 2010). Three optimal estimators are commonly used: the minimum variance estimation, the Maximum a Posteriori estimation (MAP) and the Maximum Likelihood Estimator (MLE) (Jamet and Loisel, 2009; Cosme, 2010). Before presenting in details the MLE estimator used in SEAPODYM, an example of different estimator is provided with the MAP.

In the **Maximum A Posteriori estimation (MAP)**, the estimate is defined as the most likely value of x given the observations y_{obs}, i.e, the analysis state x_a that maximize the posterior pdf: $x_a = argmax P(x|y_{obs})$ (see Fig. 6.1a). This is **illustrated with an application to track blue whales calls with ocean-bottom seismometers and an hydrophone array (section 6.3)**.

In the **Maximum Likelihood Estimator (MLE)**, the estimate is defined as the most likely value of y_{obs} given x, i.e., the analysis state x_a that maximize the observation pdf: $x_a = argmax P(y_{obs}|x))$ (see Fig. 6.1b). This is the **approach used in SEAPODYM**. Therefore in that case, we don't take into account any "a priori information" $p(x)$. Indeed, **the Maximum likelihood estimator (MLE) can be seen as the MAP estimator without any prior information.**

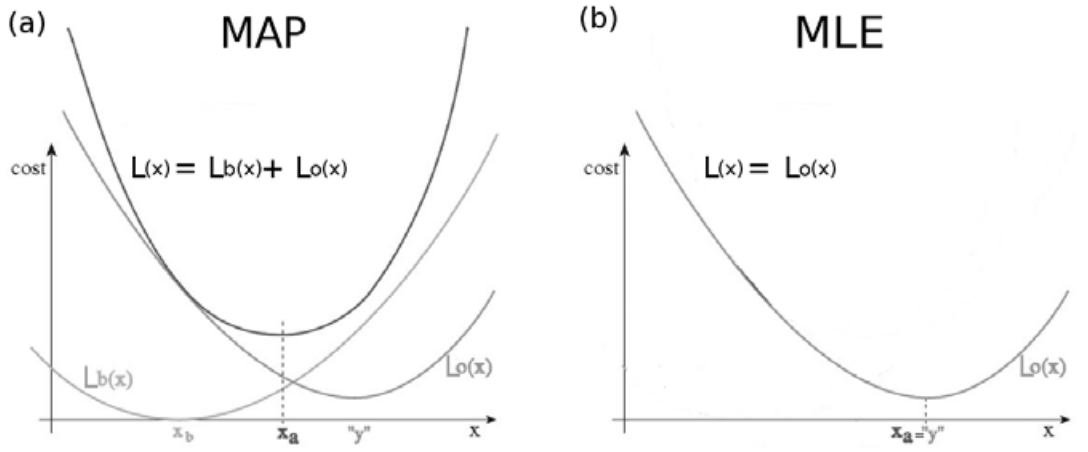

Figure 6.1: Schematic representation of a cost function depending on the optimality criteria chosen. A quadratic form (Gaussian case) has been chosen for the sake of simplicity. Adapted from Jamet and Loisel, 2009 and Bouttier and Courtier, 1999.

It should be noted that the definition of the best optimal criteria chosen is independent of the methods used to find the optimal solution x_a (solving Eq. 6.5).

Solutions of Eq. 6.5 can be roughly divided in two categories: the **variational approach** ("optimal control theory" in mathematics terms) and the statistical (or **sequential**, or "estimation theory" in mathematics terms) approach (Bertino et al., 2003; Jamet and Loisel, 2009).

In the following text, we will focus on the **variational methods** which find the analysis by minimizing the cost function L_a (or L_0, depending on the optimal criterion chosen) or maximizing the probability $P(x|y_{obs})$ (or $P(y_{obs}|x)$) using different techniques which will be discussed in the following section.

Variational techniques are typically employed to minimize the misfit between predictions and observations (a cost function) through adjustment of model inputs such as initial conditions, boundary conditions, and model parameters.

There are two fundamentally different methods to find the minimum of the cost function (Vallino, 2000; Tarantola, 2005), those based on a **local minimum**, i.e, on the local computation of the function to be optimized (gradient method), and those based on a **global minimum** that include: Grid search methods (systematic exploration of the model space), Monte-Carlo method (random exploration) (Tarantola, 2005), Genetic algorithms and Simulated annealing (Matear, 1995).

6.3 Global optimization approach: Grid search method

One example of a systematic exploration of the model space using a grid search method is given by Dunn and Hernandez (2009). In this work, we developed a grid search method (see paper in Appendix C) to represent the full probability of whales positions in a region of the ocean. In this example, we do not search for the minimum of the likelihood function $L(x|y_{obs})$ but for the maximum probability function $P(x|y_{obs})$.

Eight individual blue whales were recorded by different ocean-mounted hydrophones and seismometers during a 5-day period. They were tracked call-by-call via a Bayesian inversion procedure. In this study, our objective was to calculate the spatial coordinates of a whale (x,y) where x is the longitude and y is the latitude. Assuming that the whale is located at the surface, a grid over the model space m=(x,y), where we might find the whales, was defined. We calculated for each position a cost function P(m) (or the posterior probability density), which compared arrival time measurements of whales calls to predicted arrival times from a theoretical acoustic propagation model:

$$P(m) = P_B * P_0 = K * \rho(m) * \exp^{\Sigma_i^N \frac{(t_{obs}^i - t^i(m))^2}{\sigma_i^2}} \tag{6.6}$$

with t_{obs}^i the N (number of stations) observed arrival times of a single whales call minus a weighted average of all such times, $t^i(m)$ the theoretical travel times from a grid location in m to each station for which exists a value t_{obs}^i minus a weighted average of these times, σ_i the combination of observational and theoretical uncertainties, K a constant such that the probability of the whale being somewhere in the grid is 100% and $\rho(m)$ the priori information (P_B, the background probability) that may exist on the whale position before we calculate an estimate of the position (given the speed of the whale, the location at time i+1 cannot be further than the maximal distance travel).

The synthetic arrival times are calculated for all grid points allowing to obtain the probability density function of whale's location as illustrated on Fig. 6.2. On the left,

Gaussian distribution appears as a series of constant color contour ellipse. On the right, the probability density function shows two possibilities for the location of the whale. The location with the largest probability is chosen, but it is obvious that another location is almost as probable as the chosen location. A linearized inverse problem method would only find one or the other of these two possible locations, while our more general algorithm as the grid search shows them both. The main advantage of this method is that it provides a full representation of the probability density function. For more details, see Dunn and Hernandez, 2009 in Appendix C. In the paper, we used the L1 norm to quantify the misfit, because the solution is thereby less biased by outliers in the data (Tarantola, 2005).

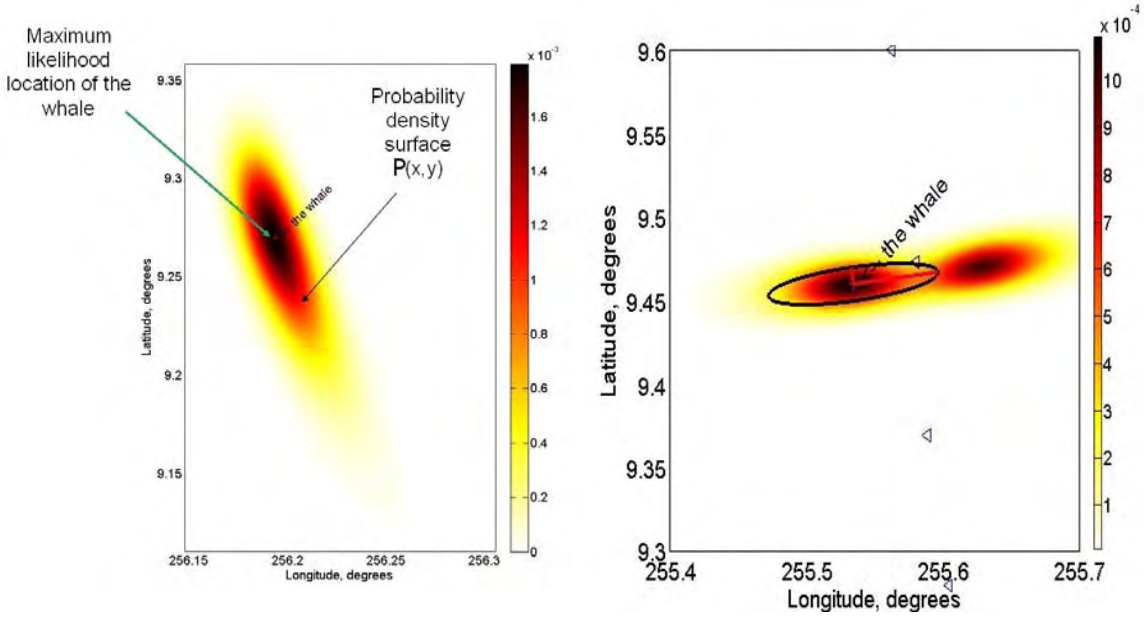

Figure 6.2: Left panel: Maximum Likelihood of whales position at time t. Location of the whale is given as the position where the cost function or posteriori probability is a maximum. Here, ρ_m has been considered as a constant, i.e, no priori information of the whales position has been used, and K was set to 1. Right panel: error ellipse for the probability density function has been plotted (1 σ = 68% of confidence level)

6.4 Local optimization approach: SEAPODYM framework

In the case of SEAPODYM, because it is too costly to explore the n-dimensional parameter space, with the number of parameters n higher than 10, the systematic exploration of the model space is not feasible. Therefore, a method based on the local computation of the cost function is used: the gradient-based method. This method is very efficient, but has the disadvantage of not automatically giving the global minimum. Indeed, for the analysis of nonlinear problems, the functions may have various different local minimum and the main issue is to find the minimum log-likelihood value ("real" or "global" minimum). But there is no simple way to evaluate the uniqueness of the

estimated parameters in non-linear problems. We will use the common approach of perturbing the model parameters and restarting the experiment but we will not be able to determine for sure if the solution found is a global minimum. We could however verify if the estimated parameters were well determined at the minimum during the minimization, via the sign of the determinant of the Hessian matrix (see section 6.4.3).

In SEAPODYM, we want to adjust model parameters (θ). We have between 10 and 20 parameters to estimate.

For simplicity in the presentation of the approach, we will focus on the Gaussian Case.

The chosen local optimization approach is the **"descent method"** that iteratively determines directions locally "descending" along the cost function surface as observed in Figure 6.3a. This Figure represents the theoretical case where the model and the observation operator H are linear, which implies that the cost function is quadratic and his minimum is unique. For non linear problems (which is is the usual case in a biology problem), the cost function L is a nonlinear function, then the hyper-surface is not necessarily quadratic and it is possible to have more than one minimum (Fig. 6.3b).

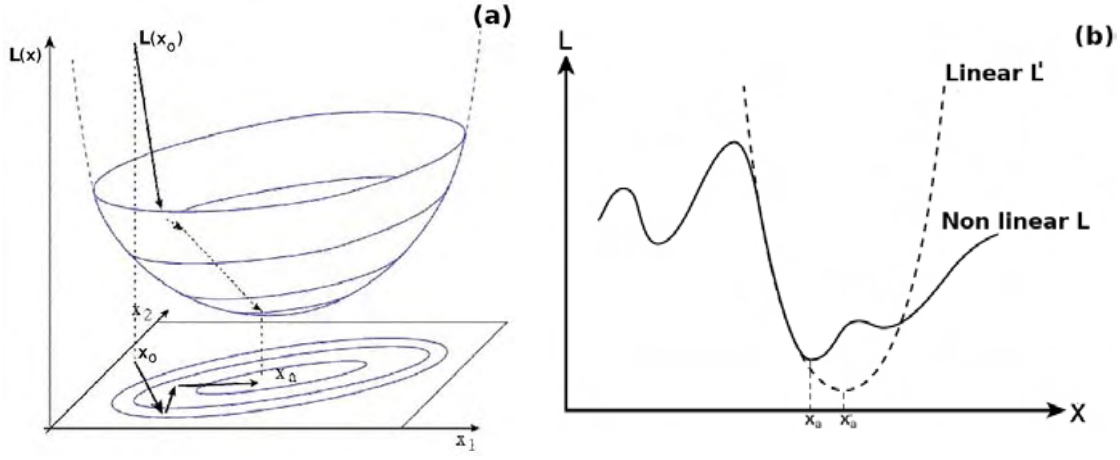

Figure 6.3: (a) Schematic representation of the cost-function minimization in 2D model space for a quadratic cost function. The quadratic cost-function has the shape of a paraboloid, with the minimum at the optimal analysis. The minimization works by performing several line-searches to move the control variable x to areas where the cost-function is smaller, usually by looking at the local slope (the gradient) of the cost-function (Bouttier and Courtier, 1999). (b) Schematic representation of the cost-function minimization in 1D model space for a linear and non linear cost function.

The objective of the optimization is to find the minimizer θ^*: $L(\theta^*) = min(L(\theta))$. Descent algorithms correspond to all algorithm where $\theta^{i+1} = \theta^i + \alpha^i * d^i$ such that $L(\theta^{i+1}) < L(\theta^i)$. α^i is the descent step at iteration i and d^i the descent direction at the iteration i.

Several methods correspond to these descent approach algorithms depending on the formulations for α^i and d^i: Optimal and fixed step methods, relaxation methods, gradient descent methods, newton methods (Tarantola, 2005; Blayo et al., 2011).

Here, we focus on the Newton method and more precisely on the Quasi-Newton method which is used in SEAPODYM. In this case, we look for the set of k parameters θ^* of the scalar function for which the gradient of L becomes zero: $\nabla L(\theta)|_{\theta=\theta^*} = 0$ (Blayo et al., 2011).

Newton's method consists of iteratively approaching the minimum of L using Eq. 6.7:

$$\theta_{i+1} = \theta_i - [\nabla^2 L(\theta_i)]^{-1} * \nabla L(\theta_i) \tag{6.7}$$

where $\nabla^2 L(\theta_i)$ is the Hessian of L at iteration i.

The descent step direction is given by: $d_i = [\nabla^2 L(\theta_i)]^{-1} * \nabla L(\theta_i)$. For each iteration i, we need to calculate the Hessian and solve a linear problem of matrix size $\nabla^2 L(\theta_i)$.

For big computational problems (here 10-20 parameters), this is too costly and thus not realizable. In order to calculate the Hessian, we used a LBFGS Quasi-Newton algorithmn (Nocedal, 1980), that gives an approximation of this matrix. In SEAPODYM, this algorithm is implemented using utilities of automatic code differentiation library (AUTODIFF - Otter Research LTd., 1994). Note also that in SEAPODYM, the likelihood function L is not normalized.

Newton methods requires an evaluation of the derivatives of the cost function (L) with respect to each control parameters. We will see now how to calculate the gradient of the cost function L.

6.4.1 Adjoint methodology

The Quasi-Newton method uses at each iteration the gradient of the cost function relative to each control parameter. We need to have an efficient and fast method to calculate this gradient. One method is to approximate the gradient by finite difference. However, if the number of parameters is high, as in our case, the adjoint method is the most efficient (Lions, 1971). Indeed, finite difference does not give an exact calculation of the gradient whereas adjoint does and the computational cost of using finite difference will *"increase linearly with the number of design variables"* (Giles et al., 2005) whereas *"the computational cost of using the adjoint approach is independent of the number of the design variables"* (Giles et al., 2005).

Now that we have seen the motivations of using the adjoint method, let's briefly explain the principle. For more details, a review of this technique is given in Bouttier and Courtier (1999).

Beginning from eq. 6.5, we note M a model such as, $\forall$ time i, $\theta_i = M_{0\to i}(\theta) = M_n M_{n-1} ... M_i ... M_1$. $M_{0\to i}(\theta)$ is a predefined model forecast operator from the initial time 0 to n. Therefore,

$$L(\theta) = L_0(\theta) = \sum_{i=1}^{n} \frac{1}{2}(y_i - H_i M_i M_{i-1} .. M_1 \theta)^T * R^{-1} * (y_i - H_i M_i M_{i-1} ... M_1 \theta) \quad (6.8)$$

Taking the gradient of L:

$$\nabla L(\theta) = -2 \sum_{i=1}^{n} M_1'^T ... M_i'^T ... M_n'^T H_i^T R^{-1}(y_i - H_i M_i M_{i-1} ... M1k) \quad (6.9)$$

with M_i' the derivative of the model M called the tangent linear model and $M_i'^T$ the transpose of M_i' called the adjoint model of M.

Using d_i, the innovation vector corresponding to the misfit to the observations, such as $d_i = y_i - H_i M_i M_{i-1} ... M_1 k$:

$$\nabla L(\theta) = -2 \sum_{i=0}^{i=n} (M_1'^T ... M_{i-1}'^T M_i'^T H_i^T R^{-1} d_i) \quad (6.10)$$

Therefore, after some factorizations:

$$-\nabla \frac{1}{2} L(\theta) = H_0^T R_0^{-1} d_0 + M_1'^T [H_1^T R_1^{-1} d_1 + M_2'^T [H_2^T R_2^{-1} d_2 + ... + M_n'^T H_n^T R_n^{-1} d_n]] \quad (6.11)$$

The adjoint model is forced by the innovations. This factorization gives us an easy way to calculate the gradient, which is therefore obtained by a backward integration of the adjoint model. Indeed, Eq. 6.11 is easily evaluated in a reverse mode, i.e, from right to left. Such factorization allows to compute L and then ∇L thanks to one direct model integration and one adjoint model. More details about practical coding of the adjoint is described in Annexe B.

Technically, the algorithm of data assimilation in SEAPODYM consist in the following steps. We initialize the parameters $\theta = \theta_0$, and the numbers of iterations n to 0. As long as the gradient of L in respect to each parameter is higher than a convergence criteria (ϵ) or before reaching a maximum number of iterations (n_{max}), i.e, as long as $\forall k$, $\frac{\partial L}{\partial \theta_k} > \epsilon$ or $n <= n_{max}$, we do the following steps:

1. Calculate L using the direct model M and the observational operator H.

2. Calculate the gradient of L using the adjoint model M'^T and the adjoint of the observational operator H^T (in reverse mode).

3. Use the gradient of L in a optimization procedure to compute the search direction towards the minimum via a Quasi-Newton method. We update θ parameters

4. Increment the number of iterations: n= n+1

To summarize, the implementation of data assimilation model in SEAPODYM can be divided in three components: the forward ecosystem model with a cost function, the backward model or adjoint model, and an optimization procedure (Fig. 6.4).

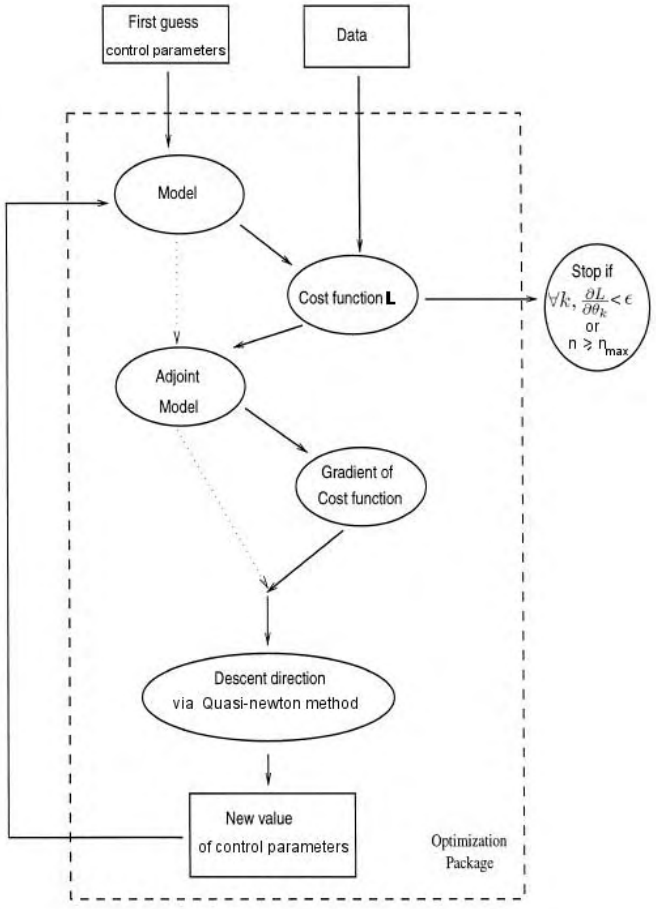

Figure 6.4: Schematic view of the steps involved in a data assimilation procedure using variation adjoint method. The solid lines indicate the main path taken during the procedure. The direct model is run with an initial guess on the set of parameters. The model output and data are then used to compute the value of the cost function. Then, the adjoint of the model runs backward in time to give the gradient of the cost function with respect to each parameter. This gradient is used in the optimization procedure to compute the search direction towards the minimum via a Quasi-Newton method. New values for the parameters are then estimated, and we re-run the model. The optimization procedure is applied until the convergence criterion has been satisfied. Adapted from: Spitz et al. (1998)

6.4.2 Parameter scaling and bounds

In order to make an efficient optimization process, it is important to limit the minimization to a range of plausible values. The Quasi-Newton algorithm itself does not deal with boundaries in the parameters. We should make sure that parameter does not go out of its limits. The instruction if ($\theta_k > \overline{\theta_k}$) then $\theta_k = \overline{\theta_k}$ is not differentiable.

We have two possibilities to deal with this problem. One method is to constrain minimization setting a penalty to the boundaries adding a **penalty term** ($L_P(\theta)$) into the cost function L when approaching boundaries ($L(\theta) = L_0(\theta) + L_P(\theta)$ (e.g., Sibert et al.,

1999; Faugeras and Maury, 2005). The second option is to constrain the minimization through **parameter scaling** (see also e.g. Vallino, 2000). This is the approach used in SEAPODYM. We use the second method, which maps the constrained parameter space to unconstrained and at the same time allows the parameter to vary within its bounds. The only problem is that the parameter can be stuck to its boundary, but the cost function will be always differentiable. Let's explain in more details the second approach.

Parameters (θ_k) are known to be on a certain range of plausible values, in a bounded space K ($\theta_k \ \epsilon$ K). Therefore in the optimization process, we will explore only the plausible parameter space. Consequently, each parameter has an upper and lower range $\underline{\theta_k} \leq \theta_k \leq \overline{\theta_k}$.

Since parameters have different units and magnitude order, they need to be rescaled within -1 and 1. So, parameters values are scaled with an arcsin function by Eq. 6.12 (see Fig. 6.5 a).

$$\theta_k^{'} = \frac{2}{\pi} * arcsin(2 * \frac{\theta_k - \underline{\theta_k}}{\overline{\theta_k} - \underline{\theta_k}} - 1) \tag{6.12}$$

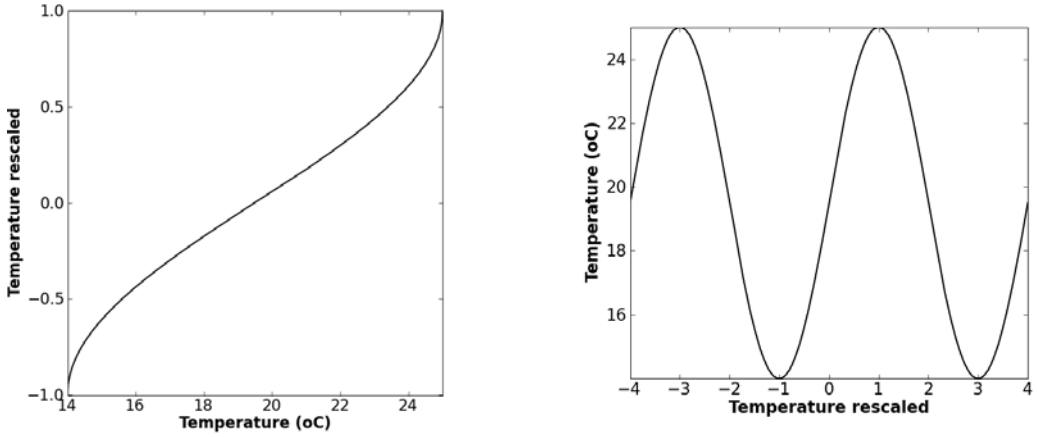

Figure 6.5: Functions from AUTODIFF library to scale parameters initially from -1 to 1 (a; Eq. 6.12), and to rescale parameters from unbounded space to bounded one (b; Eq. 6.13). These plots are made with the actual functions of AUTODIFF library ("boundpin" and "boundp" respectively). The example is made for optimal temperature parameter T, which varies from 14 to 25°C.

During the optimization process, parameters can vary in this unbounded space ($\theta_k^{'} \ \epsilon \ K'$) from $-\infty$ to $+\infty$. Minimization operates in the unbounded parametric space $\theta_k^{'}$, that is mapped to the bounded θ_k via the transformation of Eq. 6.13 (Fig. 6.5 b). Therefore, the result stays always within the bounds.

$$\theta_k = \underline{\theta_k} + (\overline{\theta_k} - \underline{\theta_k}) * (0.5 + 0.5 * sin(pi * \theta_k^{'}/2)) \tag{6.13}$$

6.4.3 Analysis of the quality of the estimate: Hessian Information

The Hessian (H) is defined as the matrix of second partial derivatives of the likelihood function in respect to each parameter ($H(\theta) = \frac{\partial^2 L(\theta)}{\partial\theta_i \partial\theta_j}$) (i,j = 1,...,k). It provides the error covariance for each estimated parameters (Bouttier and Courtier, 1999) and thus, an estimate of the quality of the outputs.

At the minimum of the negative log likelihood function, where the quadratic approximation is made (hypothesis made using the Quasi-Newton method), the Hessian of the likelihood function is equal to twice the inverse of the error covariance matrix (A):

$$A = (\frac{1}{2}L'')^{-1} \tag{6.14}$$

From this definition, the Hessian is the measure of the steepness of the likelihood surface in the region of the optimum.

This equation allows us to understand that the quality of the estimate is proportional to the convexity of the cost function. The stronger the convexity of the cost function is, the more reliable the estimate will be. In one dimensional problem, Figure 6.6 illustrates the problem. It shows the importance to define well the problem in order to obtain a very convex function.

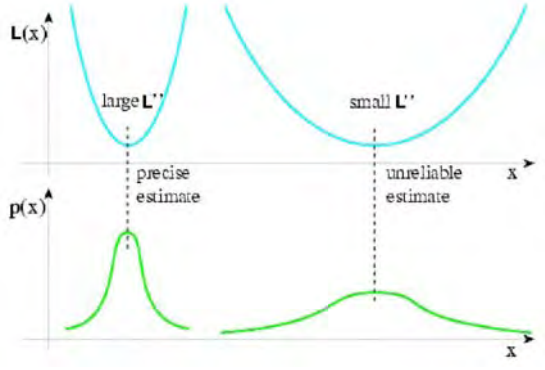

Figure 6.6: Illustration in a 1D problem of the relationship between the Hessian and quality of the analysis. If the convexity is strong (left), the probability density function (pdf) is sharper and the analysis is more reliable. If the convexity is weaker (right), the probability density function is smooth and we cannot obtain a precise estimate of the analysis (Adapted from Bouttier and Courtier, 1999).

In practice in SEAPODYM, the Hessian is approximated by central finite difference using the exact calculation of the gradient of L using the adjoint. This calculation is feasible since we have only 10-20 parameters to estimate. Knowing the matrix (H) we can compute its determinant, its eigenvalues, the matrix of variance and covariance parameters, the correlation matrix and the standard errors for each parameter.

If the determinant is positive and the eigenvalues are positives, the stationary point is a minimum; if the determinant is negative and all the eigenvalues are negatives, the stationary point is a maximum; and if the eigenvalues of H are positive and negative the stationary point is a saddle point.

The square roots of the diagonals of the inverted Hessian are the standard errors of the estimated parameters. We obtain the covariance and variance matrix by inversing the hessian matrix (equation 6.14). The correlation matrix is readily obtained by dividing the covariance of the two variables by the product of their standard deviations.

6.5 Implementation of a new likelihood function on SEAPODYM

The optimization method in SEAPODYM has been developed by Senina et al. (2008). The authors use the maximum likelihood method with fishing data to estimate the set of parameters that minimizes the differences between predictions and observations. Here we adapt the method with a data assimilation procedure using eggs and larvae data.

6.5.1 Observations and predictions

Larvae and eggs data are aggregated at the 5-day time step resolution of the model. For a given sample collected with a type of net n in a cell i, j, during the time step t, the density of eggs or larvae, $d_{a,n,t,i,j}$ is provided by IMARPE in number per square meter, after correction for the surface area of the net (see Chapter 3). In the general case, our optimization approach relies on the comparison of total abundance of eggs or larvae collected by cell and time step (S^{obs}) with the corresponding predicted values (S^{pred}) using the same observed sampling effort.

Thus, to compute the total observed abundance (in number of eggs or larvae), the density ($ind.m^{-2}$) is converted to original abundance value knowing the sampling effort e which is simply determined by the surface area of the net (m^2):

$$S^{obs}_{n,t,i,j} = \sum_a d_{a,n,t,i,j} \bullet e_n \tag{6.15}$$

The total associated sampling effort $E_{n,t,i,j}$ observed for the cell i,j and the net n is equivalent to the total surface sampled (m^2):

$$E_{n,t,i,j} = \sum_a e_n \tag{6.16}$$

The predicted abundance is the product of the observed sampling effort with the predicted density (N) in the cell and a coefficient of catchability q that characterizes the sampling efficiency of the net that should be estimated:

$$S^{pred}_{n,t,i,j} = q_n \bullet E_{n,t,i,j} \bullet N_{t,i,j} \tag{6.17}$$

In the special case of monthly climatological series of eggs and larvae (cf. Chapter 3), since the climatology provides one single average density per cell and month, the

sampling effort is constant and set to 1. The predicted abundance is the monthly average computed from the number of time steps per month (6 for a 5-day time step):

$$S^{pred}_{m,i,j} = \sum_{t=1}^{6} (q \bullet \frac{N_{t,i,j}}{6}) \tag{6.18}$$

6.5.2 Cost function

For reasons discussed in the section 6.6), the cost function is written following a Negative Binomial probability law (NB2 form):

$$\boxed{P(S^{obs}_{ntij}|S^{pred}_{ntij}, \boldsymbol{\beta}_f) = \frac{\Gamma(\beta+S^{obs}_{ntij})}{\Gamma(\beta)\Gamma(S^{obs}_{ntij}+1)} * \left(\frac{\beta}{\beta+S^{pred}_{ntij}}\right)^{\beta} * \left(\frac{S^{pred}_{ntij}}{\beta+S^{pred}_{ntij}}\right)^{S^{obs}_{ntij}}} \tag{6.19}$$

with β parameter representing the amount of overdispersion. It is the negative binomial parameter which will be estimated during the optimization process.

When using both eggs and larvae data in the optimization, the negative log-likelihood function to be minimized ($L^- = -ln(P)$) is the sum of the negative log-likelihood function for eggs and for larvae:

$$L^- = L^-_{eggs} + L^-_{larvae} \tag{6.20}$$

The adjoint method has been developed in SEAPODYM (Senina et al., 2008), but the new likelihood function required several modifications. To verify that the changes were correctly implemented, we checked that the value of the gradient $L(x)$ calculated with finite differences (using utilities of automatic code differentiation library AUTODIFF - Otter Research LTd., 1994) was identical to the gradient calculated by the adjoint code. Then we verify that equation (6.21) is correct, i.e., that the discrepancy between each gradient component obtained by analytic differentiation (adjoint code) and its finite difference approximation changes parabolically with step h (Senina et al., 2008).

$$\frac{L^-(\theta_k + h) - L^-(\theta_k - h)}{2h} - \nabla_k L^- = O(h^2) \tag{6.21}$$

An illustration (Fig. 6.7) is provided for two parameters of the spawning habitat optimal temperature (Temperature (T_0), standard deviation (σ_0)) and two catchability parameters ($q1$, $q2$) characterizing the net for eggs and larvae.

Finally, the approach was validated with a twin experiment. Starting from larvae and eggs pseudo-observations simulated with a fixed set of parameter values, we verify that after changing the parameter values, the model can converge and find the exact original values of the parameters. Six twin experiments with randomly created sets of perturbed

parameters were conducted and all successfully recovered the original values with a relative error below <1%. Furthermore, we verified that the hessian was definite positive which validate the local minimum found.

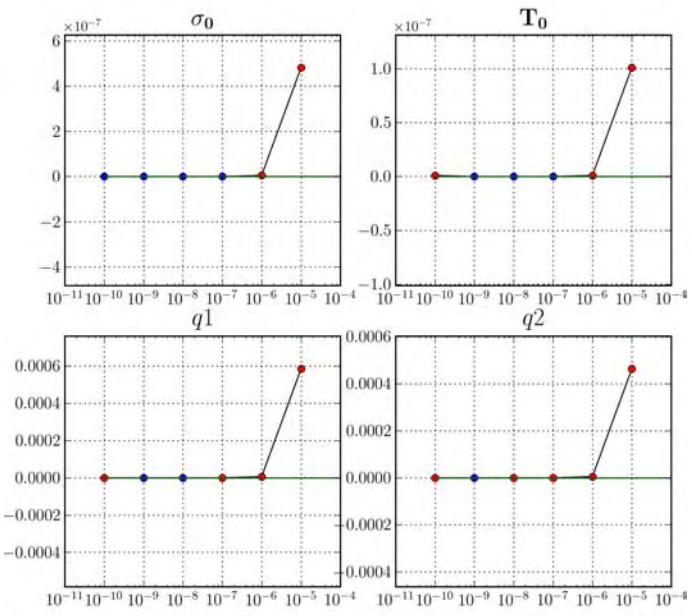

Figure 6.7: Relative error in finite difference sensitivity as a function of the step size h from 10^{-10} to 10^{-5}. Blue points correspond to values where relative error is 0, and red points were relative error is positive. For larger values of step size the dominant error is due to the second order accuracy of central finite differences (Giles et al., 2005). For smaller values, the dominant error is due to machine accuracy.

Note: a similar approach can be used for acoustical data.

6.5.3 Initial conditions

Initial distribution of the population can play an important role in parameter estimation process. In the SEAPODYM application to the entire population dynamics, it is necessary to generate a first guess of the distribution of fish population from larvae to adults. These initial conditions can be generated from a "spin-up" process starting from a uniform zero spatial distribution, from a priori spawning habitat based on first guess of parameter values, or from any known spatial pattern if there are sufficient observations to create a distribution map. In any case, it is necessary to reduce the impact of these initial conditions (IC) by excluding the longest possible initial time period of prediction from the optimization. This period will depend on the availability of the environmental forcing, the period of study, the data available and the lifespan of the species. If possible it is recommended to exclude a period of at least one lifespan. Another approach to reduce the impact of IC is to run optimization experiments several times after regenerating each time the initial conditions using the results of the last experiment.

In the particular case of this study that focus on the optimization of the early life stage's parameters, IC have very limited impact since only the first 5-day time steps of anchovy and sardine are concerned. Excluding the two first time steps from the initial conditions is enough to eliminate the impact of the initial conditions.

6.5.4 Sensitivity analysis

Sensitivity analysis is performed to reveal which parameters can be estimated from available data and which cannot. Two types of sensitivity metrics can be performed: one examines whether the predictions of the model are sensitive to its parameters, the second checks whether the objective function (L) is sensitive to model parameters.

It is recommended to run multiple sensitivity tests from experiments with different initial conditions, time periods of simulation, and estimated values of parameters. If model predictions remain insensitive to one parameter even after using all available data, then this parameter should be fixed to its best guess-estimate and removed from the optimization process.

6.5.4.1 Sensitivity of model to parameters

The first type of analysis examines how the predictions of the model are sensitive to its parameters. For this purpose we simply need to construct a function of the model solution, which represents the model predictions (see, e.g., Worley, 1991). Then, the measures of sensitivity can be computed using precise gradients obtained from adjoint calculations.

The sensitivity of model to parameters is given by Eq. 6.22.

$$R_1 = \sum_{ntij} \left(S^{pred}_{ntij} \right)^2, \tag{6.22}$$

Then we define the measure of relative sensitivity ($\xi_1(^0_k)$) for corresponding model predictions and each initial guess parameter (θ^0_k) as follows:

$$\xi_1(\theta^0_k) = \frac{1}{R_1} \frac{\partial R_1}{\partial \theta^0_i}, \tag{6.23}$$

6.5.4.2 Sensitivity of objective function to parameters

In this case, we compare values of likelihood at some found minimum $\boldsymbol{\theta}^\dagger$ to those evaluated at boundaries of parameter space (Vallino, 2000). The sensitivity of objective function to parameters is given by Eq. 6.24.

$$\xi_2(\theta^\dagger_k) = \frac{L^-(\boldsymbol{\theta}^\dagger + \delta\bar{\theta}_k \cdot \mathbf{e}_k) - L^-(\boldsymbol{\theta}^\dagger)}{L^-(\boldsymbol{\theta}^\dagger)}, \tag{6.24}$$

where $\delta\bar{\theta}_k = \bar{\theta}_k - \theta^\dagger_k$ and $\mathbf{e}_k$ is a standard basis vector with 1 in the k-th element and 0 elsewhere.

6.6 Selecting the distribution of the likelihood function

For simplicity, optimization theory was introduced in the first section of this chapter with the Gaussian case. But since our observed data are not Gaussian, different forms of the cost function and the subsequent gradient can be used. The Negative Binomial form has been chosen for the application to eggs and larvae data (cf. above). The reasons of this choice are explained in this section.

Theoretically, if data and model were perfect, the type of likelihood chosen would not matter since the model would always be able to find the parameters for which observations equal predictions. However, this perfect model does not exist and errors in data (measurement errors, methodological uncertainties, poor data coverage) and model need to be questioned when choosing the cost function.

It is important to understand which range of data is fitted best when choosing one or another likelihood function. In biology, the choice is often based on the distribution of observations. Distributions of eggs and larvae data have been shown close to a Negative Binomial distribution (Chapter 3). This means that data are highly overdispersed, and that predictions probably cannot fit the high values of biomass.

Then, the behavior of the different cost functions as a function of the observed values needs to be analyzed. The most common cost functions, ie the Normal, Log-normal, Poisson, Negative Binomial and Zero Inflated Negative Binomial distribution, are presented below in Eq. 6.25 to 6.31. The choice of additional functions is possible and can easily be implemented in the model.

We will denote S^{obs} the observations of eggs or larvae, S^{pred} the corresponding predictions and σ^2 the variance of the dataset which will be fixed to one for this example.

1 - The Normal distribution (Gaussian case):

$$P(S^{obs}_{ntij}|S^{pred}_{ntij}) = \frac{1}{\sqrt{(2\pi)}} \exp \frac{(S^{pred}_{ntij} - S^{obs}_{ntij})^2}{\sigma^2} \tag{6.25}$$

2 - The Log-Normal distribution:

$$P(S^{obs}_{ntij}|S^{pred}_{ntij}) = \frac{1}{\sqrt{(2\pi)}} \exp \frac{(ln(S^{pred}_{ntij} + 1) - ln(S^{obs}_{ntij} + 1))^2}{\sigma^2} \tag{6.26}$$

3 - The Poisson distribution:

$$P(S_{tnj}^{obs}|S_{ntij}^{pred}) = \prod_{ntij}\left(\frac{(S_{ntij}^{pred}S_{ntij}^{obs}e^{-S_{ntij}^{pred}}}{S_{ntij}^{obs}!}\right) \tag{6.27}$$

4 - The Negative Binomial distribution:
The most general way to express the Negative Binomial distribution is given by equation 6.28.

$$P(y_i|\lambda_i,\boldsymbol{\beta}) = \frac{\Gamma(\beta\lambda_i^{2-PP}+y_i)}{\Gamma(\beta\lambda_i^{2-PP})\Gamma(y_i+1)} * (\frac{\beta\lambda_i^{2-PP}}{\beta\lambda_i^{2-PP}+\lambda_i})^{\beta\lambda_i^{2-PP}} * (\frac{\lambda_i}{\beta\lambda_i^{2-PP}+\lambda_i})^{y_i} \tag{6.28}$$

where y_i is our observation, λ_i our prediction (theoretically defined as the mean of the distribution), PP a parameter describing the type of Negative Binomial distribution, and β parameter represents the amount of overdispersion. Eq. 6.28 can therefore be written in the following way:

$$P(S_{ntij}^{obs}|S_{ntij}^{pred},\boldsymbol{\beta}_f) = \frac{\Gamma(\beta(S_{ntij}^{pred})^{2-PP}+S_{ntij}^{obs})}{\Gamma(\beta(S_{ntij}^{pred})^{2-PP})\Gamma(S_{ntij}^{obs}+1)} * \left(\frac{\beta(S_{ntij}^{pred})^{2-PP}}{\beta(S_{ntij}^{pred})^{2-PP}+S_{ntij}^{pred}}\right)^{\beta(S_{ntij}^{pred})^{2-PP}} * \left(\frac{S_{ntij}^{pred}}{\beta(S_{ntij}^{pred})^{2-PP}+S_{ntij}^{pred}}\right)^{S_{ntij}^{obs}} \tag{6.29}$$

The value of the variance for this distribution is given by Eq. 6.30.

$$Var = \lambda_i(1+\frac{1}{\beta}\lambda_i^{PP-1}) \tag{6.30}$$

Two different forms of the negative binomial distribution are proposed in the literature (Hilbe, 2011). The most common is the NB2 form, with PP = 2 (Mangel and Smith, 1990; Minami et al., 2007; Vaudor et al., 2011). The other form is the NB1 function with PP=1 (Senina et al., 2008). This latter function is therefore described as a pure form of overdispersion (Hilbe, 2011). The NB2 function is also called a quadratic form since variance is a function of λ_i^2, and the NB1 is called a linear form (variance function of λ_i).

5 - The Zero Inflated Negative Binomial distribution can be described as a mixture of two distributions, the delta distribution on zero (the distribution that takes only the value zero; "perfect state") and a distribution on the non-negative integers (i.e., including the value zero; "imperfect state") (Minami et al., 2007). The definition of the ZINB function is given by Eq. 6.28 where a sample is in the perfect state with probability p and in the imperfect state with probability (1-p).

$$\begin{aligned} P(S_{ntij}^{obs}|S_{ntij}^{pred},\boldsymbol{\beta}_f) &= p+(1-p)q(0|S_{ntij}^{pred},\theta), \text{ if } S_{ntij}^{obs}=0 \\ &= (1-p)q(S_{ntij}^{obs}|S_{ntij}^{pred},\theta), \text{ if } S_{ntij}^{obs}>0 \end{aligned} \tag{6.31}$$

where q corresponds to the Negative Binomial function (NB1 or NB2) and parameter p is the probability of getting a null observation.

The optimization approach will aim at minimizing the difference between observation and predictions. According to the selected likelihood function, the approach to estimate values in a range of several orders will show different gradients as illustrated on Fig. 6.8, where the negative log likelihood function was computed for different values using each likelihood function described above.
With the Gaussian distribution (Fig. 6.8a), whatever the value of the observation the shape of the likelihood remains identical, meaning that all observations have the same weight in the minimization of the likelihood function. In the case of the Log-normal likelihood (Fig. 6.8b), the gradient is strong for estimating all low observed values but as soon as observed values increase, the gradient becomes dissymetric with steep slope when predicted values are lower than observations and weak or even flat slope for estimated values higher than observed ones. Using this function, the result should fit more correctly the low observed values since they have a strong contribution to the likelihood. For the Poisson distribution (Fig. 6.8c), the gradient is also dissymetric but the slope remains relatively steep on the left side (where predicted values are higher that observed values) which should assure a relatively good convergence toward the exact values. If the model cannot fit the exact values, then the solution of the optimization should tend to predict higher values than observed ones.

With the Negative Binomial (NB) or the Zero Inflated Negative Binomial (ZINB), Figure 6.8 shows that the minimum value of the likelihood never reach zero for non null observation; this discrepancy increasing with the value of the observation. This tendency is due to a representation by this group of functions of the overdispersion in the data, characterized by the parameter β (Eq. 6.19). In addition, the minima of the first form of these functions (NB1 or ZINB1) do not coincide with the exact observed values, but to a higher value. This is not the case for the second form (NB2 or ZINB2). The steepness of the gradients of these functions is linked to the parameter β, but the estimates for this parameter are not comparable between the two forms. Indeed, NB1 function is equal to NB2 function by replacing β with $\beta * \lambda_i$.

Small values of β correspond to high overdispersion of data. In that case, NB functions have much stronger gradient for predictions that are lower than observed values and weak or flat slope for larger predictions (Fig. 6.9). In consequence, an optimization based on these functions and parameterization of β will do not really discriminate between values ranging between several orders of magnitude. With very small values of β, the optimization is finally equivalent to fitting simply absence (zero values) and presence (positive values) since all non zero positive values tend to have the same weight.

The results are similar for the ZINB excepted that when observed values are zero the gradient and the likelihood approach zero, meaning that the optimization does not take into account the zeros from the observation.

In summary, as stated by Greene (2008), there is no simple parametric test that can simply define the best choice between NB1 and NB2 forms. This choice however can be critical in the optimization approach.

In practice, it may be important to consider the meaningful of various parts of the dataset, especially zeros and very high values. To give more weight to the small observed values, the log normal distribution can be used. Conversely, to suppress the effect of zero values, the Zero Inflated Negative Binomial (ZINB) distribution should be selected. If there is high overdispersion of high values, the Negative Binomial distribution (NB) seems more appropriate. It has the advantage of being more general than Poisson distribution.

In the case of our study based on eggs and larvae data, the NB2 distribution will be used since both zero values (absence of eggs or larvae, i.e., very unfavourable conditions) and very high values (very favourable conditions) bring key information. This distribution will also take into account the obvious overdispersion of the data and will help to search for the exact values.

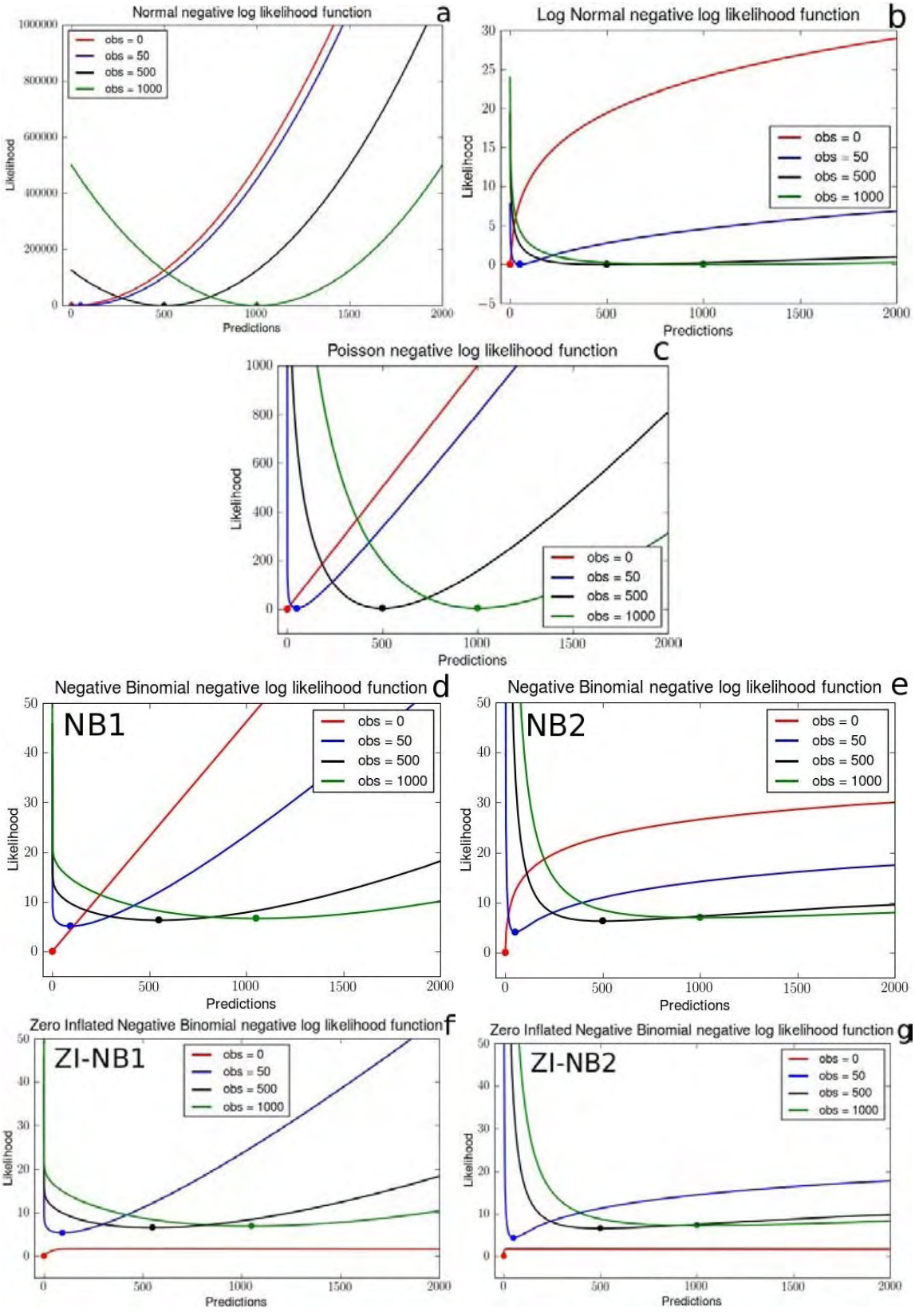

Figure 6.8: Shape of different likelihood function versus predictions in 1D for four different observed values: 0 (red), 50 (blue), 500 (black) and 1000 (green). For the NB1 and ZINB1 functions, a value of $\beta = 0.01$ has been chosen whereas for NB2 and ZINB2 functions a value of $\beta = 5$ was used. For the ZINB functions a value of $p = 0.2$ has been taken.

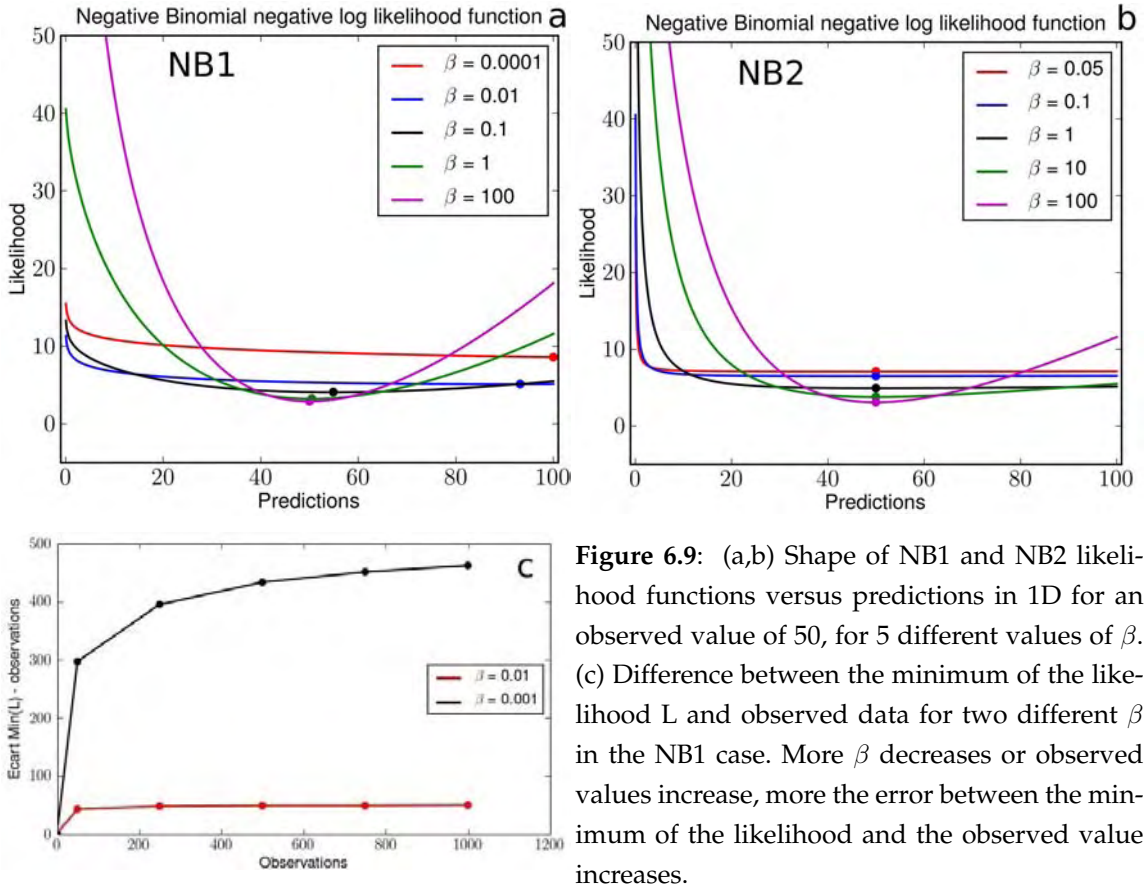

Figure 6.9: (a,b) Shape of NB1 and NB2 likelihood functions versus predictions in 1D for an observed value of 50, for 5 different values of β. (c) Difference between the minimum of the likelihood L and observed data for two different β in the NB1 case. More β decreases or observed values increase, more the error between the minimum of the likelihood and the observed value increases.

Bibliography

Bertino, L., G. Evensen, and H. Wackernagel, 2003: Sequential data assimilation techniques in oceanography. *International Statistical Review*, **71**(2), 223–241.

Blayo, E., E. Cosme, M. Nodet, and A. Vidard, 2011: Introduction to Data Assimilation. *http://ljk.imag.fr/membres/Maelle.Nodet/documents/MN_DA.pdf*.

Bouttier, F. and P. Courtier, 1999: Meteorological Training Course Lecture Series. *ECMWF*.

Brasseur, P., P. Bahurel, L. Bertino, F. Birol, J. Brankart, N. Ferry, S. Losa, E. Remy, J. Schroter, S. Skachko, et al., 2005: Data assimilation for marine monitoring and prediction: The MERCATOR operational assimilation systems and the MERSEA developments. *Quarterly Journal of the Royal Meteorological Society*, **131**(613), 3561–3582.

Cosme, E., 2010: Introduction to Data Assimilation. *http://www-meom.hmg.inpg.fr/Web/pages-perso/Cosme/doc_cours/ASSIM/coursEC.pdf*.

Dunn, R. A. and O. Hernandez, 2009: Tracking blue whales in the eastern tropical Pacific with an ocean-bottom seismometer and hydrophone array. *The Journal of the Acoustical Society of America*, **126**(3), 1084–1094.

Faugeras, B. and O. Maury, 2005: An advection-diffusion-reaction size-structured fish population dynamics model combined with a statistical parameter estimation procedure: Application to the Indian Ocean skipjack tuna fishery. *Math. Biosciences and Engineering*, **2**(4), 719–741.

Ghil, M., 1989: Meteorological Data Assimilation for Oceanographers. Part I: Description and Theoretical Framework. *Dyn. of Atmos. Oceans*, **13**, 171–218.

Giles, M., D.Gahte, and MC.Duta, 2005: Using Automatic Differenciation for Adjoint CFD Code Development. *Post SAROD Workshop*.

Greene, W., 2008: Functional forms for the negative binomial model for count data. *Economics Letters*, **99**(3), 585–590.

Griffith, A. and N. Nichols, 2000: Adjoint methods in data assimilation for estimating model error. *Flow, turbulence and combustion*, **65**(3), 469–488.

Hansen, D. and P. Poulain, 1996: Quality control and interpolations of WOCE-TOGA drifter data. *Journal of Atmospheric and Oceanic Technology*, **13**(4), 900–910.

Hilbe, J., 2011: *Negative binomial regression*. Cambridge University Press, 2nd ed. edition.

Jamet, C. and H. Loisel, 2009: Data Assimilation Methods. **187**, 307–317.

Lawson, L., Y. Spitz, E. Hofmann, and R. Long, 1995: A data assimilation technique applied to a predator-prey model. *Bulletin of Mathematical Biology*, **57**(4), 593–617.

Lions, J., 1971: *Optimal Control of Systems Governed by Partial Differential Equations*. translated from French by S. K. Mitter, Springer-Verlag, New York.

Lorenc, A., 1986: Analysis methods for numerical weather prediction, Q. J. . *Roy. Meteor. Soc.*, **112**, 1177–1194.

Mangel, M. and P. Smith, 1990: Presence-absence sampling for fisheries management. *Canadian Journal of Fisheries and Aquatic Sciences*, **47**(10), 1875–1887.

Matear, R. J., 1995: Parameter optimization and analysis of ecosytem models using simulated annealing: A case study at Station P. *Journal of Marine Research*, **53**, 571–607.

Menemenlis, D., P. Fieguth, C. Wunsch, and A. Willsky, 1997: Adaptation of a fast optimal interpolation algorithm to the mapping of oceanographic data. *Journal of Geophysical Research*, **102**, 573–584.

Minami, M., C. Lennert-Cody, W. Gao, and M. Roman-Verdesoto, 2007: Modeling shark by catch: the zero-inflated negative binomial regression model with smoothing. *Fisheries Research*, **84**(2), 210–221.

Nocedal, J., 1980: Updating quasi-Newton matrices with limited storage. *Mathematics of Computation*, **24**(151), 773–782.

Senina, I., J. Sibert, and P. Lehodey, 2008: Parameter estimation for basin-scale ecosystem-linked population models of large pelagic predators: Application to skipjack tuna. *Progress in Oceanography*, **78**(4), 319–335.

Sibert, J. R., J. Hampton, D. A. Fournier, and P. J. Bills, 1999: An advection-diffusion-reaction model for the estimation of fish movement parameters from tagging data, with application to skipjack tuna (*Katsuwonus pelamis*). *Canadian Journal of Fisheries and Aquatic Sciences*, **56**(6), 925–938.

Spitz, Y., J. M. b, M. Abbott, and J. Richman, 1998: Data assimilation and a pelagic ecosystem model: parameterization using time series observations. *Journal of Marine Systems*, **16**, 51–68.

Tarantola, A., 1987: Inverse Problem Theory: Methods for Data Fitting and Model Parameter Estimation. *Elsevier, New York*.

Tarantola, A., 2005: *Inverse problem theory and methods for model parameter estimation*. Society for Industrial and Applied Mathematics.

Vallino, J., 2000: Improving marine ecosystem models: Use of data assimilation and mesocosm experiments. *Journal of Marine Reserach*, **58**, 117–164.

Vaudor, L., N. Lamouroux, and J. Olivier, 2011: Comparing distribution models for small samples of overdispersed counts of freshwater fish. *Acta Oecologica*, **37**, 170–178.

Wikle, C. K. and L. M. Berliner, 2007: A Bayesian tutorial for data assimilation. *Physica D: Nonlinear Phenomena*, **230**(1-2), 1–16.

Worley, B., 1991: Experience with the Forward and Reverse Mode of GRESS in Contaminant Transport Modeling and Other Applications, Proc. of SIAM Workshop on Automatic Differentiation of Algorithms: Theory, Implementation, and Application, Breckenridge, Colorado, January 6-8 307–314.

Chapter 7

Optimization of early life history parameters

Sommaire

7.1 Introduction 196

7.2 Optimization with climatology time series 197

7.2.1 Sensitivity Analysis 197

7.2.2 Correlated parameters 198

7.2.3 Parameters estimation 199

7.2.4 Exploring the effect of oxygen concentration 210

7.2.5 Exploring the impact of larvae retention for older larvae cohorts . . 210

7.3 Analysis of the sources of uncertainties 212

7.3.1 Optimization at coarser resolution 212

7.3.2 Sensitivity to data uncertainties 214

7.3.2.1 Impact of extreme high data values 214

7.3.2.2 Impact of the cost function 215

7.3.3 Impact of the vertical layer definition 216

7.3.4 Impact of mesoscale variability 218

7.3.5 Impact of physical inputs 219

7.3.6 Sensitivity to simulated predator inputs. 221

7.3.7 Conclusion 222

7.4 First optimization experiments at interannual scale 223

7.4.1 Optimization with the 1992-2000 run at 1/6° of resolution 223

7.4.2 Optimization with the run 2000-2006 at 1/9° of resolution 234

Bibliography 236

7.1 Introduction

Optimization experiments were carried out for the different spawning habitats defined in Chapter 5 for Peruvian anchovy and sardine. Table 7.1 presents the control parameters of simulation with their range of plausible values. Larvae diffusion parameter was fixed to 198 $m^2.s^{-1}$, for a range of scale phenomenon between 30 km and 120 km (Okubo, 1971).

θ	Description	$\underline{\theta}$	$\overline{\theta}$	θ^0	Units
σ_0	Standard deviation in the temperature function for eggs and larvae habitat, Eq. 2.3	0.5	5	4	°C
$\mathbf{T_0}$	Optimal surface layer temperature for eggs and larvae, Eq. 2.3	14	24	18	°C
a	Slope coefficient in the prey function, defining spawning habitat index, Eq. 2.4	-1	-0.001	-0.08	-
b	Threshold value of forage (PP), defining prey function in spawning habitat definition, Eq. 2.4	0.01	100	70	-
c	Slope coefficient in the function of predators, defining spawning habitat index, Eq. 2.5	0.01	200	50	-
d	Threshold value of predators (MTL), defining predator function in spawning habitat definition, Eq. 2.5	0.001	0.25	0.08	-
$\widehat{O}$	Threshold value of dissolved oxygen, defining oxygen function in spawning habitat definition, Eq. 2.6	0.0	50	400	$\mu mol.L^{-1}$
$\mathbf{q^*_{eggs,1}}$	Net efficiency for eggs using Hensen net, Eq. Eq. 6.17	0	1	1	-
$\mathbf{q_{eggs,2}}$	Net efficiency for eggs using Calvet net, Eq. Eq. 6.17	0	1	1	-
$\mathbf{q_{larvae,1}}$	Net efficiency for larvae using Hensen net, Eq. Eq. 6.17	0	1	0.95	-
$\mathbf{q_{larvae,2}}$	Net efficiency for larvae using Calvet net, Eq. 6.17	0	1	0.95	-
$\beta_{\mathbf{e,1}}$	Negative binomial parameter for eggs Hensen data, Eq. 6.19	0	1000	5	-
$\beta_{\mathbf{e,2}}$	Negative binomial parameter for eggs Calvet data, Eq. 6.19	0	1000	5	-
$\beta_{\mathbf{l,1}}$	Negative binomial parameter for larvae Hensen data, Eq. 6.19	0	1000	5	-
$\beta_{\mathbf{l,2}}$	Negative binomial parameter for larvae Calvet data, Eq. 6.19	0	1000	5	-
ϵ	Variability of anchovy mortality with habitat quality, Eq. 2.9	0.0	10.0	1.0	-
μ	Maximal mortality rate due to predation, Eq. 2.9	0	1.0	0.063	$5days^{-1}$
R	Maximal number of larvae at large spawning biomass of adults, Eq. 2.9	1.0	1000	200	-
$\mathbf{D^*}$	Larvae Diffusion, Eq. 2.8	-	-	198	$m^2.s^{-1}$

Table 7.1: **Control parameters of the constrained optimization problem, imposed lower ($\underline{\theta}$) and upper boundaries ($\overline{\theta}$) and initial guess values (θ^0).Parameters marked by asterisks were fixed at their specified values in all experiments.**

Based on the reasonable hypothesis that eggs do not escape from the net, the efficiency of the Hensen net for eggs ($q_{e,1}$) is set to 1. Larvae have likely more physical skills to escape the net and efficiency ($q_{l,1}$) needs to be estimated. When using both Hensen and Calvet nets, the efficiency of the Calvet net for eggs was fixed to 1 and the efficiency of Hensen net was estimated. For the latter, a value lower than 1 is expected due to a more dispersed sampling effort than with the Calvet net.

7.2 Optimization with climatology time series

One optimization experiments is performed for each of the 5 definitions of the spawning habitat (see Chapter 5). These first experiments used the climatology made from the Hensen net data. For anchovy, a total of 7624 observations of eggs and 7635 observations for larvae (over the year) have been used for the optimisation process. For sardine, we had 7642 observations for eggs and 7641 for larvae.

7.2.1 Sensitivity Analysis

Before running optimization experiments, the sensitivity of predictions and cost function to parameters were tested with the anchovy dataset. Performing an exhaustive sensitivity analysis, i.e., exploring entire likelihood hyper-surface in n-dimensional parametric space is practically impossible, and here we limited the analysis to a series of 9 random sets of parameters that was created using the climatological forcing ROMS-PISCES.

The figure 7.1 presents an example of results using the habitat definition which combines all mechanisms (function Hs5; Chapter 5). The sensitivity measures for the mortality parameter ϵ were persistently low for the model predictions and for the cost function. Only one from the nine random experiences showed significant sensitivity (Fig. 7.1), suggesting that ϵ may be not estimated with reliability. The others parameters are sensitive either to model predictions or the cost function (Fig. 7.1) and thus should be estimated.

It should be noted that the sensitivity analysis gives only an indication of how the parameters may be difficult to estimate. It is therefore useful to check if all parameters can be estimated within the particular parametric space of the actual data set.

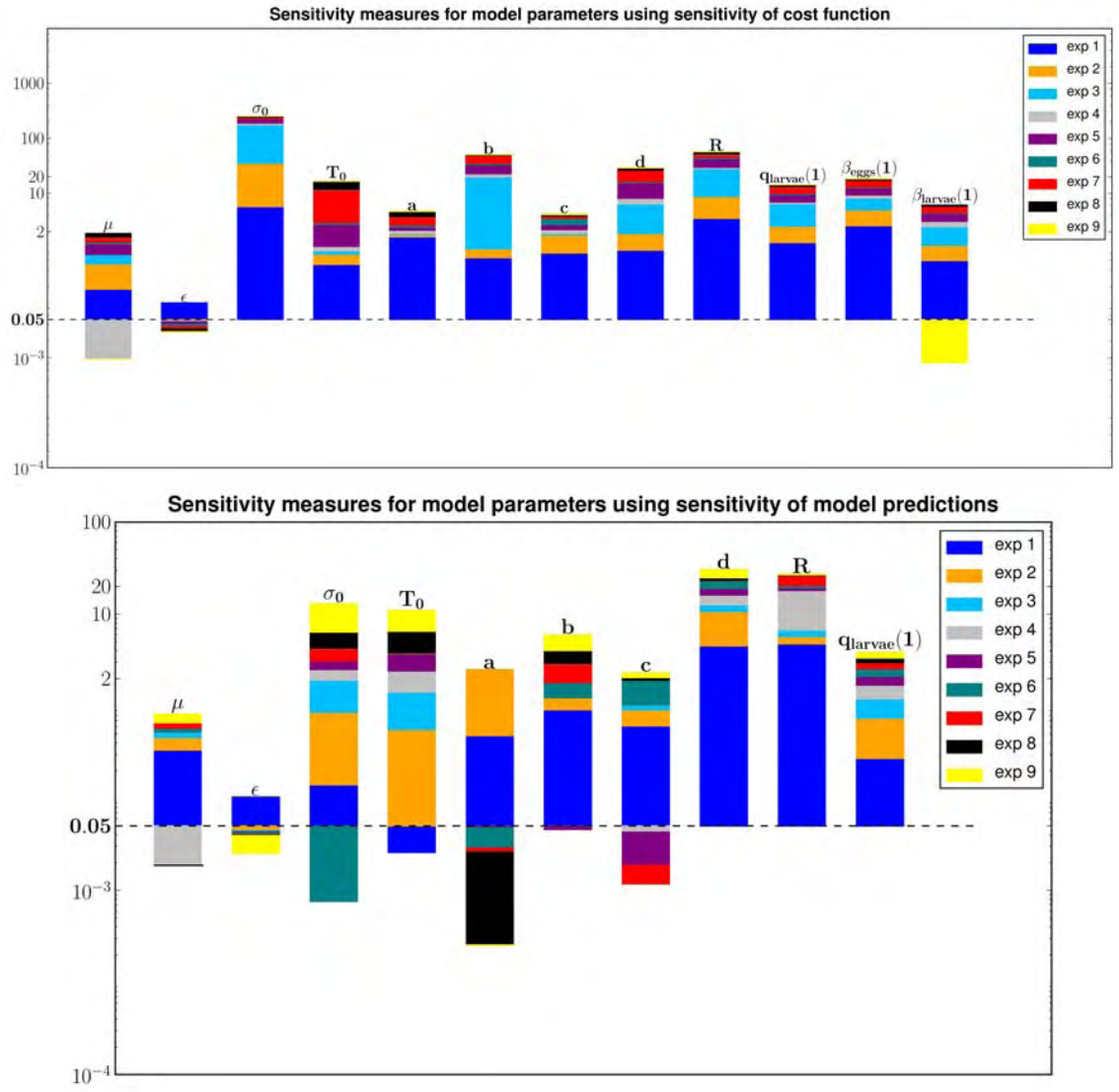

Figure 7.1: Log-scaled measures of sensitivity obtained for each parameter for 9 random experiments. The values below the dashed line correspond to less than 5% sensitivity of cost function. Top: Sensitivity measures for model parameters using sensitivity of cost function - Bottom: Sensitivity measures for model parameters using model predictions. The parameters β_n are the likelihood parameters to be estimated and are not sensitive to model predictions, then not shown.

7.2.2 Correlated parameters

Even when all parameters are sensitive, some of them can be correlated together and thus need to be fixed and estimated separately during the optimization experiments.

Correlations between pairs of parameters have been estimated from the analysis of the Hessian matrix (for more details, see section 6.4.3) which is computed at the minimum of the optimal solution.

Preliminary optimization experiments using artificial set of data from initial parameter guess (θ_0) (from Table 7.1) reveal high correlations for several pairs of parameters (Table 7.2).

	μ	ϵ	σ_0	T_0	a	b	c	d	R	$q_l(1)$
μ	1	**-0.996**	0.002	-0.089	-0.188	-0.131	0.08	0.054	-0.04	**0.997**
ϵ	**-0.996**	1	-0.006	0.103	0.196	0.121	-0.093	-0.053	0.022	**-0.987**
σ_0	0.002	-0.006	1	**-0.858**	0.134	0.271	-0.078	0.104	-0.093	-0.001
T_0	-0.089	0.103	**-0.858**	1	-0.091	-0.133	-0.013	-0.163	0.104	-0.075
a	-0.188	0.196	0.134	-0.091	1	0.771	0.09	0.163	0.081	-0.178
b	-0.131	0.121	0.271	-0.133	0.771	1	0.096	0.121	0.309	-0.137
c	0.08	-0.093	-0.078	-0.013	0.09	0.096	1	**0.804**	-0.457	0.061
d	0.054	-0.053	0.104	-0.163	0.163	0.121	**0.804**	1	**-0.763**	0.047
R	-0.04	0.022	-0.093	0.104	0.081	0.309	-0.457	**-0.763**	1	-0.057
$q_{l,1}$	**0.997**	**-0.987**	-0.001	-0.075	-0.178	-0.137	0.061	0.047	-0.057	1

Table 7.2: Correlation coefficients between optimal parameters obtained for **twin experiments simulation** with initial guess of parameters (θ_0).

The highest correlations occur between mortality parameters (μ and ϵ), and between these two parameters and the net efficiency coefficient for larvae ($q_{l,1}$). Relatively high correlations also exists between the parameters defining a same function, i.e., for the temperature (T_0 and σ_0), the prey of larvae (a and b) and the predator of larvae (c and d) and between the recruitment parameter R and d, the parameter that defines the threshold value of the predator function. This suggests a high sensitivity of the larvae survival rate to the amount of predators, likely because the value of d should be correlated to the extension of the favorable spawning habitat from the coast to offshore. However, as for the sensitivity analysis these correlations are an indication and may disappear with a real dataset.

7.2.3 Parameters estimation

A first optimization experiment using both eggs and larvae data collected with the Hensen net was conducted to estimate all parameters. However, as expected from the sensitivity analysis the parameters μ, ϵ and $q_{l,1}$ cannot be estimated due to a lack of sufficient information in the data and the correlation between these parameters. Thus, a first parameter had to be fixed. The catchability coefficient of the net for larvae being not critical in the case of this study, $q_{l,1}$ was set to 0.95, assuming that a small part of larvae (5%) can escape the net. Then, a new series of optimization experiments allowed to estimate the remaining parameters. The Hessian was definite positive for each experience indicating that a local minimum was achieved in both cases. The results for anchovy and sardine are summarized in the table 7.3 and the appendix (Table D) for the uncertainties of parameter estimates. These latter are small except for the mortality parameter ϵ with relative errors in the range of 52.4% to 17.8% respectively, confirming previous conclusions from the sensitivity analysis.

		σ_0	T_0	a	b	c	d	ϵ	μ	R	$\beta_e(1)$	$\beta_l(1)$	L
Anchovy	$Hs1$	**5**	17.44	-	-	-	-	**0.001**	0.971	222740	0.0479	0.0998	44384.7
	$Hs2$	-	-	-0.085	54.35	-	-	0.0624	0.991	274825	0.05	0.103	44243.4
	$Hs3$	-	-	-0.0687	42.37	12.07	**0.001**	1.60	0.61	796498	0.05	0.105	44180.8
	$Hs4$	**5**	17.38	-0.0992	52.47	-	-	**0.001**	0.954	313363	0.05	0.103	44232.2
	$Hs5$	**5**	19.04	-0.076	47.97	12.09	**0.001**	0.024	0.98	947739	0.05	0.104	44196.9
Sardine	$Hs1$	**5**	18.55	-	-	-	-	**0.001**	0.852	23172.3	0.012	0.021	12011.4
	$Hs2$	-	-	-0.216	39.99	-	-	1.637	1.702	22700.2	0.012	0.021	11970.0
	$Hs3$	-	-	-0.229	36.84	22.86	0.103	**2**	1.05	35866	0.012	0.022	11944.5
	$Hs4$	**5**	19.92	-0.236	39.65	-	-	1.85	1.60	25225.9	0.012	0.022	11968.2
	$Hs5$	**5**	22.16	-0.204	36.082	26.35	0.079	**2**	0.84	62875	0.012	0.022	11939.6

Table 7.3: Results of optimization experiments using eggs and larvae climatological data for anchovy and sardine (NB2 Log Likelihood). Grey shaded cells show the best likelihood solution.

The β parameter characterizing data overdispersion is two times higher for eggs than for larvae data in both species.

The standard deviation of the temperature function (σ_0) cannot be estimated and reached the upper fixed boundary value in all cases. Optimal temperature however was estimated with higher values for sardine than for anchovy in agreement with the literature (e.g., Schwartzlose et al., 1999) and chapter 3 (Figs. 3.27 and 3.28), but within a relatively large range according to the mechanisms used in the definition of the spawning habitat. These results would thus confirm those of the chapter 3 suggesting that the temperature is not the most critical parameter to define the spawning habitat, at least in this climatological representation.

In all cases, the parameterization of the prey function gives a more linear relationship between primary production (the proxy for food of larvae) and the spawning index for anchovies (Fig. 7.2 a). This is in accordance to the biological knowledge of anchovy and sardine behaviour. Indeed, sardine is known to feed on smaller zooplankton (smaller copepods and fewer euphausiids) than anchovy (Espinoza and Bertrand, 2008, Espinoza et al., 2009). Sardine is located in more oligotrophic waters, thus in accordance with the small observed threshshold of primary productivity.

For the predation function, the impact is estimated to occur earlier at low predator density for anchovies than for sardines (Table 7.3; Fig 7.2 b). This result agrees with the distribution of sardine eggs and larvae more dispersed offshore where mesopelagic organism concentration is higher than in the coastal zone.

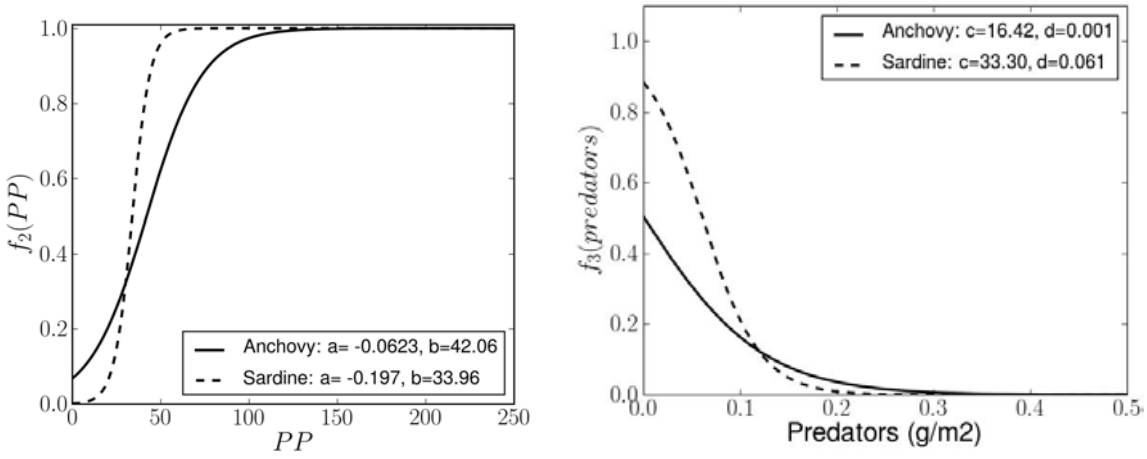

Figure 7.2: Functions of prey abundance (left) and predation (right) for the control parameters obtained after optimization experiments for anchovy and sardine for the best spawning habitat definition.

Given that the mechanisms investigated involve different number of parameters and variables, the likelihood results of the simulations are compared using the Akaike information criterion (AIC) to search for the most parsimonious model (Eq. 7.1). Amongst different models, the best one is the model leading to the minimum AIC value.

$$AIC = -2ln(P) + 2\mu = 2L + 2\mu \tag{7.1}$$

with μ the number of degrees of freedom (i.e., the total number of parameters and variables), P the maximal likelihood and L the negative log likelihood.

For both species, the AIC distinguishes the simulations Hs3 and Hs5 from the others, while the simulation Hs1 based on temperature only gives the lowest fit. The simulation Hs3 (prey-predator trade off) is better than Hs5 (prey-predator and temperature effects) for the anchovy but the situation is reversed for sardines (Table 7.4). Thus, temperature added to prey-predator trade-off mechanism can improve the final definition of the spawning habitat and larvae dynamics but seems of secondary significance. However, these results are based on a climatological series with limited changes in the seasonal cycle of temperature. The larger variability associated to El Niño events could eventually reveals a stronger effect of temperature. Thus, in the following simulations we will use the spawning definition Hs5, but releasing the upper boundary of σ to allow a maximum flexibility, i.e., no impact of temperature.

		L	Nb parameters	Nb Variables	AIC
Anchovy	$Hs1$	44384.7	7	1	88785.4
	$Hs2$	44243.4	7	1	88502.8
	$Hs3$	44180.8	9	2	88383.6
	$Hs4$	44232.2	9	2	88486.4
	$Hs5$	44196.9	11	3	88421.8
Sardine	$Hs1$	12011.4	7	1	24038.8
	$Hs2$	11970.0	7	1	23956.0
	$Hs3$	11944.5	9	2	23911.0
	$Hs4$	11968.2	9	2	23958.4
	$Hs5$	11939.6	11	3	23907.2

Table 7.4: Results of AIC for anchovy and sardine optimization experiments using different spawning habitat functions (see Table 5.1).

Annual spatial distributions obtained with the optimal parameterization are shown in Figure 7.4 for anchovy and sardine. Spatial distributions by seasons are also represented in Figures 7.6,7.7,7.8 and 7.9.

Theoretically, if the estimated parameters can explain all observations, we should obtain a linear relationship between observation and predictions. This is for example what we obtained for twin experiments. Using actual eggs and larvae datasets, we do not obtain such a linear relationship, showing that the model is (obviously) not able to achieve such a good fit with observations (see e.g., Fig. 7.3).

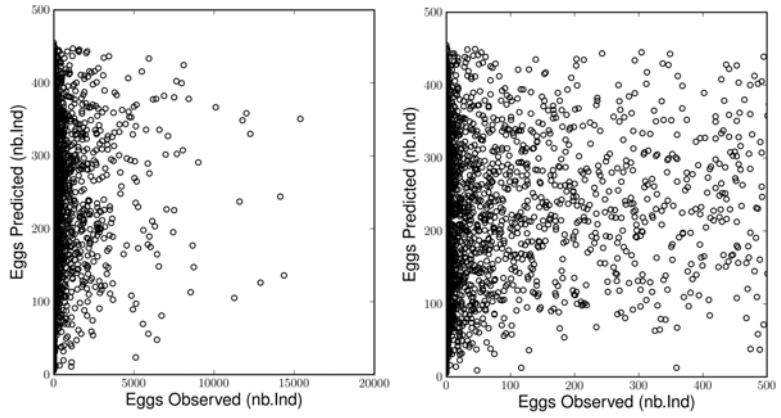

Figure 7.3: Example of observations versus predictions for eggs. A perfect fit would result in a linear relationship. The left plot is for all the samples and the right plot for a reduced range of eggs abundance.

This was expected however, due to the characteristics of the data as described in Chapter 3, with high overdispersion and heterogeneity. Thus, we do not expect the model to be able to fit exactly these datasets, but rather to capture the low frequency signal of the data. We note however that for anchovy (resp. sardine), the annual predicted

mean of eggs density is equal to 182.26 $Ind.m^2$ (resp. 20.97 $Ind.m^2$), in accordance with observations: 191.31 $Ind.m^2$ (resp. 22.15 $Ind.m^2$). For larvae density we observed also a good accordance: 49.13 $Ind.m^2$ (resp. 6.24 $Ind.m^2$) estimated versus 55.57 $Ind.m^2$ observed (resp. 6.36 $Ind.m^2$) .

From an overall spatial point of view, this low frequency signal seems effectively extracted as shown on Figure 7.4 with predictions normalized between 0 and 1 and observations log-transformed before similar normalization. Both observations and predictions display the same coast-offshore and north-south pattern. Northern regions are more favorable for larvae than for eggs in agreement with observations. For sardine, the observed maximum density north of 10°S is also well predicted despite lower data coverage.

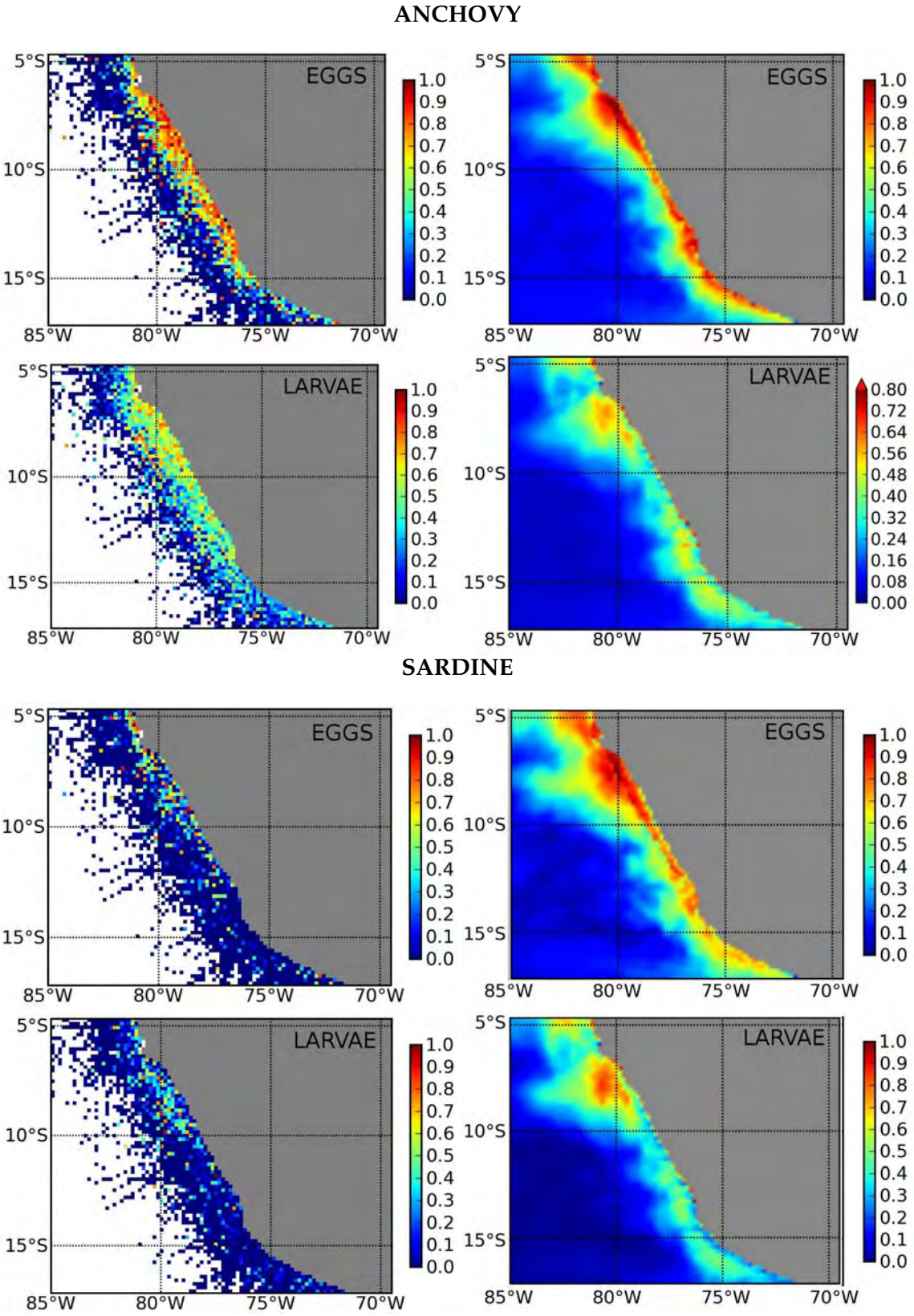

Figure 7.4: Anchovy and Sardine observed (left) and predicted (right) spawning habitat index ($\sim$ eggs abundance, top) and larvae density (bottom). Predicted density averaged is normalized between 0 and 1. Observations have been transformed using the log10 before normalization. Annual average of observed data is made with all available data from 1961 to 2008. Due to lower sampling offshore, data smoothing used for producing the average maps may introduce some biases in these regions, thus requiring cautious interpretation of these values.

Concentration of eggs and larvae occur all along the coast for both species during austral autumn and austral winter seasons (Fig. 7.4).

As far as seasonality is concerned, spawning habitat and larvae density of anchovy are predicted to be the most favorable in austral spring (October to December), whereas in observations, September and March month are the ones representing the higher abundances. Also, the very low spawning activity observed in austral autumn (April to June) is not well predicted (Figs. 7.6 and 7.7), with persistent high spawning index and larvae densities in coastal areas (Fig. 7.5). The result is better in offshore areas, with a good coincidence between observed and predicted periods of minimum spawning intensity in the second quarter (austral autumn), though here also the peak of maximum intensity is predicted to occur later (4th quarter) than in the observation (3rd quarter).

For sardines, spawning habitat and larvae density are predicted to be more favorable in austral summer in agreement with observations but the two observed periods of lowest intensity in austral autumn and austral spring (Figs. 7.8, 7.9 and 7.5) are not well predicted.

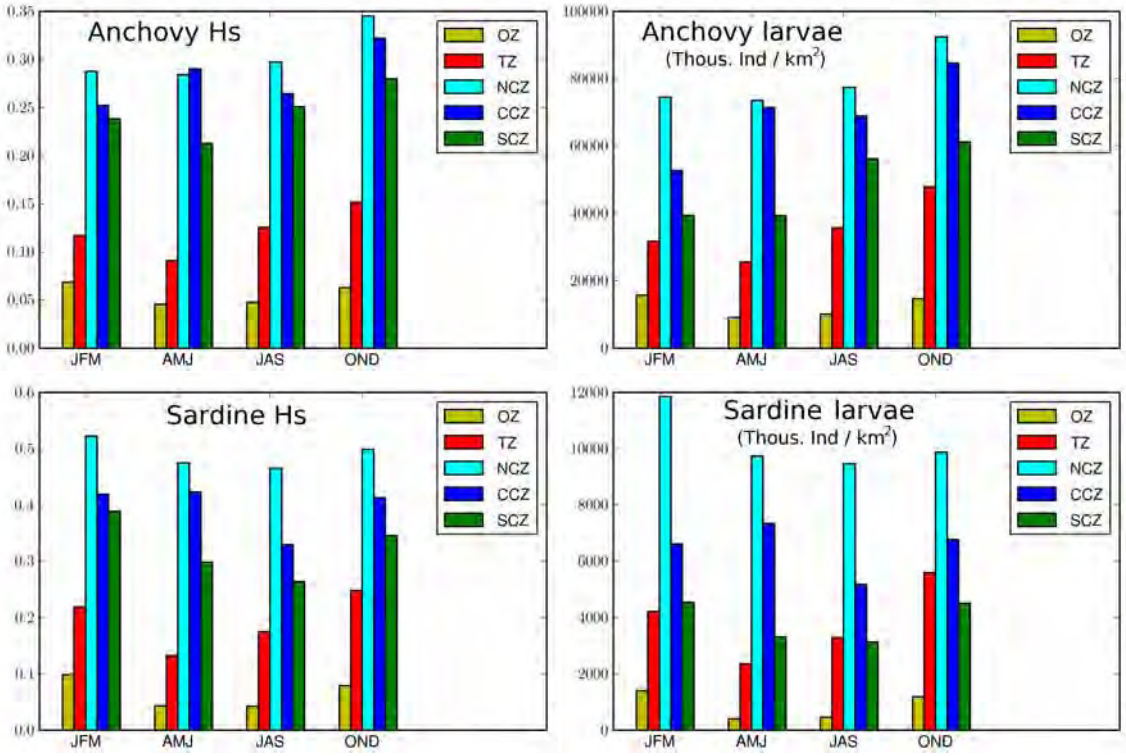

Figure 7.5: Seasonality by region and season of anchovy (top) and sardine (bottom) spawning habitat index (left) and anchovy larvae abundance (right) for the two best spawning habitat definition, Hs3 and Hs5 respectively.

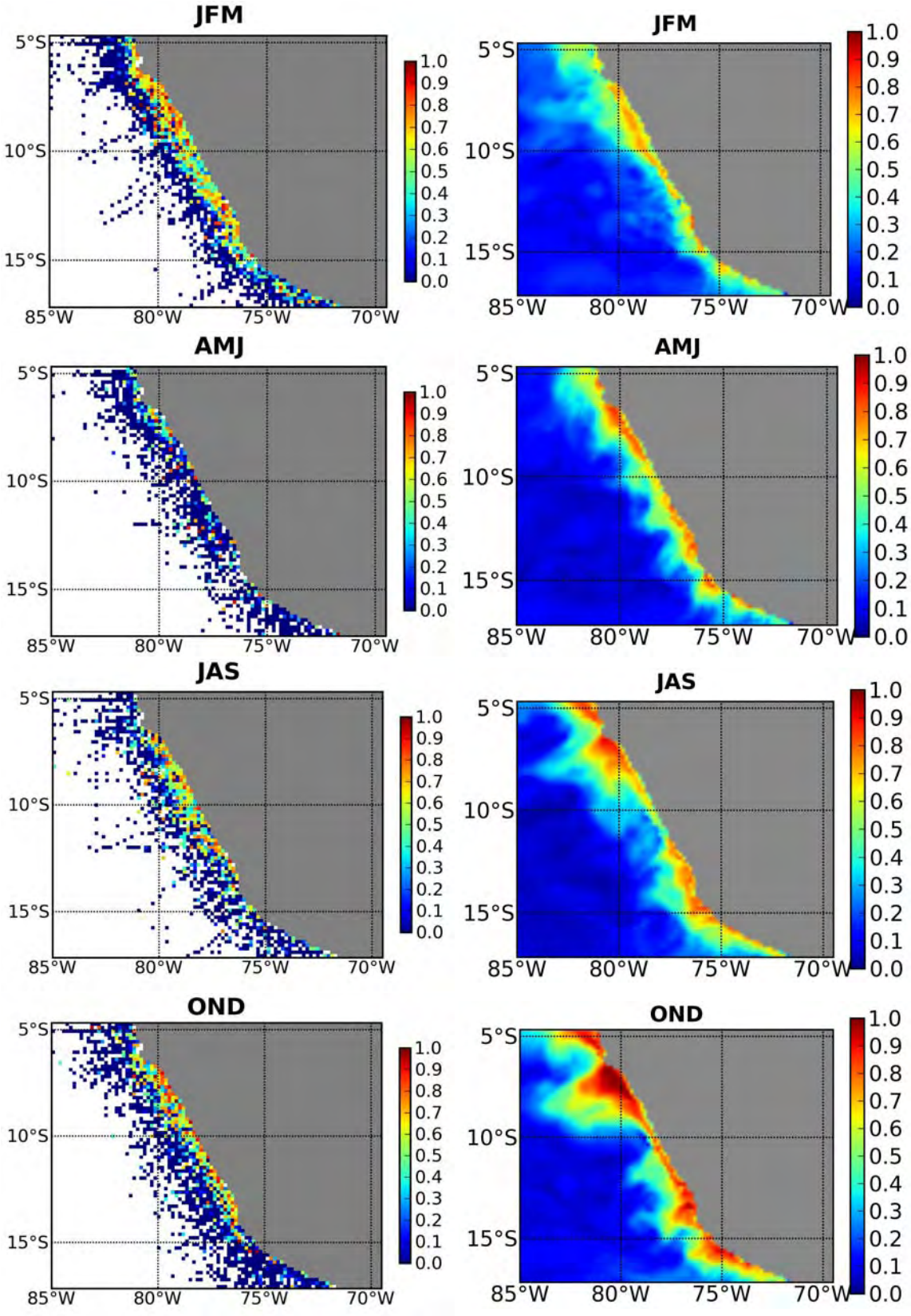

Figure 7.6: Anchovy observed (left) and predicted (right) spawning habitat index ($\sim$ eggs abundance, top) by season. Predicted density averaged is normalized between 0 and 1. Observations have been transformed using the log10 before normalization. Seasonal average of observed data is made with all available data from 1961 to 2008. Due to lower sampling offshore, data smoothing used for producing the average maps may introduce some biases in these regions, thus requiring cautious interpretation of these values.

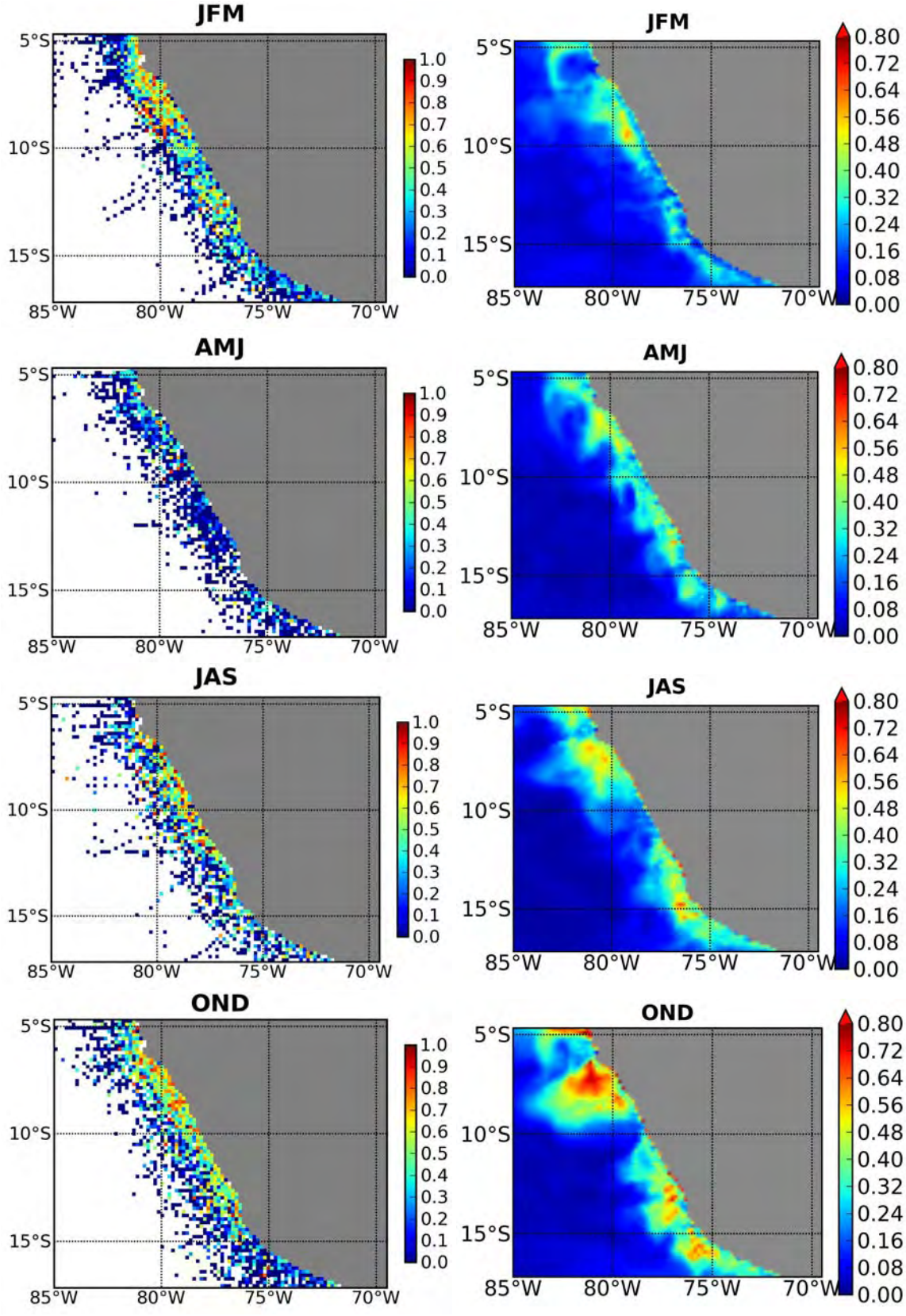

Figure 7.7: Anchovy observed (left) and predicted (right) larvae density by season. Predicted density averaged is normalized between 0 and 1. Observations have been transformed using the log10 before normalization. Seasonal average of observed data is made with all available data from 1961 to 2008. Due to lower sampling offshore, data smoothing used for producing the average maps may introduce some biases in these regions, thus requiring cautious interpretation of these values.

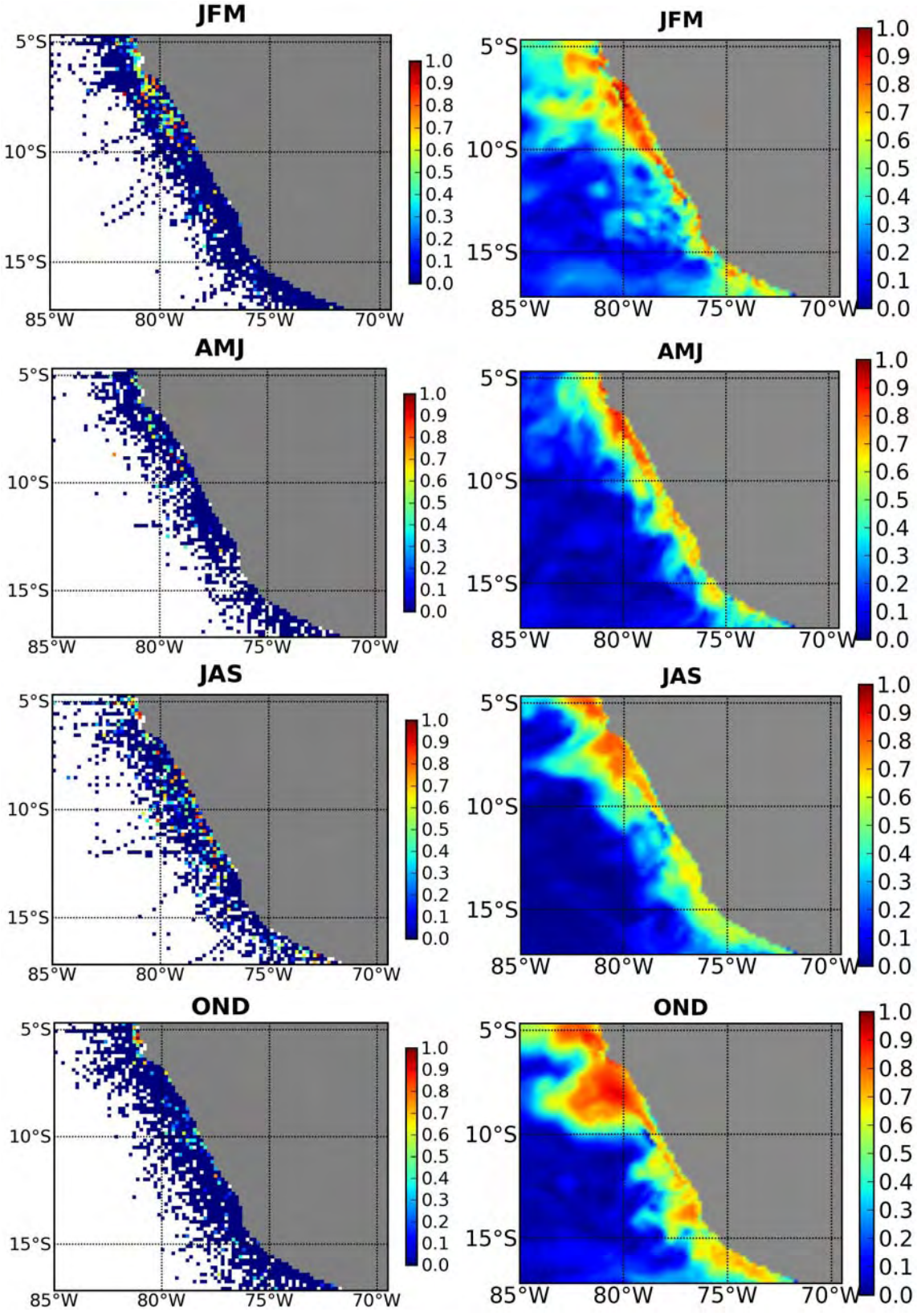

Figure 7.8: Sardine observed (left) and predicted (right) spawning habitat index ($\sim$ eggs abundance, top) by season. Predicted density averaged is normalized between 0 and 1. Observations have been transformed using the log10 before normalization. Seasonal average of observed data is made with all available data from 1961 to 2008. Due to lower sampling offshore, data smoothing used for producing the average maps may introduce some biases in these regions, thus requiring cautious interpretation of these values.

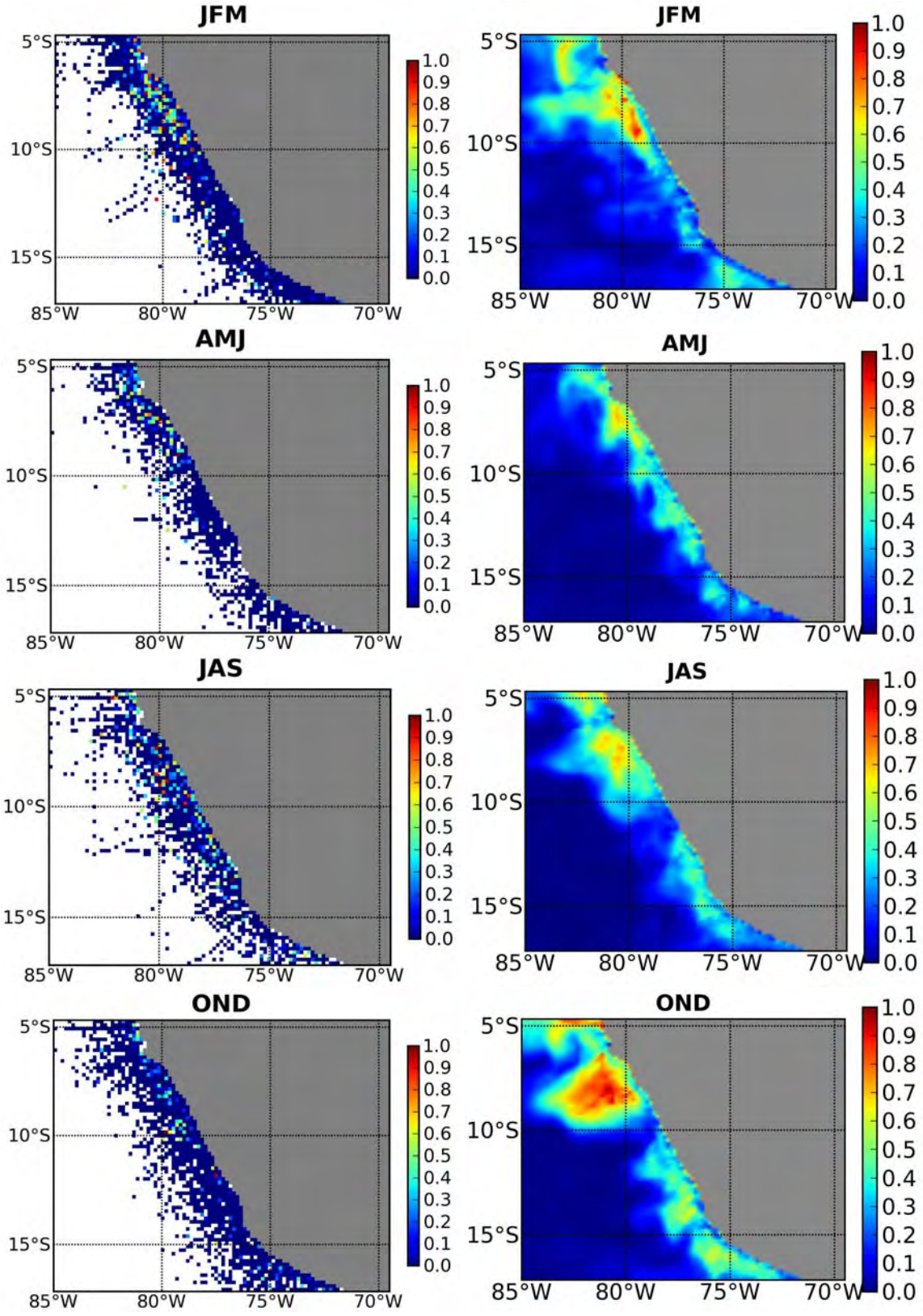

Figure 7.9: Sardine observed (left) and predicted (right) larvae density by season. Predicted density averaged is normalized between 0 and 1. Observations have been transformed using the log10 before normalization. Seasonal average of observed data is made with all available data from 1961 to 2008. Due to lower sampling offshore, data smoothing used for producing the average maps may introduce some biases in these regions, thus requiring cautious interpretation of these values.

7.2.4 Exploring the effect of oxygen concentration

In previous chapters, we have noted the importance of oxygen concentration in sardine's habitat. Despite that oxygen concentration predicted by the ROMS-PISCES model still shows disagreements with observations, we wish to test if our spawning habitat definition is sensitive to this variable. As described in Chapter 2, the effect of oxygen concentration in the spawning habitat was modeled by a sigmoid function (Eq. 2.6, Fig. 2.1d), with $\widehat{O}$, the threshold value of dissolved oxygen in $\mu mol.L^{-1}$ to be optimized.

We conducted optimization experiments for anchovy and sardine with this new spawning habitat definition (Table 7.5). Clearly, there was no sensitivity for anchovy nor sardine to this parameter. Indeed, the small improvement of the likelihood in the sardine case was not significant when studying the Aikaike information criterion (AIC). For sardine, the value of AIC for Hs5 spawning habitat definition is equal to 23904 (AIC=2L+2μ=2*11938+2*14), and for Hs6 it is equal to 23907.4 (AIC=2*11937.7+2*16).

Although no sensitivity to oxygen was found, it is too early to conclude on the global influence of oxygen due to the uncertainties on the modeled oxygen concentration in the ROMS-PISCES input. This question deserves further studies.

	σ_0	T_0	a	b	c	d	$\widehat{O}$	ϵ	μ	R	$\beta_e(1)$	$\beta_l(1)$	L
Anch - Hs5	**20**	15.04	-0.07	44.37	11.84	**0.001**	-	1.11	0.67	828494	0.05	0.105	44181.5
Anch - Hs6	**20**	15.04	-0.07	44.39	11.84	**0.001**	10.75	1.11	0.67	829148	0.05	0.105	44181.5
std	0.676	1.516	0.028	0.037	0.014	0.416	0.118	0.524	0.036	0.008	0.001	0.002	-
SDN - Hs5	7.55	24.29	-0.20	36.01	27.16	0.085	-	**2**	0.88	58988	0.012	0.0222	11938
SDN - Hs6	7.55	24.26	-0.20	36.00	27.18	0.086	15.75	**2**	0.88	58968	0.012	0.022	11937.7
std	0.149	0.186	0.075	0.022	0.035	0.087	0.049	0.419	0.039	0.004	0.001	0.002	-

Table 7.5: Results of optimization experiments using eggs and larvae climatological data for anchovy and sardine (NB2 Log Likelihood) using Hs5 (see Table 5.1 and Hs6 (Eq. 2.2) spawning habitat definition.

7.2.5 Exploring the impact of larvae retention for older larvae cohorts

The age of observed larvae was assumed to be around 5 days, which is one time step in the model. Here, we analyse the spatial distribution of the 3 older larvae's cohorts, which are respectively 10, 20 and 30 days old for the anchovy. Their distribution is based on the ADR equation 2.8, using a fixed mortality term during all larval stages of life.
As shown in Figure 7.10, increasing age leads to an extension offshore of the area of high larvae densities. In average, the density is lower, but the spatial extension of high density values is larger. However, the seasonal cycle of abundance is not changed (Figure 7.11).

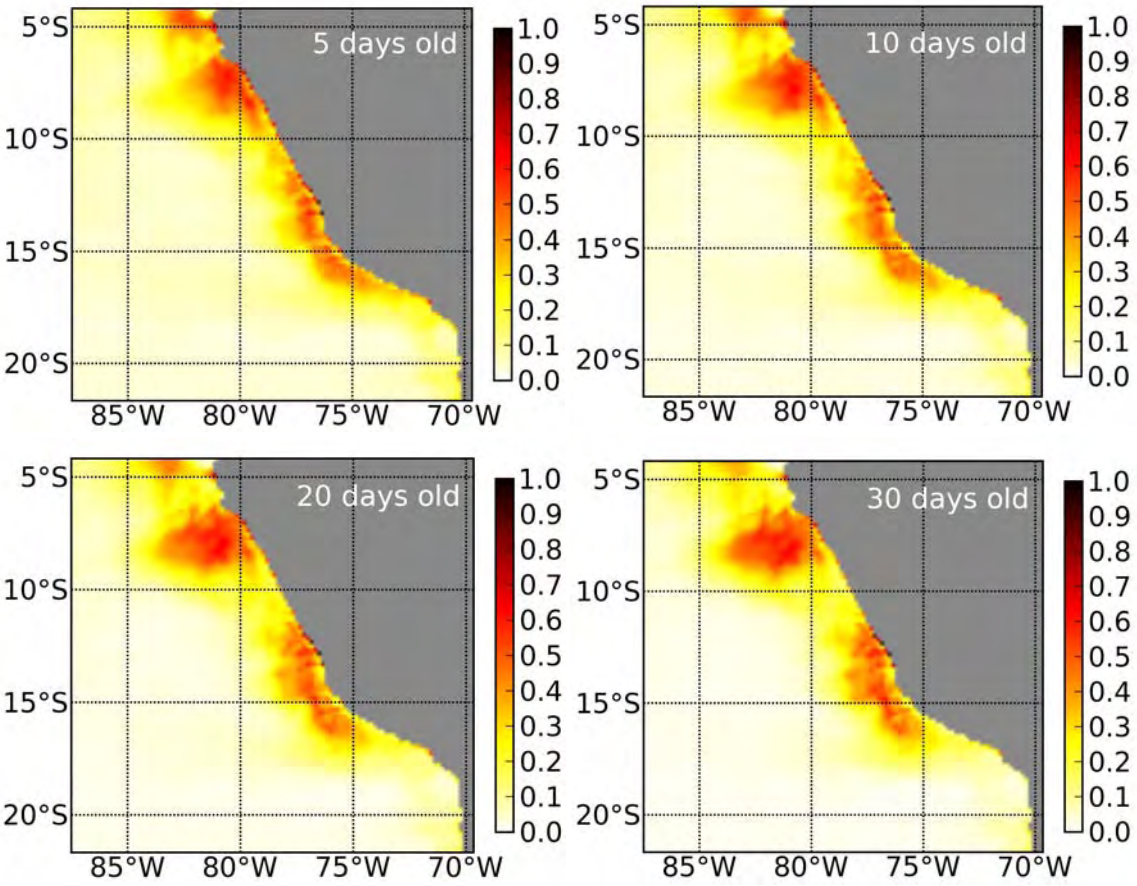

Figure 7.10: Comparison of the annual mean of larvae density spatial distribution old of 5, 10, 20 and 30 days old.

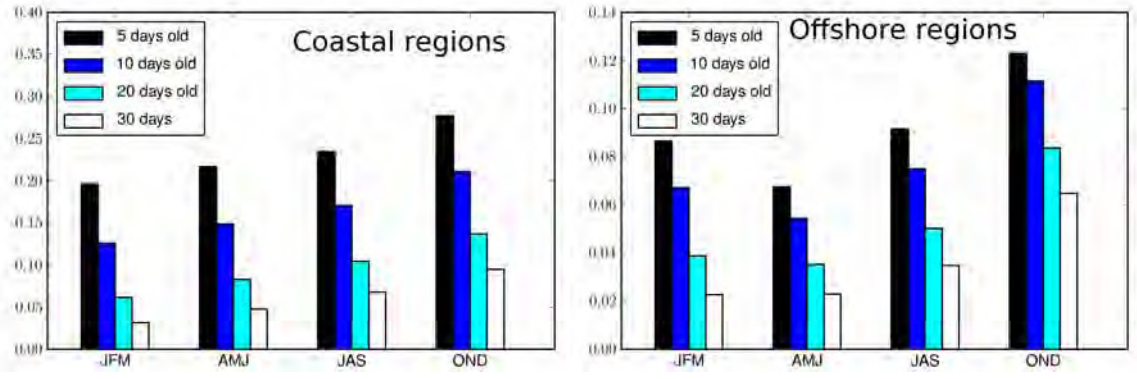

Figure 7.11: Seasonality of anchovy larvae abundance in coastal regions (left) and offshore regions (right) for larvae of 5,10, 20 and 30 days old. Each cohort of larvae density was normalized by its maximum over the year.

7.3 Analysis of the sources of uncertainties

The work presented in this study is the first attempt of modelling spawning and larvae dynamics of small pelagic species with a regional model based on an Eulerian approach and using data assimilation methods. However, in addition to missing mechanisms (e.g., stock - larval recruitment), various sources of uncertainties can explain the difficulty of simulating the range of absolute values of observed eggs and larvae densities.

Several variables are used to drive SEAPODYM simulations (Fig. 1.24) that each comes with its own uncertainty. They are not only the physical and biochemical variables provided by the ROMS-PISCES, but also the micronekton density distribution simulated with these environmental fields and the observations themselves used in data assimilation. In addition of course, the model also relies on assumptions and necessary simplified mechanisms. All these uncertainties are independent from each others, but can finally lead to a sum of biases and errors in the final predictions.

Sensitivity of optimization results to the main source of uncertainties is tested using the case of anchovy to search for potentially major source of errors that could provide guidance for further improvements in this approach.

Results of simulation experiments are compared through the changes in the likelihood value but also through qualitative analysis of changes observed in spatial and temporal distributions.

7.3.1 Optimization at coarser resolution

Observed eggs and larvae have been aggregated at 1/6° resolution (ROMS-PISCES model resolution) to be compared with predictions. Here, an optimization experiment was conducted using a coarser resolution of 2/3° for the data but keeping the same resolution (1/6°) for the model.

The mortality and recruitment parameters estimates differ, resulting in different absolute values of spawning index and larvae density. However, the other new parameter estimates (Table 7.6) are close to those obtained with the original resolution of 1/6°, and spatial distribution patterns for eggs and larvae (Fig. 7.12) as well as the seasonality did not change significantly (Fig. 7.13).

	σ_0	T_0	a	b	c	d	ϵ	μ	R	$\beta_e(1)$	$\beta_l(1)$	L
Anch $1/6$	**20**	15.04	-0.07	44.37	11.84	**0.001**	1.11	0.67	828494	0.05	0.105	44181.5
Anch $2/3$	10.13	14	-0.056	57.06	13.15	0.001	0.05	1.043	$1.08 * 10^6$	0.059	0.128	27609.5

Table 7.6: Results of optimization experiments using eggs and larvae data at $1/6°$ and $2/3°$ resolution using Hs_5 definition of spawning habitat.

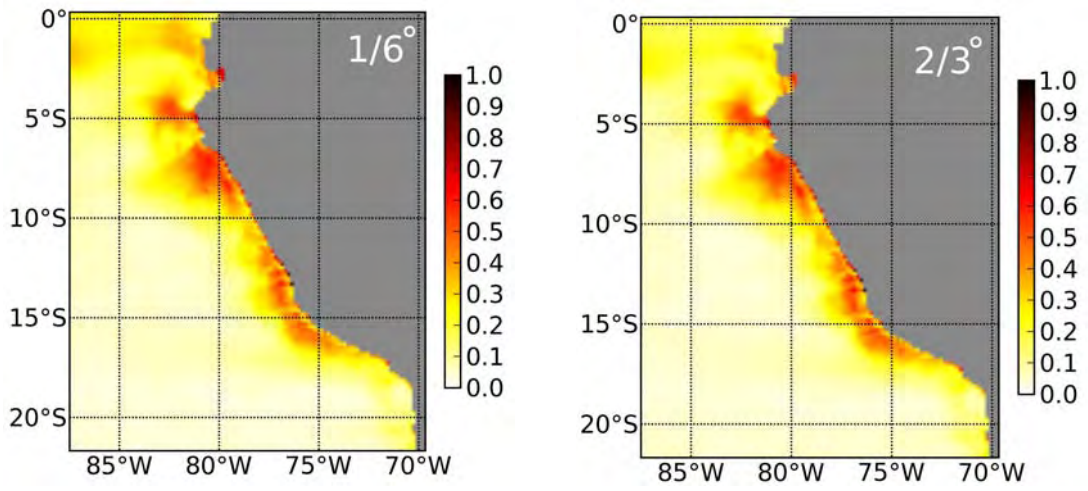

Figure 7.12: Comparison of the annual mean of spawning habitat spatial distribution for estimated parameters obtained at $1/6°$ of resolution (left, S1) and at $2/3°$ of resolution (right, S2).

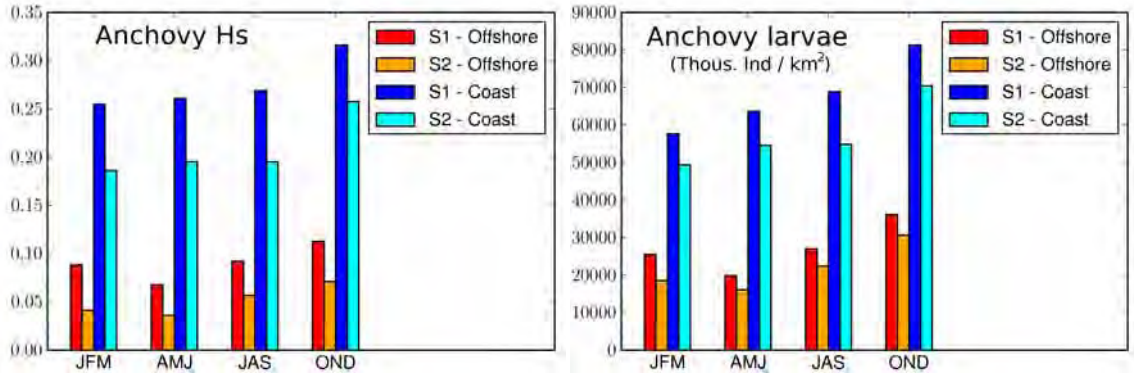

Figure 7.13: Seasonality of anchovy spawning habitat index (left) and larvae abundance (right) in coastal regions and offshore regions with estimated parameters obtained at $1/6°$ of resolution (S1) and at $2/3°$ of resolution (S2).

Finally, besides the high heterogeneity of observed data, the resolution used to make the optimization does not have a high impact on spatial and seasonal predicted patterns. In both cases, the model only captures the low frequency pattern.

7.3.2 Sensitivity to data uncertainties

7.3.2.1 Impact of extreme high data values

In some cases, a few extremely high values in the data may strongly influence the final likelihood. To test this potential effect, three experiments were conducted after removing respectively 0.2% (S2), 0.5% (S3) and 1% (S4) of highest values from the original climatological datasets (S1) of eggs and larvae (cf. chapter 3).

Estimated spawning habitat parameters were similar for each 3 experiments (Table 7.7), that is, in the range of uncertainties of estimated parameters (see Table 7.7). Indeed, the annual average of larvae density displays the same spatial patterns between all experiments. An example is shown with the parameter estimates of simulations S2 and S4 (Fig. 7.14).

	σ_0	T_0	a	b	c	d	ϵ	μ	R	$\beta_e(1)$	$\beta_l(1)$	L
S1	**20**	15.04	-0.07	44.37	11.84	**0.001**	1.11	0.67	828494	0.05	0.105	44181.5
S2	**20**	17.77	-0.079	46.04	10.94	**0.001**	0.79	0.83	748767	0.05	0.11	43841.7
S3	**20**	**14**	-0.062	48.12	10.9	**0.001**	0.76	0.82	703369	0.05	0.11	43202.6
S4	11.45	**14**	-0.095	49.60	10.48	**0.001**	0.98	**0.001**	617501	0.05	0.11	42211.8

Table 7.7: Results of optimization experiments after removing different percentages of highest data values in eggs and larvae density datasets.

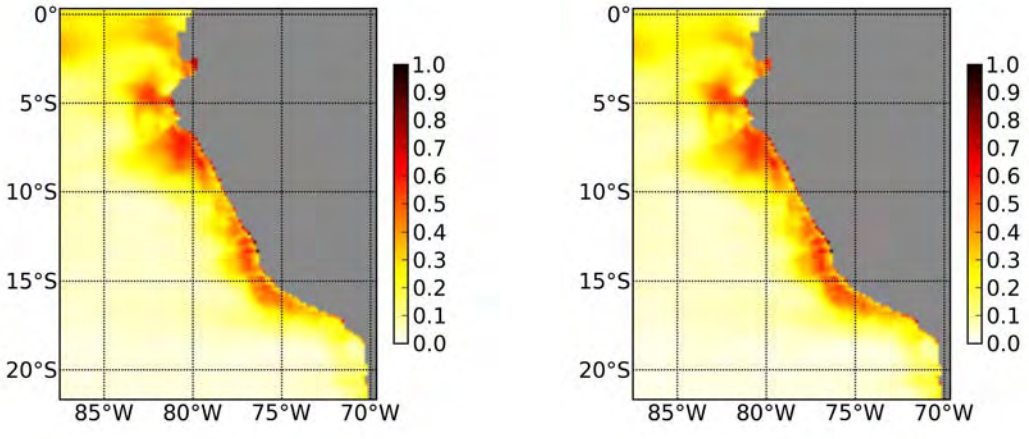

Figure 7.14: Comparison of annual average spatial distribution of spawning habitat for estimated parameters obtained with S2 (left) and S4 (right) experiment.

Removing extreme values lets the spatial distributions unchanged but decreases absolute estimated density (Fig. 7.15), as indicated also by differences in mortality and

recruitment parameters.

We note that these results could be expected, since by using Negative Binomial distribution, the weight of extreme values in the likelihood is drastically reduced.

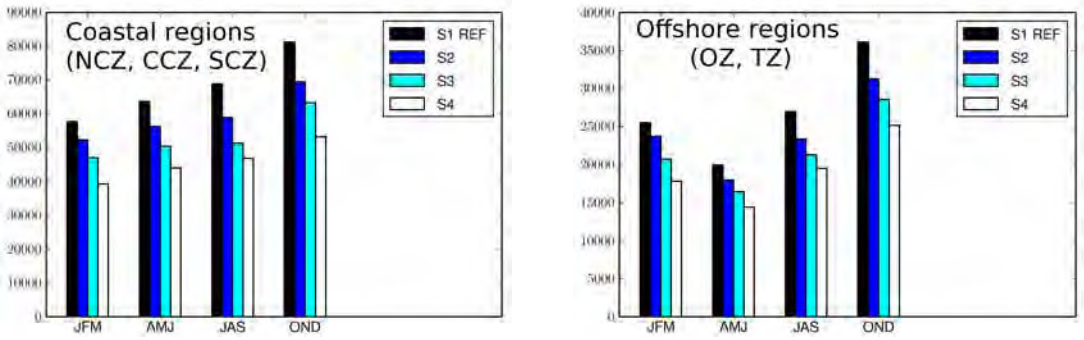

Figure 7.15: Seasonality of anchovy larvae abundance in coastal regions (left) and offshore regions (right) for the 4 experiments.

7.3.2.2 Impact of the cost function

The choice of the cost function has been discussed in chapter 6. The Negative Binomial distribution was chosen to account for overdispersed data, and because the model was not expected to predict very heterogeneous data with extreme range of densities. Clearly, the Normal, Log-Normal, Poisson and zero inflated negative binomial distributions were not suitable for the eggs and larvae datasets. We selected the second form of the Negative Binomial distribution (NB2) because in theory it allows to estimates the exact values of the data. Here we have tested the first form (NB1).

Parameter estimates and predictions with NB1 showed substantial differences with those achieved with the NB2 form (Table 7.8). However results with NB2 give a better agreement with spatial distributions of data, especially the coast-offshore gradient (Fig. 7.16). The seasonality predicted using these two cost functions present large differences (Fig. 7.17). None of them simulate the lowest spawning intensity in austral autumn. In the case of the simulation with the NB1 form, the maximum intensity is even in opposite phase with observation.

Exp	σ_0	T_0	a	b	c	d	ϵ	μ	R	$\beta_e(1)$	$\beta_l(1)$	L
NB1	12.14	**14**	-0.0186	123.67	9.575	**0.001**	**0.001**	1.135	$2.083{*}10^6$	0.00027	0.0022	43773.9
NB2	**20**	15.04	-0.07	44.37	11.84	**0.001**	1.11	0.67	828494	0.05	0.105	44181.5

Table 7.8: Results of optimization experiments using eggs and larvae climatological data for anchovy an sardine using Hs_5 as spawning habitat definition with NB1 and NB2 cost function.

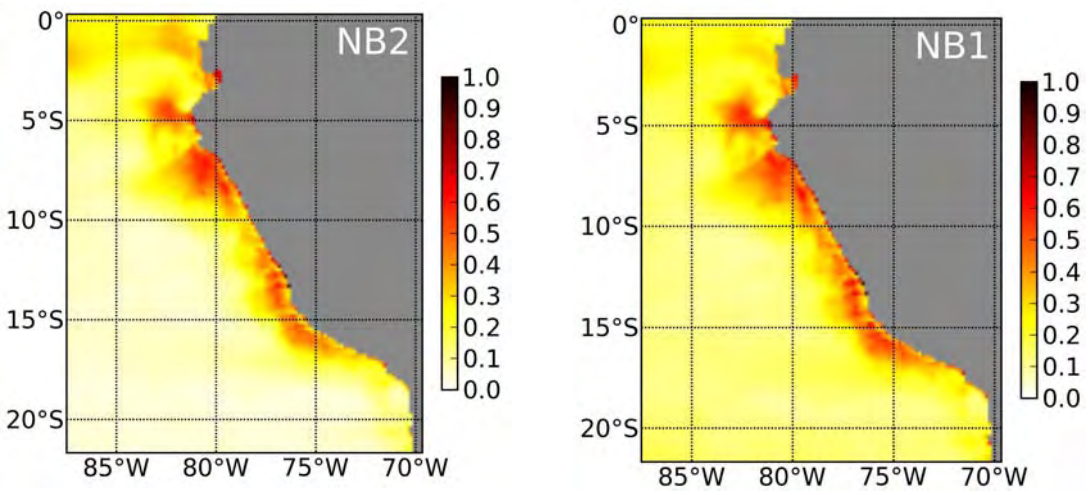

Figure 7.16: Comparison of the annual mean of larvae density for estimated parameters obtained with NB2 (left) and NB1 (right) cost functions.

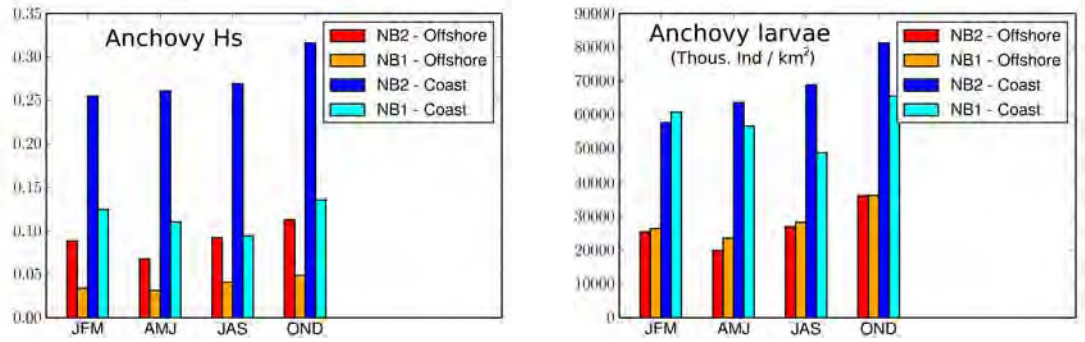

Figure 7.17: Comparison of the seasonality of spawning habitat (left) and anchovy larvae abundance (right) between NB1 and NB2 cost function.

In this case, the choice of the NB2 seems to provide the most plausible results. However, the way of taking into account the uncertainties of our dataset by using one or another cost function appears to be highly sensitive and should be probably further explored with different datasets or interannual runs.

7.3.3 Impact of the vertical layer definition

Since currents and temperature are averaged in the vertical layers (Chapters 3 and 4), their definitions can be critical in the simulated dynamics of eggs and larvae. This impact was tested with an optimization experiment using the euphotic depth instead of the mixed-layer, the euphotic layer being deeper everywhere.

In this case, the likelihood was slightly improved (Table 7.9), suggesting that physics averaged on the euphotic depth provides better dynamics for our observations. However, changes in parameters estimates of the spawning habitat are very limited; the differences remaining in the range of one standard deviation of the first parameter estimates. This

may be not too surprising since currents only affect larvae dynamics. Larvae density distribution is effectively modified, but only between latitudes 0°S and 5 °S where higher larvae density are predicted offshore (Fig. 7.19). The predicted seasonality is not very different from the simulation using mixed layer (Fig. 7.18).

	σ_0	T_0	a	b	c	d	ϵ	μ	R	$\beta_e(1)$	$\beta_l(1)$	L
S1	**20**	15.04	-0.07	44.37	11.84	**0.001**	1.11	0.67	828494	0.05	0.105	44181.5
S2	**20**	**14**	-0.09	38.22	12.43	**0.001**	1.25	0.74	799067	0.049	0.105	44173

Table 7.9: Anchovy optimization using vertical definition based on mixed layer depth (S1) and euphotic depth (S2) using Hs_5 definition of spawning habitat.

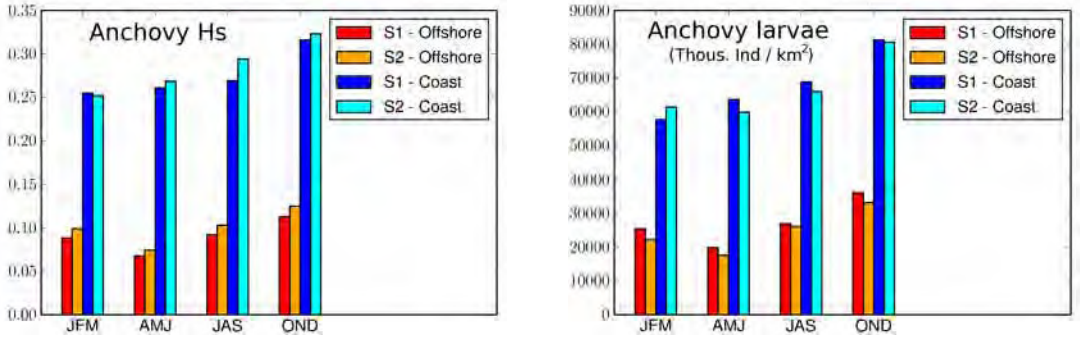

Figure 7.18: Seasonality of anchovy spawning habitat index (left) and anchovy larvae abundance (right) in coastal regions and offshore regions with estimated parameters obtained by averaging currents over the mixed layer depth (S1) or over the euphotic depth (S2).

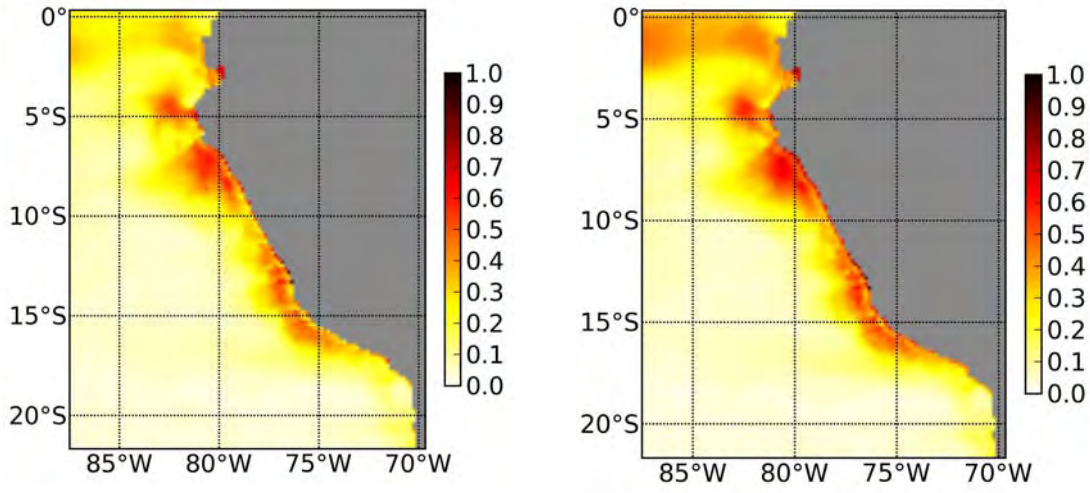

Figure 7.19: Comparison of annual average spatial distribution of larvae density for estimated parameters obtained by averaging currents over the mixed layer depth (S1, left) or the euphotic depth (S2, right).

Thus, predicted spawning habitat is rather insensitive to the definition of the vertical layers while there are only slight changes in the predicted distribution of larvae.

7.3.4 Impact of mesoscale variability

The climatological ROMS-PISCES simulation used to drive the SEAPODYM model produced intrinsic mesoscale variability due to the non linearity of the model. Thus, each simulated year has slightly different mesoscale activity which could impact the parameter estimation.

All previous experiments were run using the year 8 of the 10 years of climatological outputs from the ROMS-PISCES model (REF Year). The impact of intrinsic mesoscale variability was evaluated with three optimization experiments forced by three other climatological years (Table 7.10).

	σ_0	T_0	a	b	c	d	ϵ	μ	R	$\beta_{e,1}$	$\beta_l(1)$	L
REF Year 8	**20**	15.04	-0.07	44.37	11.84	**0.001**	1.11	0.67	828494	0.05	0.105	44181.5
Year 7	6.07	16.01	-0.05	65.53	11.28	**0.001**	0.001	0.965	$1.33*10^6$	0.05	0.105	44121.7
Year 9	7.75	**14**	-0.076	61.03	15.51	0.075	**0.001**	0.967	758742	0.05	0.106	44059.6
Year 10	9.19	**14**	-0.082	52.89	10.81	**0.001**	0.28	0.86	997155	0.05	0.105	44146.0

Table 7.10: Results of optimization experiments using different years of the 10 years of climatological run ROMS-PISCES.

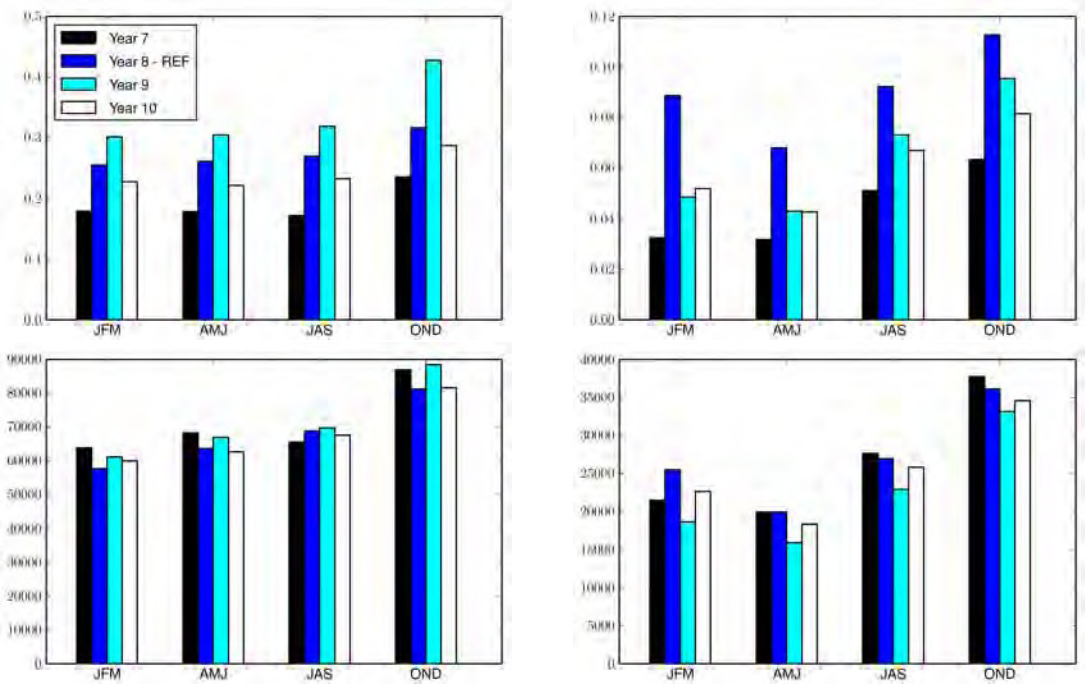

Figure 7.20: Seasonality of anchovy spawning habitat index (top) and larvae abundance (bottom) in coastal regions (left) and offshore regions (right) with estimated parameters obtained from 4 different years of the 10 years of the climatological run.

There were limited changes in estimated parameters of the spawning habitat (same order of uncertainty). Mortality and recruitment parameters differ, increasing the result-

ing eggs and larvae densities (Fig. 7.20), but the seasonality remains unchanged. These parameters are highly correlated (Table D.2), therefore variability in recruitment and absolute values can change significantly. The predicted distributions of eggs and larvae densities have similar large scale spatial patterns (Fig. 6.7.21).

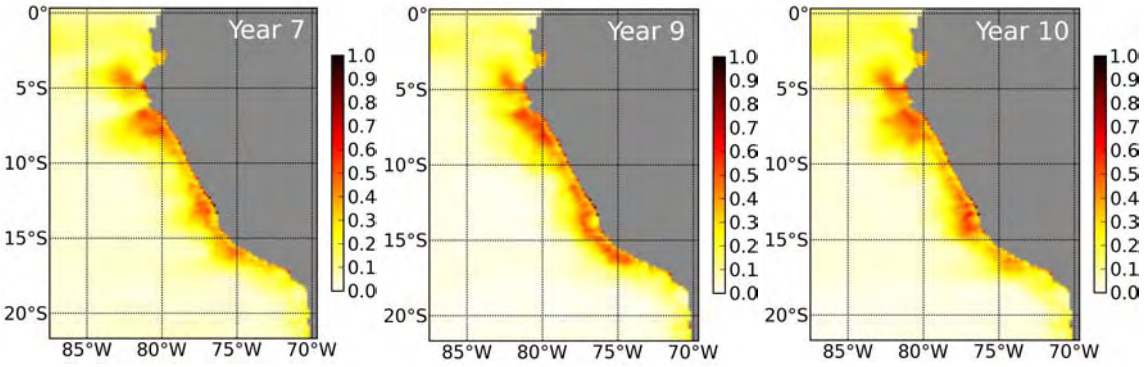

Figure 7.21: Comparison of annual average spatial distribution of larvae density for estimated parameters obtained for the different climatological years.

In summary, the model optimization relies more on large scale spatial variability and its seasonality than on mesoscale variability. Either the model, the forcing or the data (or a combination of all) do not allow to capture more than the first order of variability.

7.3.5 Impact of physical inputs

As noted in Chapter 4, some differences appear between sea surface temperature (SST) and primary productivity (PP) modelled by ROMS-PISCES and those derived of satellite data. Therefore, we made three optimization experiments using different combinations of physical forcing: (1) satellite SST and PP, (2) satellite PP and modeled SST, and (3) modeled PP and satellite SST.

The likelihood was slightly improved (Table 7.11) by using satellite PP and SST (simulation 1). Spatially, the use of satellite PP allowed to obtain a good coast-offshore gradient (Fig. 7.22) with no larvae predicted offshore, in good agreement with observations. However, the fit with observed seasonality was degraded both offshore (predicted minimum occurred in austral winter instead of austral autumn) and in coastal regions (predicted minimum in austral winter when observation indicate the highest densities) (Fig. 7.23).

The parameter estimation is highly sensitive to the type of primary productivity fields.

	σ_0	T_0	a	b	c	d	ϵ	μ	R	$\beta_e(1)$	$\beta_l(1)$	L
REF	**20**	15.04	-0.07	44.37	11.84	0.001	1.11	0.67	828494	0.05	0.105	44181.5
Anch PP sat	6.6	**14**	-0.026	157.99	24.00	0.13	**0.001**	0.900	626727	0.05	0.10	43959.6
Anch PP,SST sat	6.3	**14**	-0.026	164.116	31.01	**0.001**	0.923	0.148	688110	0.05	0.11	43934
Anch SST sat	17.1	**14**	-0.064	45.88	15.9	0.003	0.28	0.86	$1.06*10^6$	0.05	0.11	44149

Table 7.11: Results of optimization experiments using eggs and larvae climatological data for anchovy and 4 different physical forcing inputs.

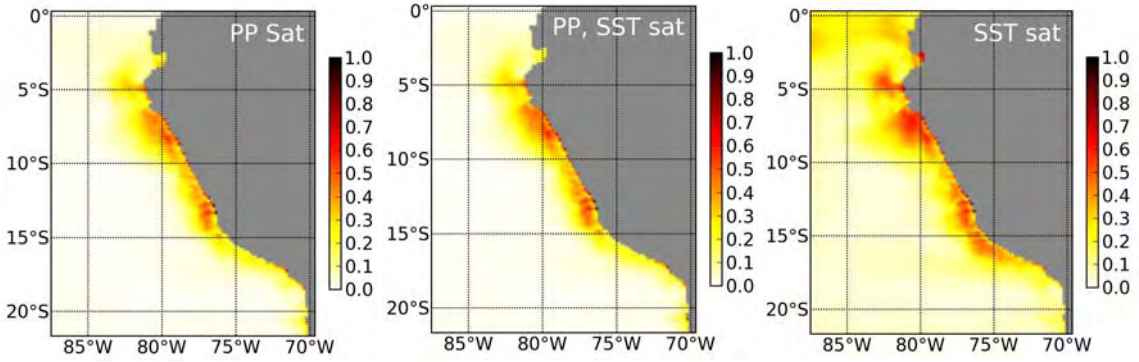

Figure 7.22: Comparison of the annual mean of larvae density spatial distribution for estimated parameters obtained by using satellite derived primary productivity (left), by using satellite derived primary productivity and satellite sea surface temperature (middle), and by using satellite sea surface temperature (right).

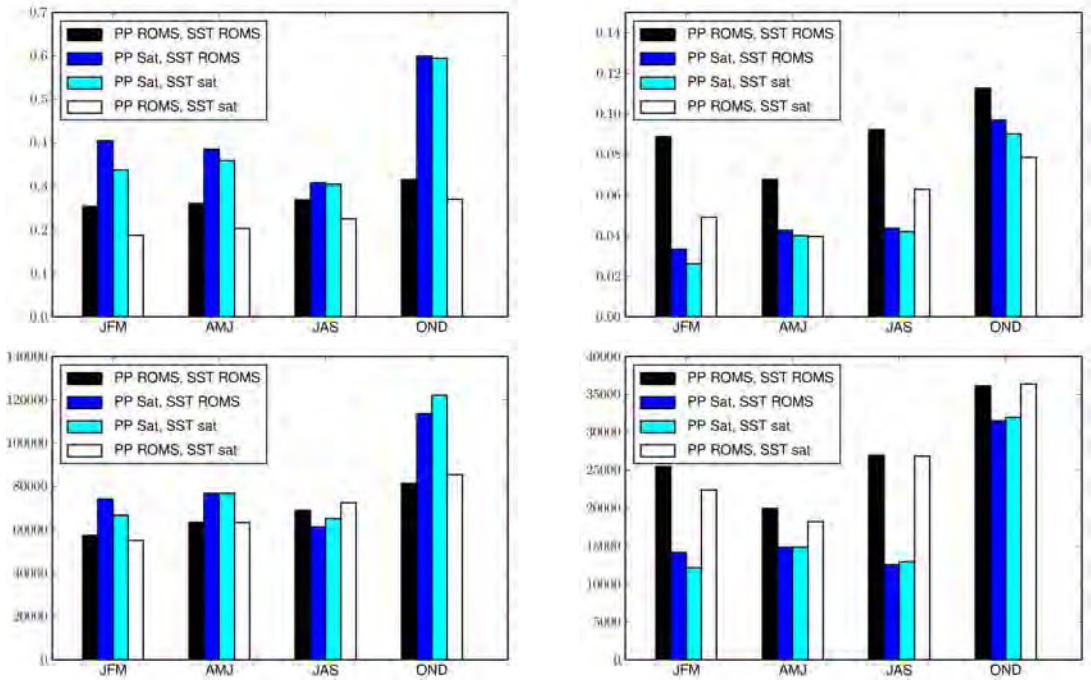

Figure 7.23: Seasonality of anchovy spawning habitat index (top) and larvae abundance (bottom) in coastal regions (left) and offshore regions (right) with estimated parameters obtained from 4 different physical forcings.

7.3.6 Sensitivity to simulated predator inputs.

As noted in Chapter 4, the definition of vertical layers strongly impacts the micronekton predictions and consequently the larvae's predator fields.

To analyze the impact of different micronekton outputs, we performed 3 new experiments (Table 7.12). In the first one (S2), the micronekton model used the satellite primary productivity and the euphotic depth to define the epipelagic layer, in the second (S3) it uses the model productivity and two times the euphotic depth, the third one (S4) it uses the model productivity and the mixed layer depth.

Estimated parameters for the prey function where modified in each simulation. However, none of the 3 new simulations improved either the observed seasonality (Fig. 7.25) or the spatial patterns (Fig. 7.24).

	σ_0	T_0	a	b	c	d	ϵ	μ	R	$\beta_e(1)$	$\beta_l(1)$	L
S1 - REF	**20**	15.04	-0.07	44.37	11.84	**0.001**	1.11	0.67	828494	0.05	0.105	44181.5
S2	10.09	**14**	-0.086	51.73	3.33	**0.001**	**0.001**	0.977	821204	0.05	0.103	44211.1
S3	**20**	**14**	-0.13	36.7	7.38	**0.001**	**2**	0.6	793067	0.05	0.107	44092.8
S4	8.08	**14**	-0.057	52.74	38.57	0.082	0.001	0.97	408145	0.05	0.104	44199.6

Table 7.12: Results of optimization experiments using eggs and larvae climatological data for anchovy and 4 different parameterizations of the micronekton model.

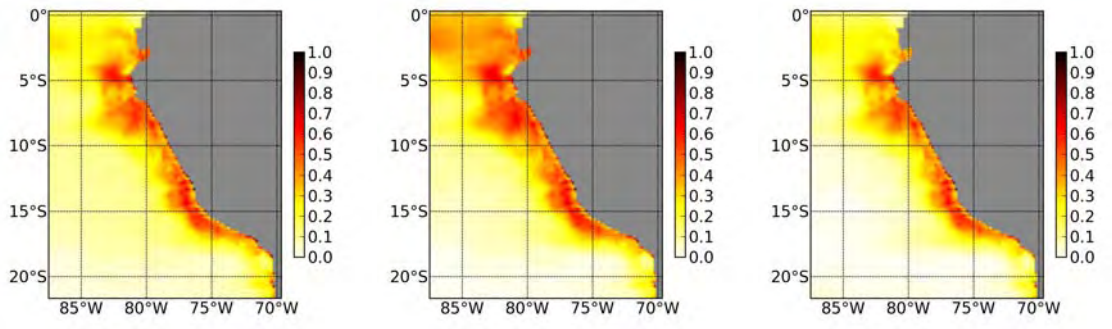

Figure 7.24: Comparison of the annual mean of larvae density spatial distribution for estimated parameters obtained by using the 3 parameterizations of the micronekton model S2 (left), S3 (middle) and S4 (right).

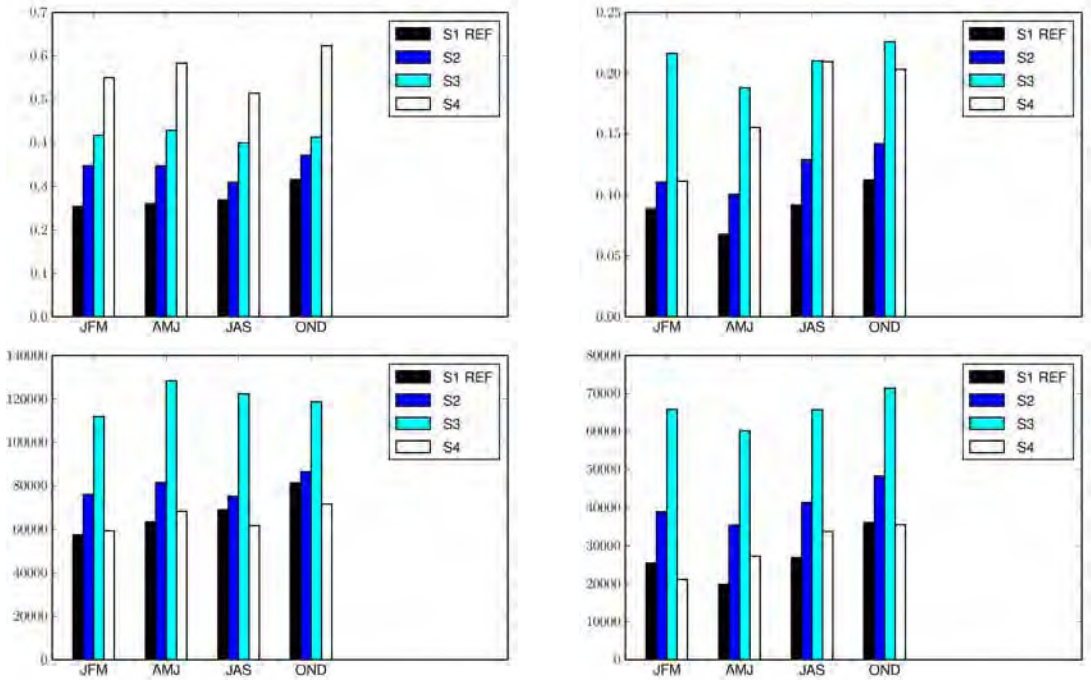

Figure 7.25: Seasonality of anchovy spawning habitat index (top) and larvae abundance (bottom) in coastal regions (left) and offshore regions (right) with estimated parameters obtained from 4 different parameterizations of the micronekton model.

Predicted simulations are highly sensitive to simulated predators inputs and by the definition of the epipelagic layer in the micronecton model. This impact should be further explored.

7.3.7 Conclusion

This sensitivity analysis shows that the main sources of uncertainty for the model predictions are primarily those on the data and how they are taken into account into the model. The others sources of major uncertainty are linked to the quality and definition of primary productivity and predators fields, especially through the definition of vertical layers and the average of currents in these layers.

Nevertheless, the parameter estimates in all optimization experiments remain quite stable. The best definition of spawning habitat cannot be easily selected from these series of experiments that clearly indicate that more than one mechanism is needed to approach observations.

No experiment improved significantly the fit between predicted seasonality and the observed cycle. Much higher resolution with realistic environmental forcing and additional mechanisms, like the relationship between adult spawning stock and larval recruitment (cf Fig. 3.34), are likely needed to get a more accurate predicted seasonality.

7.4 First optimization experiments at interannual scale

The following section presents the preliminary results of optimization experiments using Negative Binomial likelihood function (NB2) with the ROMS-PISCES environmental forcings 1992-2000 and 2000-2006.

7.4.1 Optimization with the 1992-2000 run at 1/6° of resolution

The data coverage for this series is shown on Figure 7.26. For anchovy, a total of 6722 observations of eggs and 6701 observations for larvae (from 1992 to 2000) have been used for the optimisation process. For sardine, we had 6677 observations for eggs and 6680 for larvae. The sampling with the Hensen net becomes more frequent in the second part of the series after 1996 and there is no data collected between January 1998 and March 1998 that is during the maximum development of the strongest El Niño of the series and the highest anomaly of the gonadosomatic index (Fig. 3.23). Consequently, the series of data contains almost no information characterizing the conditions during the strongest El Niño anomaly. However, a test carried out after adding false data with zero values during the months of January 1998 to March 1998 did not modify significantly the results.

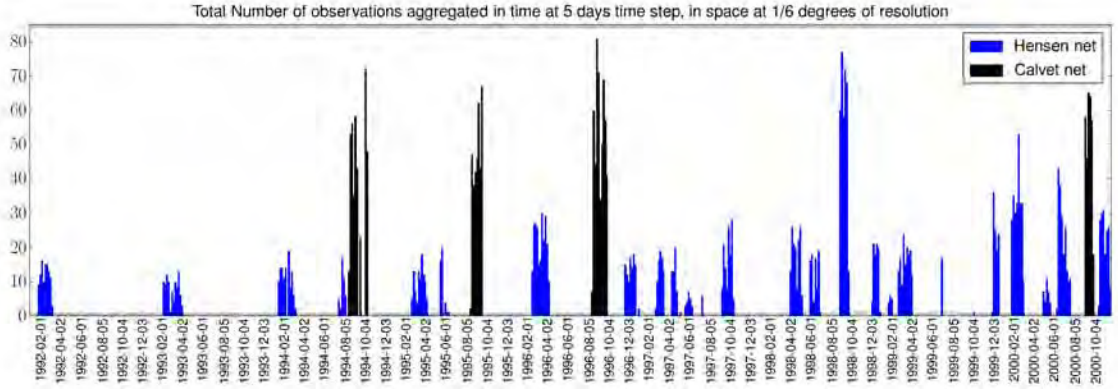

Figure 7.26: Total number of observations of eggs and larvae from 1992 to 2000 by net aggregated at the time and space model resolution.

The results for anchovy (A) and sardine (S) are summarized in the table 7.13 and the appendix (Table D.5 and D.6) for the correlation between parameters. Only Hensen net data were used for sardine (cf. Chapter 3).

Table 7.13: Results of optimization experiments using eggs and larvae data from Hensen and Calvet net from 1992 to 2000 for anchovy and only Hensen net for sardine species. Estimated parameters and corresponding standard deviation (std).

	σ_0	T_0	a	b	c	d	ϵ	μ	R	$q_{e,1}$	$q_{l,1}$	$\beta_{e,1}$	$\beta_{l,1}$	$\beta_{e,2}$	$\beta_{l,2}$	L
A	4.32	22.2	-0.05	46.9	**100**	1.09	**4**	0.66	$1.68*10^6$	0.27	0.15	0.055	0.12	0.13	0.28	24460
std	0.026	0.221	0.024	0.031	0.013	0.028	3.706	0.038	0.03	0.029	0.016	0.001	0.001	0.001	0.001	
S	3.06	23.1	-0.18	33.3	5.34	**0.0001**	**0.0001**	5.35	19242	-	-	0.014	0.03	-	-	2705
std	0.113	0.266	0.048	0.139	0.142	0.049	0.023	0.122	0.009	-	-	0.001	0.001	-	-	

The optimization with the interannual run produced a parameterization different from the one achieved with the climatological run (Fig. 7.4.1a). The parameter σ was estimated for both species suggesting a higher sensitivity to temperature than with the climatological run. It is worth noting that optimal temperature for sardine is estimates slightly higher than for anchovy as expected from the literature. For both species, parameters of the prey function estimated with the interannual run 1992-2000 lead to relationships very similar to those obtained with the climatology (Fig. 7.4.1b). This results corroborates the estimated parametrization where sardine is located in more oligotrophic waters. For the predators function however, the results are different. For anchovy, the prediction of spawning habitat and larvae densities are not at all sensitive to predators field (Fig. 7.4.1c). For sardine, the effect becomes more linear than with the climatological run (Fig. 7.4.1c).

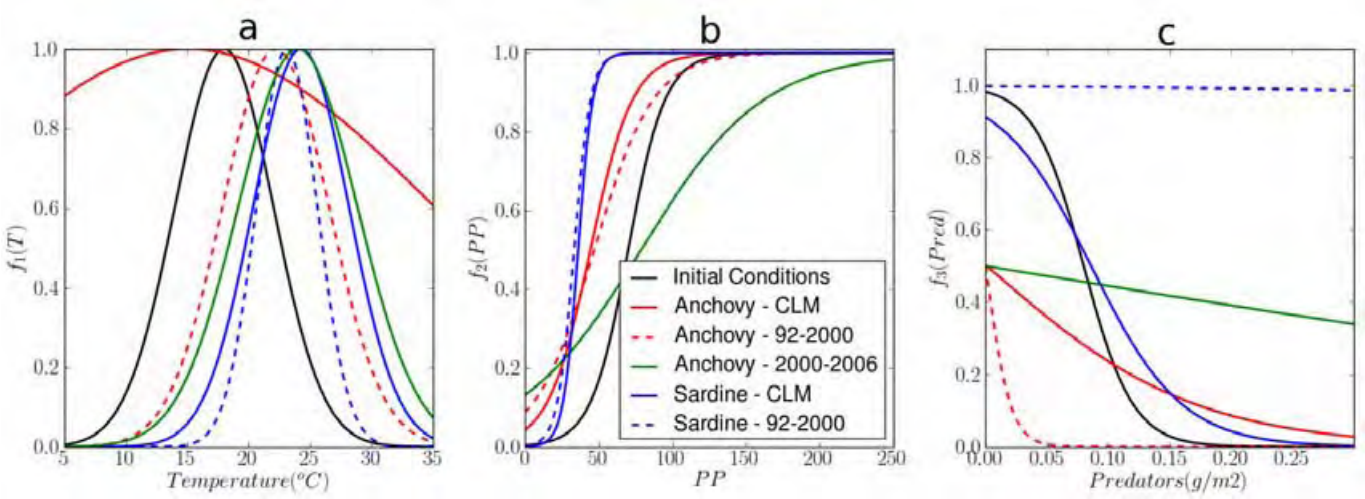

Figure 7.27: Functional relationships defining the spawning habitat: functions of temperature (a), prey availability (b), predation (c). resulting from optimization experiments with the climatological run, the interannual run 1992-2000 and the interannual run 2000-2006 forr anchovy and sardine.

Spatial distributions of anchovy and sardine resulting from this new optimization experiments (Figs. 7.28 and 7.29) are in fairly good agreement with observations. Anchovy is predicted to be located along the Peruvian coast. The northern coastal region (NCZ) is the most favorable region. Another favorable area is predicted in the vicinity of the Galapagos Islands. However, no anchovy have been found in this area (Grove and Lavenberg, 1997). For sardines, spawning habitat is predicted to be less coastal in accordance with the literature, as described in Bertrand et al., 2004: *"In fact, the remaining stock of sardine can be observed far offshore, as reported by Russian research vessels (M. Gutiérrez, pers. com.), or can take refuge in specific areas such as the Galapagos (Chalen, 2000)"*. However, spawning grounds appears very favorable in the equatorial region that

seem rather unrealistic.

During El Niño event, spawning habitat and larvae density for both species are predicted to become more concentrated in the very coastal zones with a shift towards the central and southern coastal regions (CCZ, SCZ), for anchovy more particularly (Figs. 7.30 and 7.32). These predictions agree very well with the knowledge on the behavior of adult anchovy during El Niño event (Fig. 3.15). However, the habitat is still perhaps not enough coastal since Galapagos Islands are still too favorable. For sardine, spawning habitat and larvae density are also predicted to be more coastal, though remaining favorable much far offshore than for anchovy and still including the region of the Galapagos Islands.

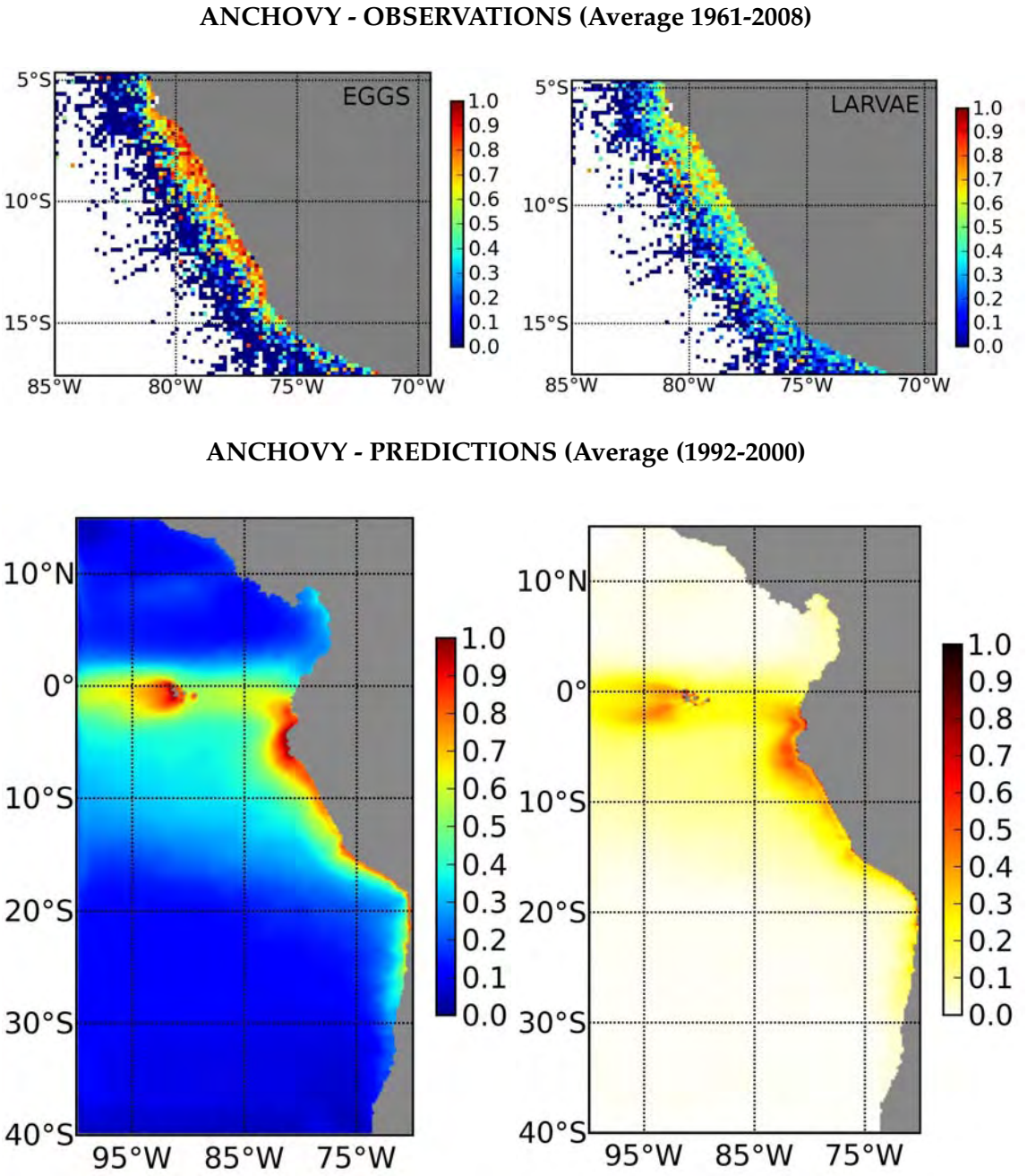

Figure 7.28: Anchovy observed (top) and predicted (bottom) spawning habitat index ($\sim$ eggs abundance, left) and larvae density (right). Predicted density averaged from 1992 to 2000 is normalized between 0 and 1. Observations have been transformed using the log10 before normalization. Annual average of observed data is made with all available data from 1961 to 2008. Due to lower sampling offshore, data smoothing used for producing the average maps may introduce some biases in these regions, thus requiring cautious interpretation of these values.

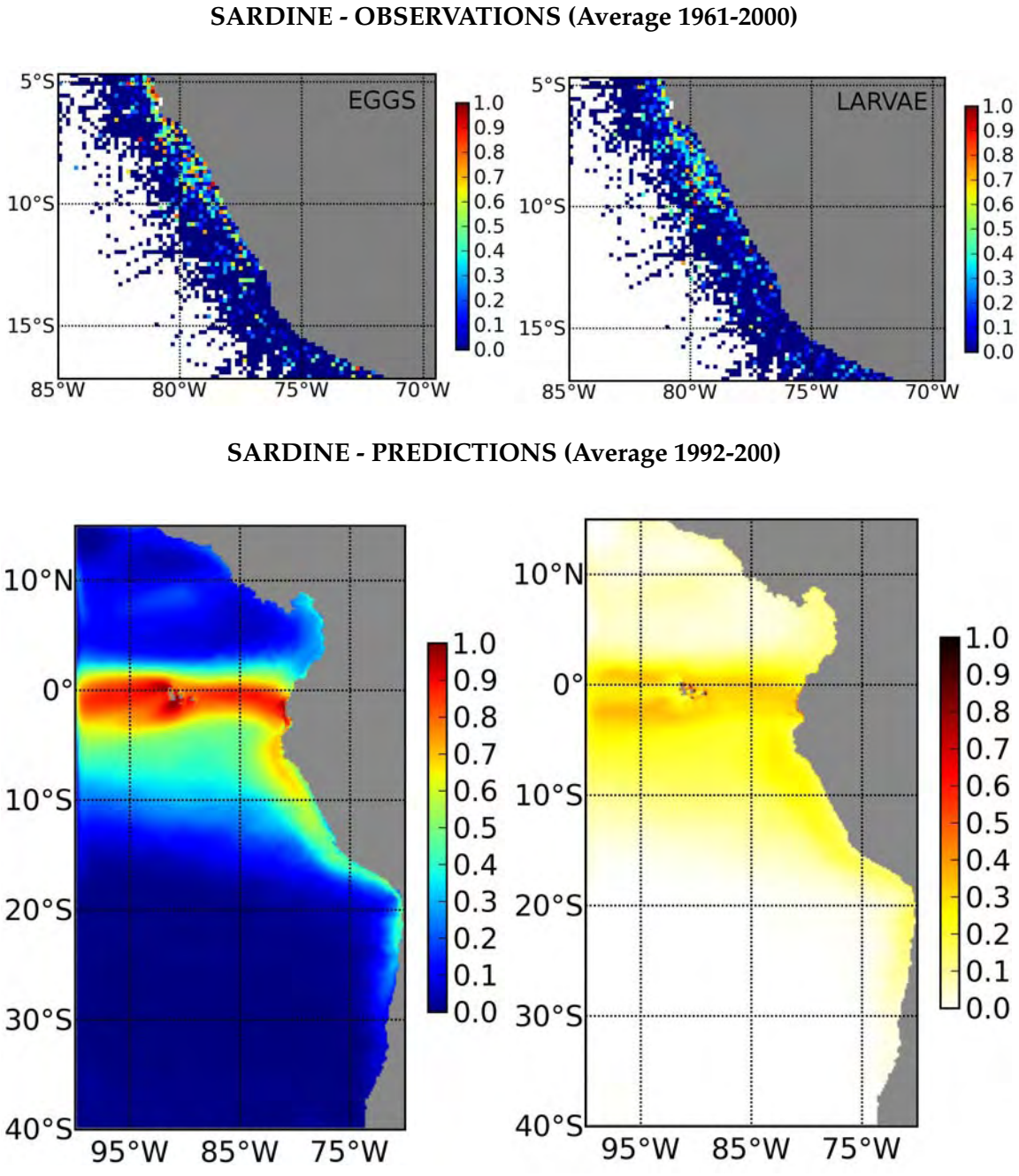

Figure 7.29: Sardine observed (top) and predicted (bottom) spawning habitat index (∼ eggs abundance, left) and larvae density (right). Predicted density averaged from 1992 to 2000 is normalized between 0 and 1. Observations have been transformed using the log10 before normalization. Since, 94% of sardine data contains zero, only absence (blue) and presence (red) are presented here to be more representative of the spatial pattern. Annual average of observed data is made with all available data from 1961 to 2000. Due to lower sampling offshore, data smoothing used for producing the average maps may introduce some biases in these regions, thus requiring cautious interpretation of these values.

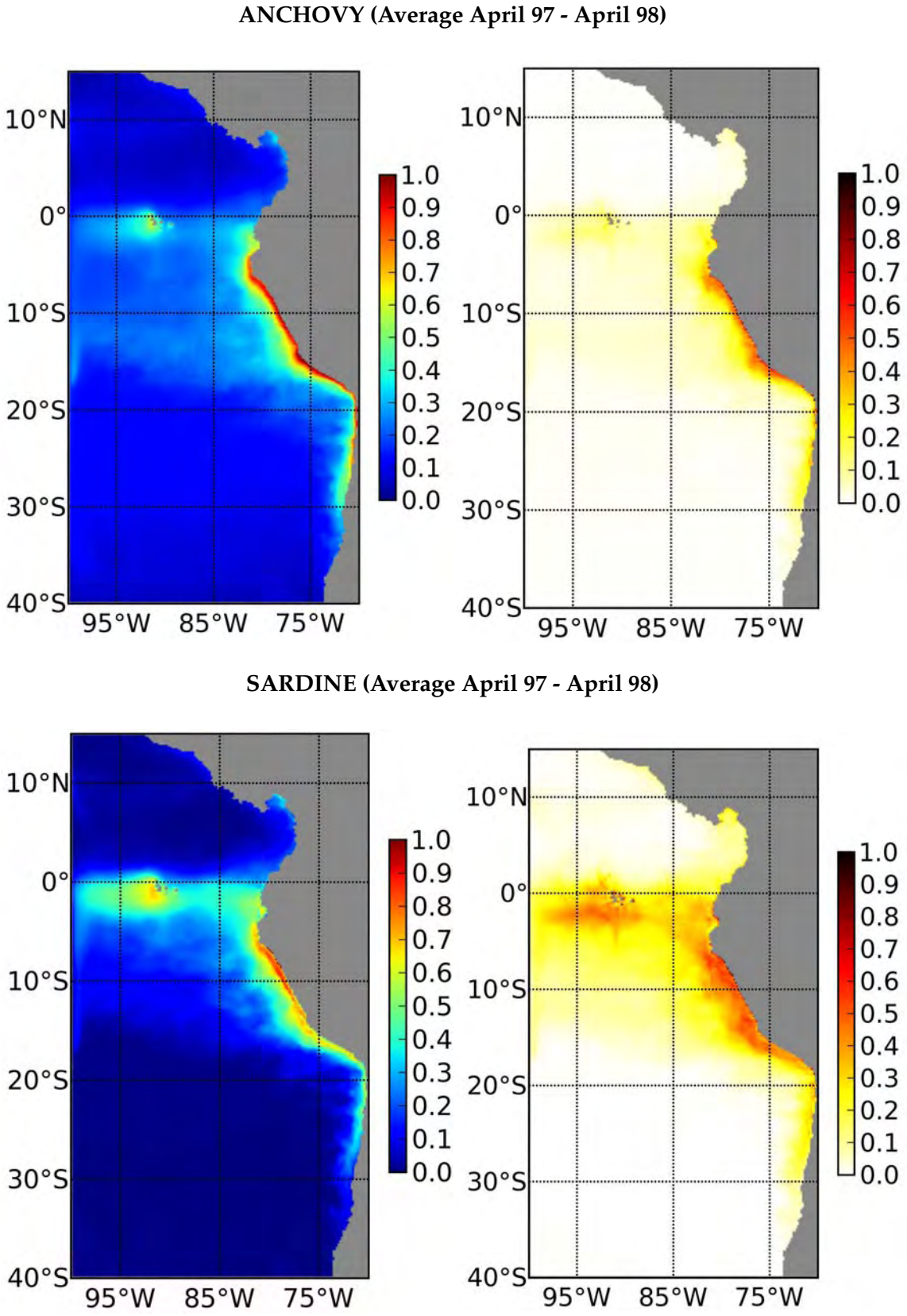

Figure 7.30: Anchovy (top) and sardine (bottom) predicted spawning habitat index (left) and larvae density habitat averaged from April 97 to April 98, during El Niño event. Data is normalized between 0 and 1.

At large scale, the simulated temporal variability is also satisfying (Fig. 7.31). The spawning rest period observed between April and June is well predicted with low abundances in July as it was also observed in eggs and larvae datasets. Note that as described in section 3.4.1, the GSI began one month earlier than observed in eggs and larvae datasets. We note however that spawning habitat and larvae density of anchovy are predicted to be the most favorable at the end of austral spring (November to January) whereas for observations March and September month appears to have the maximum abundances (Fig. 3.19). We note also that at large scale the simulated variability is well reproduced, but when studying the seasonality only in coastal regions (NCZ, CCZ, SCZ), the spawning rest period is not well predicted and shifted in time with climatological run (Fig. 7.32).

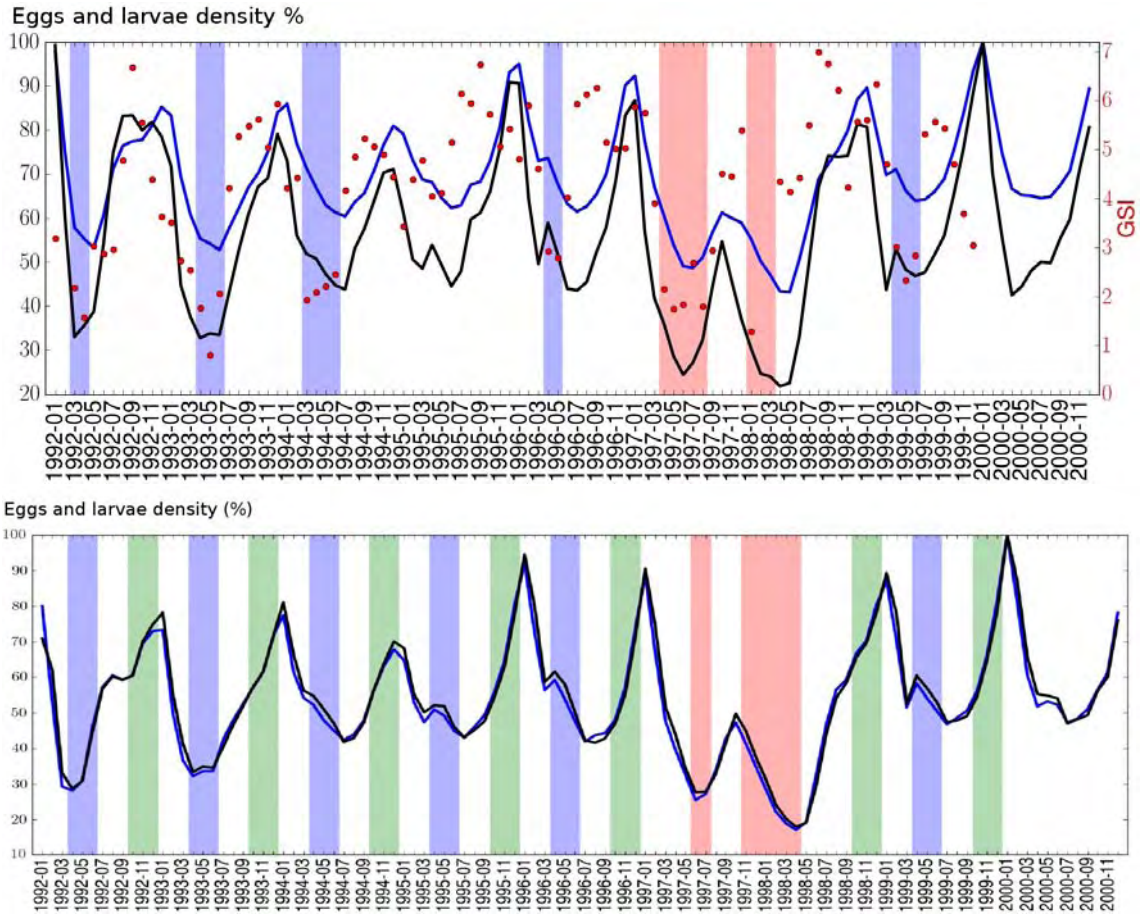

Figure 7.31: Time serie of predicted mean eggs density (black) and mean larvae density (blue) from 1992 to 2000 for anchovy (top) and sardine (bottom) over all the domain. Blue shaded bands indicate the minimal values of gonado-somatic index (Fig. 3.23) in austral autumn for anchovy. Blue and green shaded bands indicate the two minimum of gonado-somatic index (Fig. 3.24 observed for sardine, during the austral autumn and spring respectively. Red bands indicate the minimum of gonado-somatic index during the El Niño period. For anchovy and sardine, the maximum of eggs density predicted is 412 Ind/m^2 and 2.1 Ind/m^2 respectively. The maximum of larvae density predicted is 139 Ind/m^2 and 1.8 Ind/m^2.

The expected anomaly associated to the El Niño event of 1997-98, and characterized by a minimum in the gonadosomatic index (Fig. 3.23), is well predicted. Spawning habitat and larvae density show the same pattern than the one obtained with the anomaly of the GSI, which was minimum from Jul to Oct 1997, strongly increased until late 1997, then decreased to a minimum in Jan 1998 that probably lasted until March 1998 (in accordance to spawning fraction, Buitrón and Perea, 2000).

Sardine seasonality is predicted to be close of anchovy seasonality. Spawning habitat and larvae density are predicted to be more favorable in January, in agreement with observations but the two observed periods of lowest intensity in austral autumn and austral spring (Fig. 3.21) are not well predicted. We predicted that sardine did not do better than anchovy during this "El Niño" (Bertrand et al., 2004). However, there is no evidence of a strong decrease in the favorability of spawning habitat that could explain the disappearance of the species after 1999.

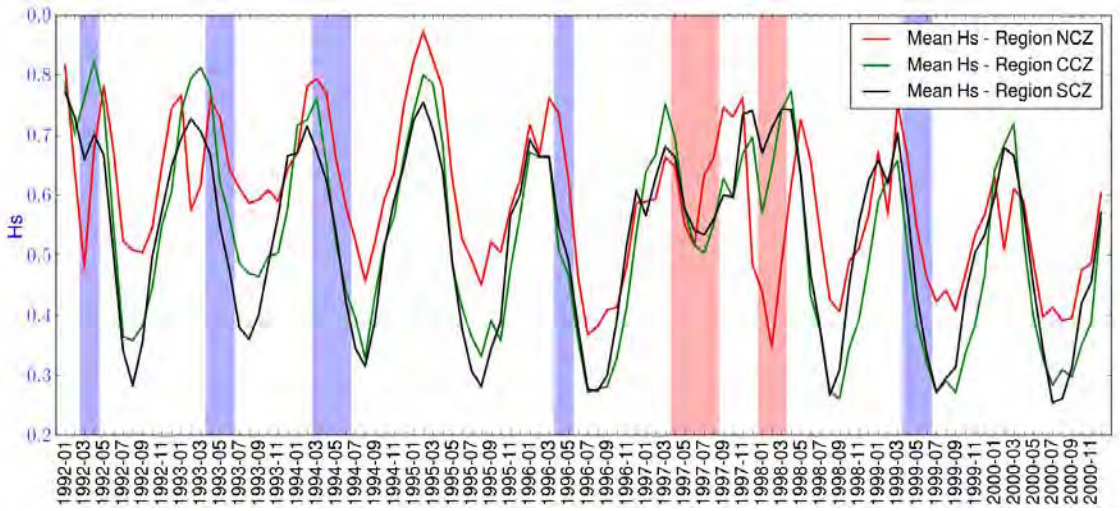

Figure 7.32: Time series of predicted mean spawning habitat index from 1992 to 2000 for anchovy average on the NCZ, CCZ and SCZ coastal areas. Blue shaded bands indicates the minimum of gonado-somatic index (Fig. 3.23 observed for anchovy during the austral autumn. Red bands indicate the minimum of gonado-somatic index during the El Niño period. During El Niño event, central and southern regions became more favorable. However, as coastal regions stays always favorable, we need to take into account the whole domain to see the impact of El Niño event in the decreasing of mean eggs and larvae density.

Parameterizations achieved either with the climatological run or the interannual run cannot predict observed seasonality in coastal areas. However, the interannual run providing contrasted environmental changes associated to El Niño events, allowed to get a parameterization producing correct interannual and seasonal variability patterns over all the model domain.
In this new parameterization, the temperature parameters were estimated with an optimum of 22.1°C (std er. 4.32) and 23.1°C (std er. 3.01) for anchovy and sardine respectively. For anchovy, the other key mechanism would be the prey abundance (i.e. primary productivity) for which the result is similar to previous experiments with climatological data, and there is no more sensitivity to the predation effect. The result is also similar for sardine about the relationship to the prey abundance but there is still a possible impact of the predation. There is no evidence of environmental changes that could explain the decadal variability and the decline of the stock at the end of 1999.

Fig. 7.33 shows the comparison between observations and predictions during the study period. Predictions (in red) are in the same order of values than observations (in blue), but as expected, predictions are not able to reproduce the high heterogeneity of observed data.

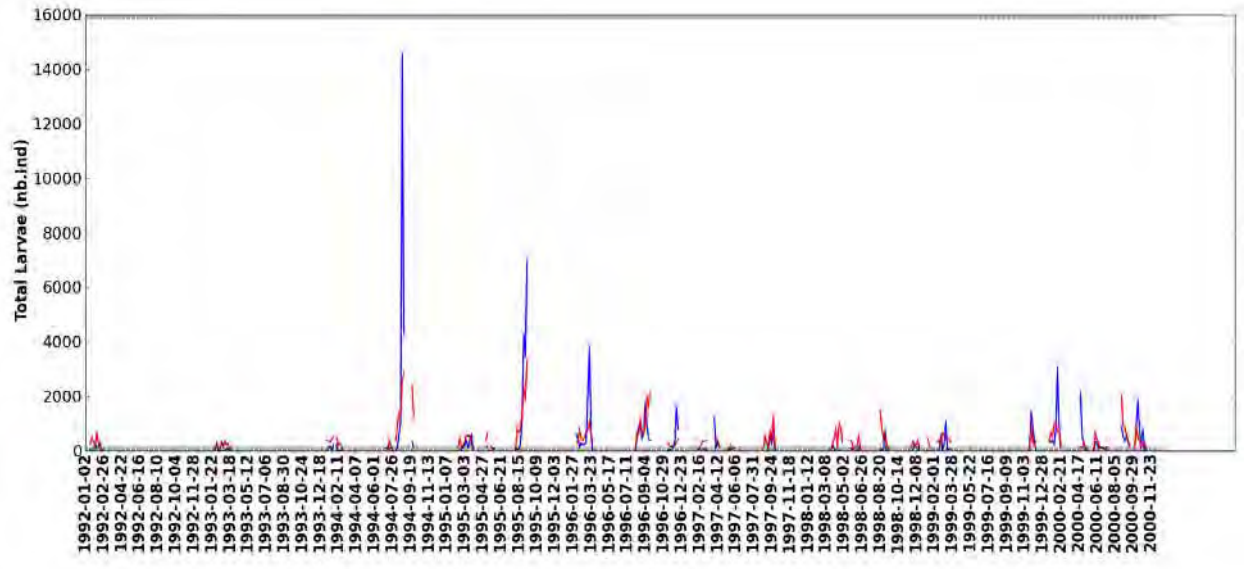

Figure 7.33: Total larvae abundances at 5 days model resolution of the observed (blue) and predicted (red) cells. Data from the Hensen net. Note that number of observed cells varied in time (Fig. 7.26) and therefore total larvae abundances is not representative of um of maximum of abundances.

An optimization experiment using only eggs data (Table 7.14) for anchovy allow to improve spatial and temporal pattern of spawning habitat index. Spatially (Fig. 7.35), spawning habitat is predicted to be more coastal in better aggrement with observation.

This also is the case during the El Niño events. Estimated parameters (Table 7.14) differ slightly for temperature which is predicted to be 2 degrees cooler with similar standard deviation and for the prey relationship which is predicted to be more linear. Absence of predator impact remains. Temporally, predicted minimum of eggs density in austral autumn 1994 is improved (Fig. 7.34) and in coastal areas the minimum of spawning habitat during the second anomaly of GSI is also better represented.

	σ_0	T_0	a	b	c	d	R	$\beta_{e,1}$	$\beta_{l,1}$	L
A	4.32	20.6	-0.03	62.66	0.47	0.092	$2.86*10^6$	0.055	0.12	24460

Table 7.14: Results of optimization experiments using only eggs data from Hensen and Calvet net from 1992 to 2000 for anchovy species.

Finally, to improve anchovy and sardine spatial dynamics, others mechanisms are likely missing, for instance, the role of the oxygen, the presence of adults and the effect of the catch. These points will be further discussed in the conclusion chapter.

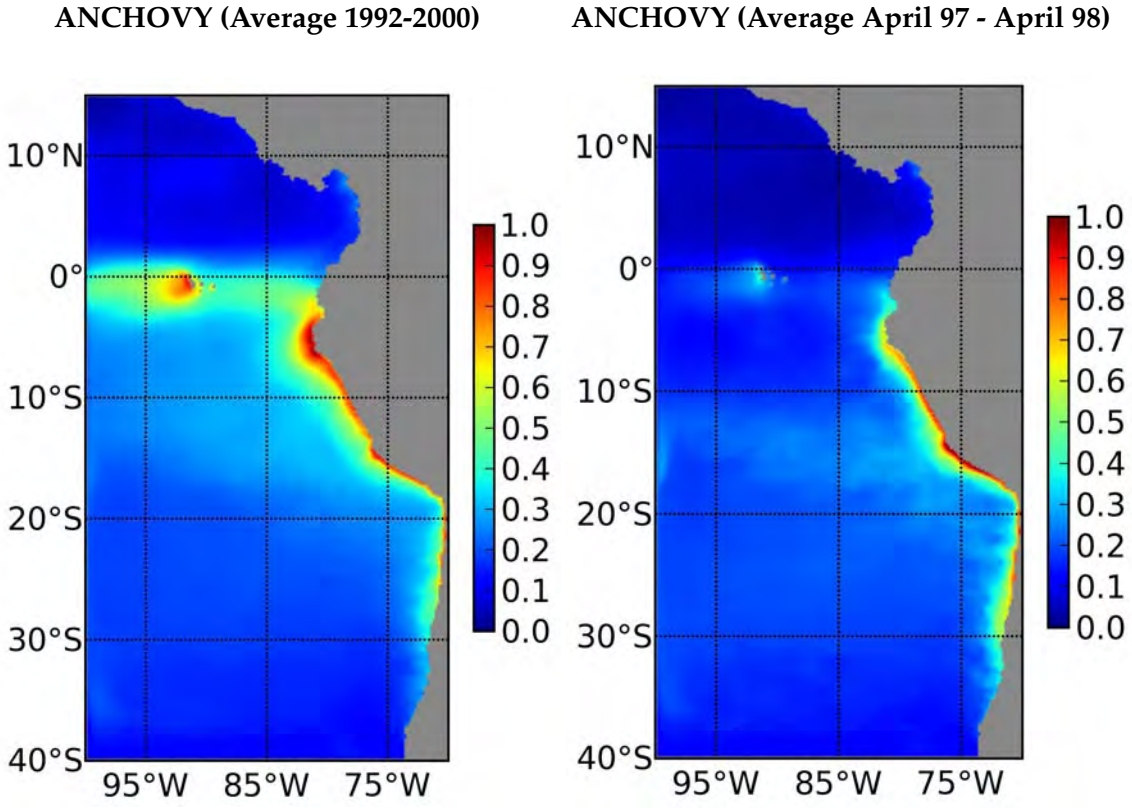

Figure 7.35: Anchovy observed (top) and predicted (bottom) spawning habitat index ($\sim$ eggs abundance, left) and larvae density (right). At left, predicted density averaged from 1992 to 2000 is normalized between 0 and 1. At right, predicted density averaged from April 1997 to April 1998.

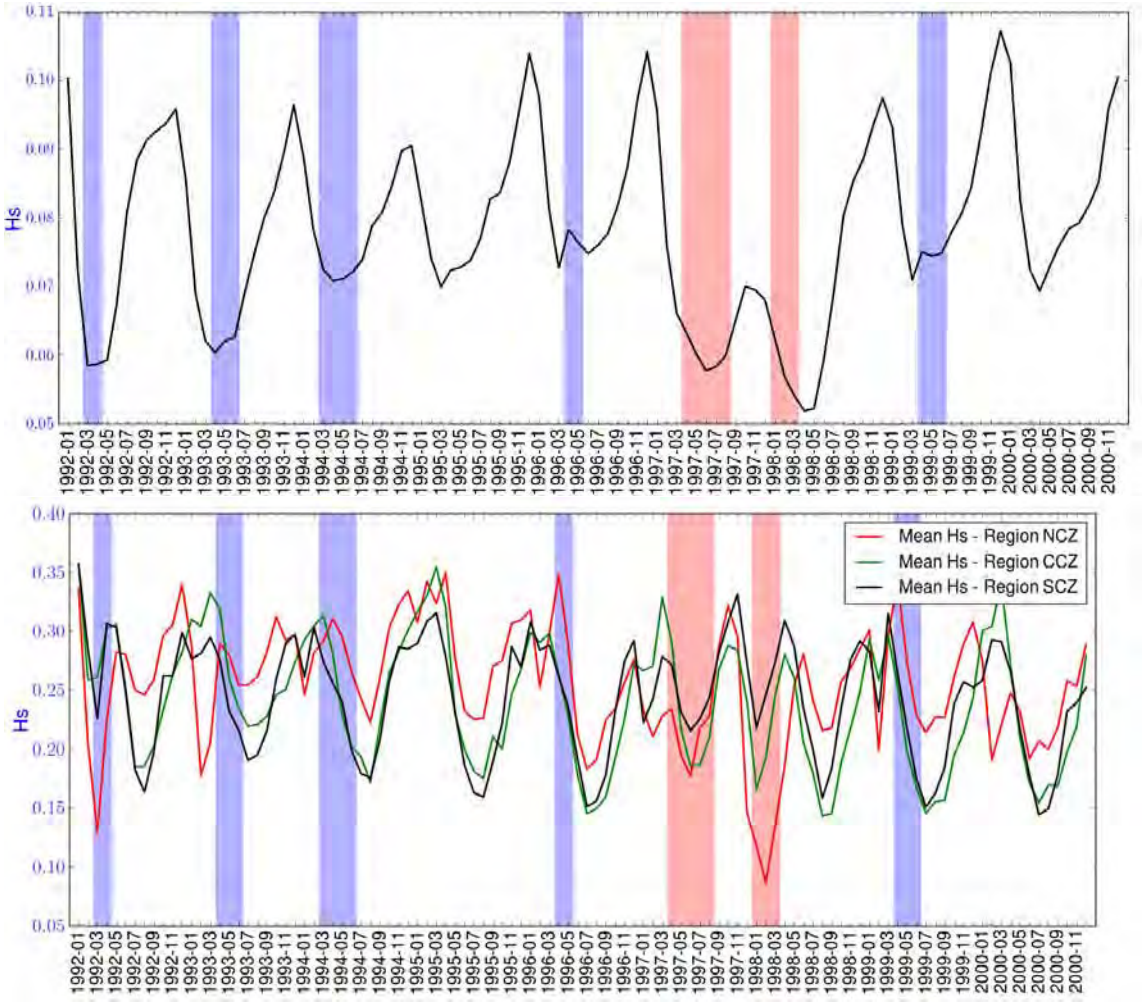

Figure 7.34: Time series of predicted spawning habitat index from 1992 to 2000 for anchovy averaged over all the domain (top) and for coastal regions (bottom). Blue shaded bands indicates the minimum of gonado-somatic index (Fig. 3.23 observed for anchovy during the austral autumn. Red bands indicate the minimum of gonado-somatic index during the El Niño period.

7.4.2 Optimization with the run 2000-2006 at 1/9° of resolution

Optimization experiment has been run from 2000 to 2006 using the available data from this period (Fig. 7.36). For anchovy, a total of 9777 observations of eggs and 9776 observations for larvae (from 2000 to 2006) have been used for the optimisation process. This run was only made for the anchovy case, since no sardine is observed since 1999.

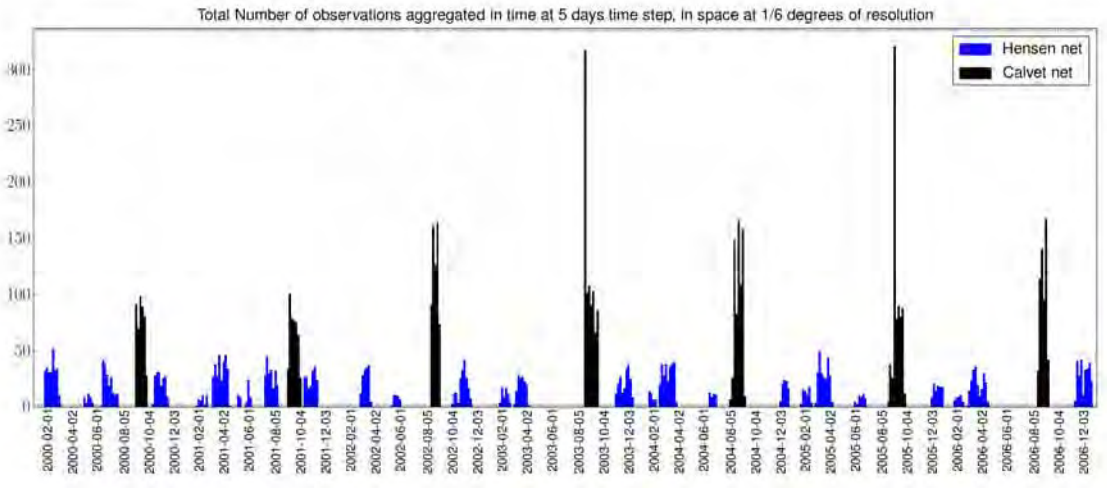

Figure 7.36: Total number of observations of eggs and larvae from 2000 to 2006 by net aggregated at the time and space model resolution.

The table 7.15 shows the results of estimated parameters using eggs and larvae dataset for anchovy from 2000 to 2006. Unlike with the run 1992-2000, temperature parameters cannot be estimated correctly. Parameter estimates for the prey and predators function also strongly differ from those obtained with the others physical forcings. A more linear relationship was predicted for the prey function (Fig. 7.4.1) and there is a low sensitivity to predator fields.

	σ_0	T_0	a	b	c	d	ϵ	μ	R	$q_e(1)$	$q_l(1)$	$\beta_e(1)$	$\beta_l(1)$	$\beta_e(2)$	$\beta_l(2)$	L
	11.67	24	-0.024	78.63	2.22	0.0001	0.0001	1.64	$3.32{*}10^6$	0.55	0.73	0.07	0.16	0.14	0.22	46370.2
std	0.074	0.361	0.342	0.409	0.026	0.184	0.036	5.959	0.742	0.06	0.073	0	0.001	0.001	0.001	

Table 7.15: Results of optimization experiments using eggs and larvae data from Hensen and Calvet net from 2000 to 2006 for anchovy.

This parameterization produced results in disagreement with observed spatial and temporal patterns. If spawning habitat is predicted to be more favorable in coastal areas and in the northern region (NCZ) as in previous optimization experiments, it is much too favorable offshore, leading in too high density of larvae offshore (Fig. 7.37). Temporally, the spawning rest period observed between April and June (Fig. 7.37) is not predicted. Furthermore, the model simulates low densities of eggs and larvae in August and September in disagreement with observations.

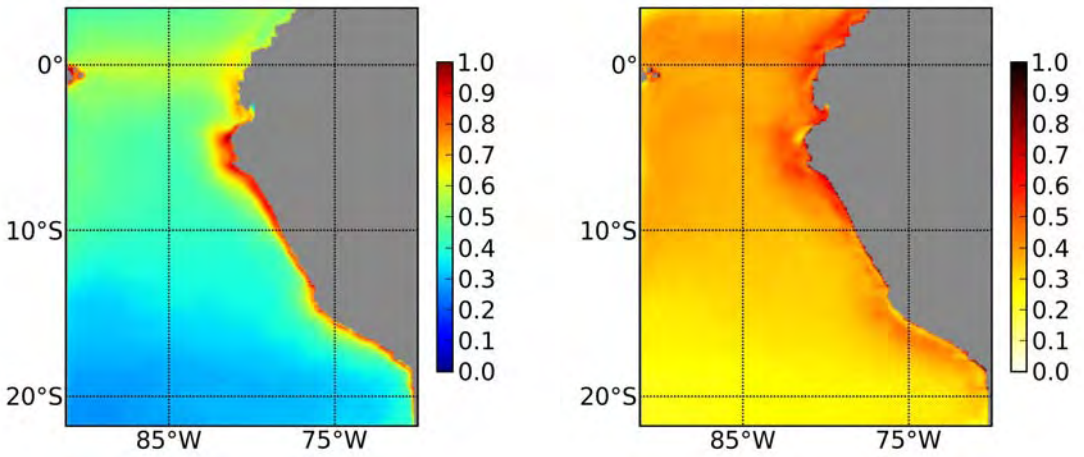

Figure 7.37: Anchovy predicted spawning habitat index (left) and larvae density (right) averaged from 2000 to 2006 and normalized between 0 and 1.

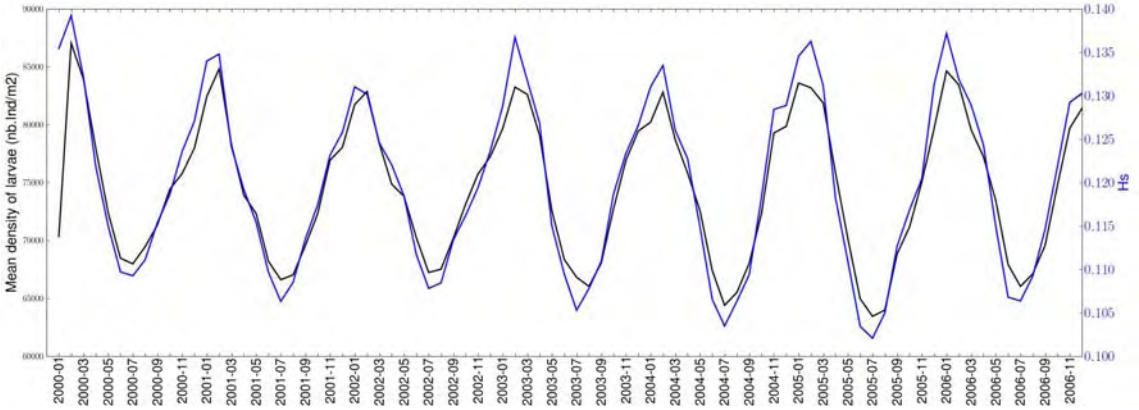

Figure 7.38: Time serie of predicted larvae abundance and spawning habitat index from 2000 to 2006 for anchovy.

The main difference between this run and the previous one is the lack of contrasted environmental conditions, since no major El Niño event occurred during the period 2000-2006. Clearly, long interannual environmental forcings including major El Niño events bring key information that help in the optimization, since the impact of extreme conditions on the fish population are strong and obvious. Despite the better resolution of the 2000-2006 run, optimization results are not improved and the use of long-term reanalysis has more impact. This is an important point to consider for further optimization experiments.

Bibliography

Bertrand, A., M. Segura, M. Gutiérrez, and L. Vasquez, 2004: From small-scale habitat loopholes to decadal cycles: a habitat-based hypothesis explaining fluctuation in pelagic fish populations off Peru. *Fish and Fisheries*, **5**(4), 296–316.

Buitrón, B. and A. Perea, 2000: Aspectos reproductivos de la anchoveta peruana durante el periodo 1992-2000. *Bol. Inst. Mar Perú*, **19**, 45–54.

Chalen, 2000: Estimación de la biomasa de peces pelágicos pequeños entre 1994 y 1999 a través de cruceros de prospección hidroacustica. In: Características oceanográficas y pesqueras en el Ecuador durante 1994-1999. . *Boletín Especial Instituto nacional de Pesca, Ecuador* 133–157.

Espinoza, P. and A. Bertrand, 2008: Revisiting peruvian anchovy (*Engraulis ringens*) trophodynamics provides a new vision of the humboldt current system. *Progress in Oceanography*, **79**(2-4), 215–227.

Espinoza, P., A. Bertrand, C. D. van der Lingen, S. Garrido, and B. R. de Mendiola, 2009: Diet of sardine (*Sardinops sagax*) in the northern humboldt current system and comparison with the diets of clupeoids in this and other eastern boundary upwelling systems. *Progress In Oceanography*, **83**, 242–250.

Grove, J. and R. Lavenberg, 1997: *A list of the fishes of the Galapagos Islands. In The fishes of the Galapagos Islands*. Stanford University Press., 1-17.

Okubo, A., 1971: Oceanic diffusion diagrams. *Deep-Sea Res.*, **18**, 789–802.

Schwartzlose, R., J.Alheit, A. Bakun, T. Baumgartner, R.Cloete, and R. Crawford, 1999: Worldwide large scale fluctuations of sardine and anchovy populations. *S. Afr. J. Mar. Sci*, **21**, 289–347.

Chapter 8

Conclusion

Sommaire

8.1 Conclusion (English version) 238
8.1.1 Synthesis of the study 238
8.1.2 Main results 239
8.1.3 Limitations of the study 240
8.1.4 Perspectives of the study 241
8.1.4.1 Improvements of the optimization method 242
8.1.4.2 Taking into account adults and catch 242
8.1.4.3 Long term perspectives 243
8.2 Conclusion Générale (French version) 245
8.2.1 Synthèse de l'étude 245
8.2.2 Résultats majeurs 246
8.2.3 Limitations de l'étude 247
8.2.4 Perspectives de l'étude 249
8.2.4.1 Perspectives d'amélioration de la méthode d'optimisation . 249
8.2.4.2 Prise en compte des adultes et de la pêche 250
8.2.4.3 Perspectives à long terme 251
Bibliography 252

8.1 Conclusion (English version)

8.1.1 Synthesis of the study

Throughout this work, we focused on anchovy and sardine population dynamics in the upwelling Humboldt current system (HCS) in Peru. These two species have a central role both in the HCS ecosystem, and the peruvian fishing industry. In order to better understand the variability of their stocks, and to develop tools permitting a sustainable management of these resources, we contributed to the development of a modelisation approach of the spatial and temporal dynamics of the early life stages of these species.

The originality of this work lies in three main points :

- We described the spatio-temporal dynamics of anchovy and sardine (eggs, larvae and adults distribution) at the seasonal, interanual and decadal scale, as well as the main characteristics of their lifestyles (habitat, larval stock-recruitment relationships).
- We implemented a data assimilation framework in order to constrain the parameters used to define the spawning habitat and larvae dynamics, from a unique dataset of eggs and larvae densities sampled over the last 40 years by IMARPE.
- In order to simulate the dynamics of the early life stages of anchovy and sardine in the HCS, we adapted the eulerian model SEAPODYM, initially developed for big pelagic fishes such as tuna, at the basin scale. Lagrangian approaches such as Indivudual Based Models (IBM) have been widely used to study the small pelagics dynamics (e.g Lett et al., 2007; Brochier et al., 2008), but this is the first time, to our knowledge that an eulerian approach is developed for such a study.

The first thesis objective was to adapt the SEAPODYM model to small pelagic fishes in a coastal domain. Spawning mechanisms et larval recruitment are crucial for the whole population dynamics, so we focused on the first life stages. We managed to adapt SEAPODYM to the specificities of these early life stages, and to implement a new data assimilation method.

The second broad objective was to explore to which extent environmental variability and the own dynamics of each species is responsible for the successive periods of high abundances of anchovy and sardine documented in the HCS. Our first results are promising, but need to be confirmed. In particular, optimization with the longest interanual series including the strong 1997-1998 Niño event permitted to restrain the uncertainty on the temperature parameter, and to reproduce satisfyingly the spatial distribution of eggs and larvae during this event.

8.1.2 Main results

The main results of this thesis are as follows :

- For the first time, we propose a detailed analysis of eggs, larvae, and adults densities in the HCS, which allows to realize a precise description of their habitat. Anchovy eggs, larvae and adults are indeed localized all along the peruvian coast between 6°S and 14°S, larvae being mainly present in the northern part of the HCS, between 6°S et 9°S. Sardine eggs and larvae are also localized along the coast, but adults are more present offshore, at the limit with the continental shelf. Thus, anchovy eggs and larvae seem to share globally the same habitat as adults, while in the case of sardine, spawning habitat and feeding habitat seem to be different. A seasonal spawning rest is observed during the austral autumn for anchovy, while for sardine another rest appears during spring.

- At the seasonal scale, we show that temperature, which was often considered as an crucial factor for larval recruitment, is not sufficient to explain, alone, the spatial and seasonal variations of small pelagics dynamics in the HCS. Besides, the match-mismatch mechanism, as defined by (Cushing, 1975, 1990), does not apply to HCS, because the larvae maximum abundance, which occurs in september, does not coincidate with the maximum primary production (March-April). However, spatial and temporal abundance of preys remains a crucial factor to explain the spawning habitat and larvae survival. Presence or absence of larvae predators seem to be another important mechanism, probably more for anchovy than for sardine, but the different optimization experiments gave sensibly varying results, and this relationship will have to be confirmed, while validating, as far as possible, the micronekton prediction in this region.
  The major and recurrent problem in all experiments based on the climatological run is the impossibility to correctly predict the coastal seasonal cycle, and particularly the biological rest in April-May.
  At the interanual scale, based on the 1992-2000 period, we show however that temperature plays an important role, because the strong environmental variability allows the model to better estimate the thermal characteristics of the spawning habitat. This interanual simulation permits to correctly reproduce the main spatial patterns of anchovy and sardine eggs and larvae dynamics during and outside El Niño periods. At the scale of the whole studied region, the model correctly predicts a drop in eggs and larvae density observed during El Niño events. But we were not able to propose an explanation concerning the disappearance of sardines since 1999.
  Finally, these results suggest that environmental variability permits to explain the main spatial patterns of anchovy and sardine eggs and larvae dynamics, the interanual variability, and to a certain extent the seasonality observed at the whole

region scale, but not at the coast. Different factors may explain the difficulty to represent the seasonal cycle at the coast, but we may consider first the influence of dissolved oxygen, of adult fishes dynamics, and of fishing mortality which is extremely high and limited to short periods.

- It has recently been proposed that the oxygen dissolved concentration could be an important parameter controlling interanual variability, anchovy being more tolerant to anoxic conditions (in terms of oxygen concentration, and height of the homogeneously oxygenated layer) than sardine (Bertrand et al., 2010, 2011). At the seasonal scale, our results did not demonstrate the oxygen influence on anchovy and sardine. It is too early to conclude on the global influence of oxygen due to the uncertainties on the modeled oxygen concentration in the ROMS-PISCES input. This question deserves further studies.

- We tested the strength of our modelisation approach, and of the parameters retrieved by mean of the data assimilation method, by studying the sensibility of our model to physical forcing, forcing of the micronecton model, data resolution, and finally to uncertainty and errors on eggs and larvae data. It appears that the cost function selection significantly impacts predictions. Besides, the way larvae predators, and primary productivity (a proxy for larvae food) are simulated, can considerably modify spatial and temporal dynamics.

8.1.3 Limitations of the study

Every model naturally contains approximations and simplifications. In ecosystemic modelisation, uncertainties are even bigger, since ecology in general, and fish behaviour more specifically, can hardly be translated into mathematical equations, and of course because the uncertainties of all input data and forcing sum up with the own uncertainties of the ecosystemic model. It is is therefore necessary to know the limitations and uncertainties of our model to correctly interpret its results.

The main limitations of our model are described below :

- SEAPODYM is limited by a lack of vertical resolution. Indeed, we average currents over the vertical habitat of anchovy eggs and larvae. However, in an upwelling system, the currents dynamics is very heterogeneous in the first 50 meters, and this averaging may be excessive. Recent studies (Brochier et al., 2008) have shown the importance of eggs flotability and larvae vertical micration on dispersing patterns. Our study of the IMARPE dataset confirms these results. It shows that without taking into account vertical resolution and larval migration below Ekman depth (Chapter 3), we can hardly explain how larvae can remain in a favorable environment. The vertical movements of larvae could in part be compensated in

SEAPODYM by considering a deeper vertical layer (which remains to be defined). This would permit to average the opposed surface and subsurface currents, in order to eventually predict correctly the resulting overall larvae transport.

- Anchovies are encountered in a narrow band of 50 to 100km from the coast (Fig. 3.14). The coastal processes simulated at the 1/6° to 1/9° resolution are probably not yet well resolved, and could explain the retention of eggs and larvae in some bays along the seashore. Such processes could have an important role during El Niño events, when anchovy and sardine are concentrated close to the coast. A better spatial resolution with finer resolution nearshore could probably help to reproduce the small pelagics dynamics. Resolutions of the order of 1/24° to 1/36° would be necessary for this purpose.

- Our predictions depend on the model used to simulate larvae predators, i.e the micronekton model. We only consider mesopelagic and bathypelagic migrants in the micronekton, so we do not take into account all larvae and eggs predators nearshore, at depths lower than 100m. This micronekton model was validated at the basin scale (Lehodey et al., 2010), but its predictions were not validated in the HCS because of a lack of observations (Chapter 4). It is therefore necessary to improve this model in order to improve the accuracy of our predictions.

- Data assimilation is done from a dataset whose uncertainties have been exposed in Chapter 3. It is possible to choose a cost function that takes into account to a certain extent these uncertainties, as well as data overdispersion, but we do not know how to take into account the results of significativity tests (Mann-Withney test, cf section 3.4.1.

- Another limit is the absence of adults in the model. We do not take into account the adults habitat, their feeding preferences, or female fecondity, which may also impact the seasonality of spawning. This point will be more specifically discussed in the perspectives section.

- Last, we shall recognize the difficulty for current biogeochemical models to realistically simulate dissolved oxygen concentration. As long as this essential variable for the marine species life cycle is not better described, it will be difficult to test and take into account the influence of this parameter.

8.1.4 Perspectives of the study

Each of the limitations exposed above calls for improvements perspectives. At short run however, I propose two major improvements axis. The first one deals with data assimilation methods and corresponding validation tools. The second one concerns the complete version of the SEAPODYM code (from eggs to adults), which would take into account the spatio-temporal distribution of adults and the consequences of fishing.

8.1.4.1 Improvements of the optimization method

We defined our habitat index Hs as a function of temperature, food presence, and predators absence:

$$H_S = f_1(T) * f_2(PP) * f_3(Pred) \tag{8.1}$$

Each of the f_1, f_2 and f_3 functions is comprised between 0 and 1, but their maximum do not necessarily coincidate in the spatio-temporal space of optimization experiments. Consequently, H_S maximum is not necesarily equal to 1, and may vary between different optimization experiments.
Optimization experiments have shown that some parameters are strongly correlated, and particularly the recruitment parameter with the different parameters defining the spawning habitat (Table D.1). We shall recall that $d_{eggs} = R * H_s$, and because H_s value is not normalized, the optimization process may find different (R; H_s) couples explaining a given observed eggs density.
By normalising H_s between 0 and 1, we would reduce the correlations existing between different optimized parameters, which would ease the obtention of an optimal solution. The function permitting this normalisation should be differentiable, and it remains to be found.

In a second step, we propose to perform the optimization procedure in a gaussian framework.
I showed that our assimilation method of eggs and larvae density data by the adjoint method is technically able to retrieve the parameters of our model, and allow to reproduce at the first order the spatial distribution and interanual variability of the dataset. It does not however explain satisfyingly the observed eggs and larvae seasonal variability. Numerous uncertainties on the dataset itself, and on the model may explain these differences between observations and predictions. Moreover, we encountered a lack of analysis tools to quantitatively evaluate the quality of our results, in particular because of the non-gaussian distribution of our data.
I propose to render our optimization problem gaussian, via an anamorphosis transformation (Simon and Bertino, 2009). This would permit to use analysis tools and techniques developed mainly in the fields of meteorology and physical oceanography (e.g Simon and Bertino, 2009). In this case, we could use a gaussian cost function instead of negative binomial, and observations as well as predictions could transformed with mathematical functions in this gaussian space.

8.1.4.2 Taking into account adults and catch

Adults dynamics is not considered in our model of the early life stages of small pelagics. The underlying hypothesis is that adults are present everywhere, and we

only intended to define the parameters controlling the spawning habitat. This of course is an important bias of our simulations, in the sense that predicted habitats only are potential habitats. In the absence of mature adults, it is obvious that there can be no reproduction. Furthermore, we highlighted in chapter 3 a relationship at the small scale (1°x 1°) between larvae density and adults density.
Spatio-temporal dynamics of adults may thus strongly modify the potential habitat index, in particular through aggregation phenomena during spawning. It seems indeed that adults of anchovy, and mainly sardine, tend to concentrate nearshore during reproduction peaks, and then disperse more offshore to feed. Numerous acoustical observations on the distribution of adult biomass are available, which should provide valuable informations to calibrate the additional parameters that are required to describe adults dynamics, following the same optimization method as developped for eggs and larvae.

Another key factor which may modify adults dynamics, and therefore eggs and larvae's, is the fishing impact. In the HCS, it can have a very brutal impact on the reproductory stock. In 2007 for instance, about 7 million tons of anchovy were captured in barely 50 days. In 4 days, the peruvian fishermans can fish up to 600 000 tons of anchovy, which corresponds roughly to the total annual catches of the french fisheries. Taking into account this impact is really crucial, and should be possible within the framework of a complete SEAPODYM model describing the dynamics of the whole populations.

8.1.4.3 Long term perspectives

Despite the problems encountered throughout this study to adapt the SEAPODYM model to small pelagics populations, the progresses realized with the optimization approach are encouraging, and let me think that building the complete version of the model exposed above is a realistic objective.
Additional data could be integrated in the Maximum Likelihood Estimation approach of the completed model. These include the acoustical data cited above, but also spatially resolved fishing data. The exact positions of fishing boats and of the fishing effort can indeed be deduced from the real time tracking system of the fleet (Bertrand et al., 2008). Not only could this information allow a better estimation of the fishing mortality, but it would bring an additional direct and precise signal for model optimisation, the capture being also estimated by the model.

There remains a lot to do in order to understand the origin of the strong variability of small pelagics populations in the Humboldt Current System, but the eulerian modelisation framework developped during this thesis work with a rigourous optimisation approach seem to be an appropriate tool to pursue this research.
In the longer term, we may consider that such tools would enable the evaluation of the

impacts of climate change on these central species for marine ecosystems, and we hope that it could be used to implement a real ecosystemic stocks and fisheries management.

8.2 Conclusion Générale (French version)

8.2.1 Synthèse de l'étude

Au cours de ce travail de thèse, nous nous sommes intéressés à la dynamique des populations d'anchois et de sardine dans le système d'upwelling du courant de Humboldt (SCH) au Pérou; deux espèces qui occupent un rôle central à la fois dans l'écosystème du SCH, et dans l'industrie de la pêche péruvienne. Dans le but de mieux comprendre la variabilité de leurs stocks, et de développer des outils permettant une gestion durable de ces ressources, nous avons contribué au développement d'une approche de modélisation permettant d'étudier la dynamique spatiale et temporelle des premiers stades de vie de ces espèces.

L'originalité de ce travail réside en trois points principaux :

- Nous avons décrit les dynamiques spatio-temporelles des anchois et des sardines (distribution des oeufs, larves et adultes) à l'échelle saisonnière, interannuelle et décennale ainsi que les principales caractéristiques de leurs traits de vie (habitat, relation stock-recrutement larvaire).
- Pour simuler la dynamique des premiers stades de vie des anchois et sardines dans le système du Courant de Humboldt, nous avons utilisé un modèle de type eulérien, adapté du modèle SEAPODYM, initialement développé pour les grands poissons pélagiques comme le thon à l'échelle du bassin océanique,. Si les approches lagrangiennes du type modèles individus-centrés (IBM) ont largement été utilisées pour étudier la dynamique des petits poissons pélagiques (e.g Lett et al., 2007; Brochier et al., 2008), c'est la première fois, à notre connaissance, qu'une approche eulérienne est implémentée pour ce type d'étude.
- Nous avons mis en oeuvre une méthode d'assimilation de données afin de contraindre les paramètres utilisés pour définir l'habitat de ponte et la dynamique larvaire, à partir d'un jeu unique de données de densité d'oeufs et de larves échantillonnés au cours des 40 dernières années par l'IMARPE.

Le premier objectif de la thèse était d'adapter le modèle SEAPODYM aux petits poissons pélagiques dans un domaine côtier. Les mécanismes de ponte et de recrutement larvaires étant déterminants pour la dynamique de population des petits pélagiques, nous nous sommes focalisés plus particulièrement sur les premiers stades de vie. Nous avons réussi à adapter SEAPODYM aux spécificités des premiers stades de vie de ces espèces et à y implémenter une nouvelle méthode d'assimilation de données.

Le second objectif était d'explorer dans quelle mesure la variabilité environnementale et la dynamique propre à chaque espèce peuvent expliquer les périodes successives de

forte abondance de l'anchois ou de la sardine observées dans le SCH. Sur ce point, il reste à confirmer des résultats non définitifs mais prometteurs. En particulier, l'optimisation avec la série interannuelle la plus longue et incluant le très fort évènement El Niño 1997-98 a permis de restreindre l'incertitude sur les paramètres de température et de reproduire de manière satisfaisante la distribution spatiale des oeufs et larves durant cet évènement.

8.2.2 Résultats majeurs

Les principaux résultats de la thèse sont les suivants:

- Nous effectuons pour la première fois une analyse détaillée de données de densités d'oeufs, de larves et d'adultes (Chapitre 3), ce qui permet de réaliser une description minutieuse de leur habitat. Les oeufs, larves et adultes d'anchois sont ainsi localisés tout le long de la côte entre 6°S et 14°S, les larves étant présentes préférentiellement dans la partie Nord du HCS entre 6°S et 9°S. Les oeufs et larves de sardines sont également localisées le long de la côte, mais les adultes sont présents plutôt plus au large, à la limite du plateau continental. Ainsi, les oeufs et larves d'anchois semblent globalement partager le même habitat que les adultes tandis que dans le cas de la sardine, habitat de ponte et habitat d'alimentation semblent être différenciés. Un repos saisonnier de la ponte est observé pendant l'automne austral pour l'anchois, tandis que pour la sardine un second repos apparaît pendant le printemps.

- A l'échelle saisonnière, nous montrons que la température, souvent considérée comme un facteur essentiel au recrutement larvaire, ne permet pas, seule, d'expliquer les variations spatiales et saisonnières de la dynamique des petits pélagiques dans le HCS. Par ailleurs, le mécanisme de "match-mismatch" tel que défini par (Cushing, 1975, 1990), ne s'applique pas au HCS puisque le maximum d'abondance des larves (Septembre) ne coïncide pas avec le maximum de production primaire (Mars-Avril). En revanche, l'abondance spatiale et temporelle des proies demeure un facteur primordial pour expliquer l'habitat de ponte et la survie larvaire. La présence ou absence des prédateurs des larves semble être aussi un mécanisme important, probablement plus pour les sardines que pour les anchois, mais les différentes expériences d'optimisation donnent des résultats sensiblement différents et il faudra confirmer cette relation tout en validant autant que possible la simulation du micronecton dans cette région.
  Le problème majeur récurrent dans toutes les expériences réalisées avec la climatologie est l'impossibilité de prédire correctement à la côte le cycle saisonnier et notamment la période de repos biologique entre avril et juin.
  A l'échelle interannuelle, pendant la période 1992-2000, nous montrons en revanche que la température joue un rôle beaucoup plus important car la forte variabilité permet au modèle d'estimer les caractéristiques de l'habitat thermal de ponte. Cette

simulation interannuelle permet de modéliser correctement les principaux schémas spatiaux de la dynamique des oeufs et larves d'anchois et sardines hors et pendant périodes El Niño. A l'échelle de la région d'étude, le modèle prédit correctement une chute de la densité des oeufs et larves observée pendant les périodes El Niño. Toutefois, aucune explication n'est fournie par ces simulations quant à la disparition des sardines depuis 1999.
Finalement, ces résultats suggèrent que la variabilité environnementale permet d'expliquer les principaux schémas spatiaux de la dynamique spatiale des oeufs et larves d'anchois et sardines, la variabilité interannuelle, et dans une certaine mesure la saisonnalité observée à l'échelle du domaine d'étude mais pas à la côte. Parmi les facteurs pouvant expliquer la difficulté de représenter le cycle saisonnier à la côte, on peut considérer en priorité l'impact de l'oxygène, la prise en compte des dynamiques des poissons adultes et de celle de la mortalité par pêche qui est extrêmement forte et concentrée dans le temps.

- Il a récemment été proposé que la concentration en oxygène dissous pourrait être un paramètre important dans l'origine de cette variabilité interannuelle, l'anchois étant plus tolérant aux conditions anoxiques (en termes de concentration en oxygène et hauteur de la couche homogène oxygénée) que la sardine (Bertrand et al., 2010, 2011). A l'échelle saisonnière, nos résultats n'ont pas montré d'impact de l'oxygène sur l'anchois et la sardine. Nos résultats ne sont toutefois pas concluants, peut être du fait que la concentration en oxygène n'est pas suffisamment bien prédite dans le modèle PISCES. Cet axe de recherche est encore à explorer.

- Nous avons testé la robustesse de notre approche de modélisation et des paramètres biologiques obtenus par assimilation de données en étudiant la sensibilité de notre modèle aux forçages physiques utilisés, à ceux du modèle de micronecton, à la résolution des données, et enfin aux erreurs sur les données d'oeufs et de larves. Il apparaît que le choix de la fonction coût a un impact significatif sur les prédictions. De même, la manière dont le champ de prédateurs des larves, et la productivité primaire (proxy de la nourriture des larves), sont modélisés peut modifier considérablement la dynamique temporelle et spatiale.

8.2.3 Limitations de l'étude

Tout modèle contient naturellement des approximations et simplifications. Dans la modélisation écosystémique, les incertitudes sont encore plus grandes, puisque la biologie et plus particulièrement le comportement des poissons est plus difficilement modélisable par de simples équations que ne peuvent l'être les processus physiques, et bien entendu, parce qu'aux modèles écosystémiques s'ajoutent les incertitudes de tous les forçages et données d'entrées. Il est ainsi indispensable de connaitre les limites et possibles incertitudes de notre modèle pour pouvoir interpréter correctement ses

résultats.

Les principales limitations de notre modèle sont les suivantes:

- SEAPODYM est limité par un manque de résolution verticale. Nous moyennons en effet les courants sur l'habitat vertical des oeufs et larves d'anchois. Cependant, dans un système d'upwelling, la dynamique des courants est très hétérogène dans les 50 premiers mètres, et cette approximation peut se révéler excessive. Certaines études (Brochier et al., 2008) ont montré l'importance de la flottabilité des oeufs et de la migration verticale des larves sur les schémas de dispersions. L'étude des données corrobore ces résultats. Elle montre que sans prise en compte de la résolution verticale et de la migration des larves sous la profondeur d'Ekman (Chapitre 3), nous pouvons difficilement expliquer comment les larves peuvent rester dans un environnement favorable. Ce mécanisme de retour des larves en subsurface pourrait, peut être, être compensé en partie dans SEAPODYM par la prise en compte d'une couche verticale plus profonde (qui reste à définir). Cela aurait pour effet de moyenner les courants opposés de surface et subsurface et donc au final de prédire correctement le transport final des larves qui en résulte.

- La concentration d'anchois s'observe dans une bande de 50 à 100km à la côte (Fig. 3.14). Les processus côtiers à l'échelle d'un modèle au 1/6 ème ($\sim$ 18 km) ou 1/9 ème ($\sim$ 12.3 km) ne sont probablement pas encore bien résolus et pourraient expliquer la rétention des oeufs et larves dans certaines baies le long du littoral. Ces processus pourraient avoir un rôle important pendant les années "El Niño" où anchois et sardines se retrouvent « collés » à la côte. Une meilleure résolution spatiale avec des échelles plus fines près des côtes pourraient peut être permettre de reproduire plus correctement la dynamique de ces petits pélagiques. Il faudra très probablement atteindre des résolutions de l'ordre du 1/24 ème ou 1/36 ème pour que ces processus puissent être pris en compte.

- Nos prédictions sont dépendantes du modèle utilisé pour simuler les prédateurs des larves, i.e. le modèle micronecton. En ne prenant en compte que les migrants mésopélagiques et bathypélagiques, nous ne prenons pas en compte la globalité des prédateurs des oeufs et larves d'anchois et sardines à la côte, dans des fonds inférieurs à 100m. Ce modèle micronecton a fait ses preuves à l'échelle du basin (Lehodey et al., 2010), mais le réalisme de ses prédictions dans la zone d'upwelling de Humboldt reste non validé par manque d'observations (Chapitre 4). Ainsi, il est nécessaire d'améliorer ce modèle afin d'améliorer nos prédictions.

- L'assimilation de données est effectuée à partir d'un jeu de données brutes dont nous avons mis en évidence les incertitudes dans le chapitre 3. Il est possible de

choisir une fonction coût qui prenne en compte dans une certaine mesure ces incertitudes ainsi que la surdispersion des données, mais nous ne savons pas comment prendre en compte les résultats des tests de significativité (test de Mann-Withney, cf section 3.4.1).

- Une autre limite est l'absence d'adultes dans le modèle. En effet nous ne prenons pas en compte l'habitat de présence des adultes, leur alimentation, la fécondité des femelles, facteurs qui peuvent aussi affecter la saisonnalité de la ponte. Ce point sera discuté avec plus de détails dans les perspectives.

- En dernier point, nous noterons la difficulté des modèles biogéochimiques à modéliser de façon réaliste la concentration en oxygène dissous. Tant que cette variable essentielle au cycle de vie des espèces biologiques ne sera pas mieux décrite, il sera difficile de tester et prendre en compte ce mécanisme de manière optimale.

8.2.4 Perspectives de l'étude

Chacune des limites exposées précédemment appelle des perspectives d'améliorations. A court terme cependant, je propose deux axes majeurs d'amélioration. Le premier porte sur les méthodes d'assimilation de données et les outils de validation correspondants. Le second porte sur la version complète du modèle SEAPODYM (des oeufs aux adultes), qui prend en compte la distribution spatio-temporelle des adultes et l'effet de la pêche.

8.2.4.1 Perspectives d'amélioration de la méthode d'optimisation

Nous avons défini notre indice d'habitat H_s comme une fonction de la température, de la présence de nourriture et de l'absence de prédateurs:

$$H_s = f_1(T) * f_2(PP) * f_3(Pred) \tag{8.2}$$

Chacune des fonctions f_1, f_2 et f_3 est comprise entre 0 et 1. Le maximum de ces fonctions ne coïncidant pas nécessairement d'un point de vue spatio-temporel, le maximum de H_s n'est pas nécessairement égal à 1, et peut varier selon les expériences d'optimization.
Les expériences d'optimisation ont montré que certains paramètres sont fortement corrélés (cf Table D.1). Plus particulièrement, on trouve que le paramètre de recrutement est ainsi corrélé avec différents paramètres définissant l'habitat de ponte. En effet, $d_{oeufs} = R * H_s$, et par conséquent, la valeur de H_s n'étant pas normalisée, le processus d'optimisation peut aboutir à différents couples R ; H_s permettant d'expliquer une observation de densité d'oeufs.
Je propose de normaliser l'indice d'habitat Hs afin qu'il soit effectivement toujours compris entre 0 et 1. Il faudrait donc trouver une fonction pour normaliser cet indice, qui soit également différentiable. Cette approche permettrait de réduire les corrélations

existant entre les paramètres optimisés, et par conséquent faciliterait la recherche d'une solution optimale.

Dans un second temps, une autre approche d'optimisation en se plaçant dans un cadre gaussien pourrait être envisagée. J'ai montré que notre méthode d'assimilation de données d'oeufs et larves par méthode adjointe est techniquement capable d'obtenir les paramètres de notre modèle, et permet de reproduire au premier ordre la distribution spatiale des données et la variabilité interannuelle. Elle ne permet cependant pas d'expliquer de façon entièrement satisfaisante la variabilité saisonnière de la dynamique des oeufs et larves observée. De nombreuses incertitudes sur les données elles-mêmes, ainsi que sur le modèle, peuvent expliquer ces différences entre observations et prédictions. De plus, nous nous sommes heurtés au cours de cette thèse au manque d'outils d'analyses pour évaluer quantitativement la qualité de nos résultats, notamment en raison de la distribution non gaussienne des données.
Je propose, via une transformation d'anamorphose (transformation d'image, Simon and Bertino, 2009), de rendre notre problème d'optimisation gaussien, ce qui nous permettrait d'utiliser les outils d'analyses et techniques développées principalement dans les domaines de la météorologie et de l'océanographie physique (e.g Simon and Bertino, 2009). Dans ce cas, on n'utiliserait plus une fonction cout négative binomiale mais gaussienne, et observations et prédictions seraient transformées par des fonctions mathématiques dans cet espace gaussien.

8.2.4.2 Prise en compte des adultes et de la pêche

La dynamique des adultes n'est pas prise en compte dans notre modèle des premiers stades de vie. L'hypothèse sous-jacente est que les adultes sont présents partout, et nous avons essayé simplement de définir l'habitat de ponte favorable. C'est évidemment un biais important dans nos simulations et les habitats prédits sont des habitats potentiels. Il est bien évident qu'en l'absence totale d'adultes matures, il ne peut y avoir de reproduction. De plus, nous avons mis en évidence au chapitre 3 qu'il existe une relation à une échelle spatiale fine (1°x 1°) entre les quantités de larves et d'adultes.
Les dynamiques spatiales et temporelles des adultes sont donc susceptibles de modifier fortement l'indice d'habitat potentiel, notamment par des phénomènes d'agrégation de ponte. Il semble en effet que les adultes d'anchois et surtout de sardine se concentrent à la côte au moment des pics de reproductions puis se dispersent plus au large pour s'alimenter. Les nombreuses observations acoustiques disponibles sur la distribution des biomasses adultes devraient fournir une information précieuse pour calibrer les paramètres supplémentaires pour décrire ces dynamiques, en suivant la même approche d'optimisation que celle développée pour les oeufs et les larves.

L'autre facteur clé qui peut modifier les dynamiques des adultes, et donc celle des oeufs et larves, est bien sûr la pêche. Elle peut avoir un impact sur le stock reproducteur extrêmement brutal. En effet rappelons par exemple qu'en 2007, environ 7 millions de tonnes d'anchois furent capturés en à peine 50 jours. En 4 jours, les pêcheurs péruviens peuvent capturer jusqu'à 600 000 tonnes d'anchois, ce qui équivaut environ aux captures annuelles de pêche française. La prise en compte de cet impact est absolument primordiale et pourra être analysée une fois la version finale du modèle décrivant la dynamique complète de la population mise au point.

8.2.4.3 Perspectives à long terme

En dépit des problèmes rencontrés au cours de cette étude pour adapter le modèle SEAPODYM aux populations de petits pélagiques, les progrès réalisés avec l'approche d'optimisation et les premiers résultats encourageants obtenus permettent de penser que la version complète du modèle évoquée plus haut est un objectif réalisable. D'autres types de données pourraient être avantageusement intégrées dans l'approche d'Estimation du Maximum de Vraisemblance de la version complète du modèle. Il s'agit des données acoustiques déjà citées plus haut, mais aussi des données de pêches. Les positions exactes de l'effort de pêche et des captures peuvent notamment être déduites du système de suivi des flottilles en temps réel (Bertrand et al., 2008). Non seulement cette information permettrait une meilleure prise en compte de la mortalité par pêche, mais elle apporterait un signal supplémentaire direct et précis dans l'optimisation du modèle (la capture étant estimée par le modèle).

Il reste donc encore beaucoup à faire pour comprendre l'origine de la forte variabilité des populations de petits pélagiques dans le système de Humboldt, mais le cadre de modélisation eulérienne développé ici avec son approche rigoureuse d'optimisation me semble un outil approprié pour poursuivre cette recherche. A plus long terme, on peut envisager que de tels outils permettent d'évaluer l'impact du changement climatique sur ces espèces centrales de certains écosystèmes marins, et nous espérons qu'ils puissent être utilisés dans le cadre d'une véritable gestion écosystémique des stocks et des pêcheries.

Bibliography

Bertrand, A., M. Ballón, and A. Chaigneau, 2010: Acoustic observation of living organisms reveals the upper limit of the oxygen minimum zone. *PlosOne*, **5**, e10330.

Bertrand, A., A. Chaigneau, S. Peraltilla, J. Ledesma, M. Graco, F. Monetti, and F. Chavez, 2011: Oxygen, a fundamental property regulating pelagic ecosystem structure in the coastal Southeastern tropical Pacific. *PlosOne*, **6**(12), e29558.

Bertrand, S., E. Diaz, and M. Lengaigne, 2008: Patterns in the spatial distribution of Peruvian anchovy (*Engraulis ringens*) revealed by spatially explicit fishing data. *Progress in Oceanography*, **79**(2-4), 379–389.

Brochier, T., C. Lett, J. Tam, P. Fréon, F. Colas, and P. Ayón, 2008: An individual-based model study of anchovy early life history in the northern Humboldt Current system. *Progress in Oceanography*, **79**(2-4), 313–325.

Cushing, D., 1975: Marine ecology and fisheries. *Cambridge Univ. Press, Cambridge, England* 278.

Cushing, D., 1990: Plankton production and year-class strength in fish populations: an update of the match/mismatch hypothesis. *Adv. Mar. Biol.*, **9**, 295–354.

Lehodey, P., R. Murtugudde, and I. Senina, 2010: Bridging the gap from ocean models to population dynamics of large marine predators: A model of mid-trophic functional groups. *Progress In Oceanography*, **86**(1-2), 302–315.

Lett, C., P. Penven, P. Ayón, and P. Fréon, 2007: Enrichment, concentration and retention processes in relation to anchovy (*Engraulis ringens*) eggs and larvae distributions in the northern Humboldt upwelling ecosystem. *Journal of Marine Systems*, **64**(1-4), 189–200.

Simon, E. and L. Bertino, 2009: Application of the Gaussian anamorphosis to assimilation in a 3-D coupled physical-ecosystem model of the North Atlantic with the EnKF: a twin experiment. *Ocean Science Discussions*, **6**(1), 617–652.

Appendix A

Spatial Map for Anchovy and Sardines

- Anchovy maps
  Figure A.1,A.2,A.3,A.4,A.7,A.8 show the monthly climatological maps for eggs, larvae and adults for anchovy. Figure A.5 and A.6 show the results of the climatology for August and September using the Calvet net. All these data is used for data assimilation on chapter 6.

- Sardine Maps
  As for anchovy, figure A.9, A.10,A.11,A.12,A.13,A.14 show the monthly climatological maps for eggs, larvae and adults for anchovy. All these data is used for data assimilation on chapter 6.

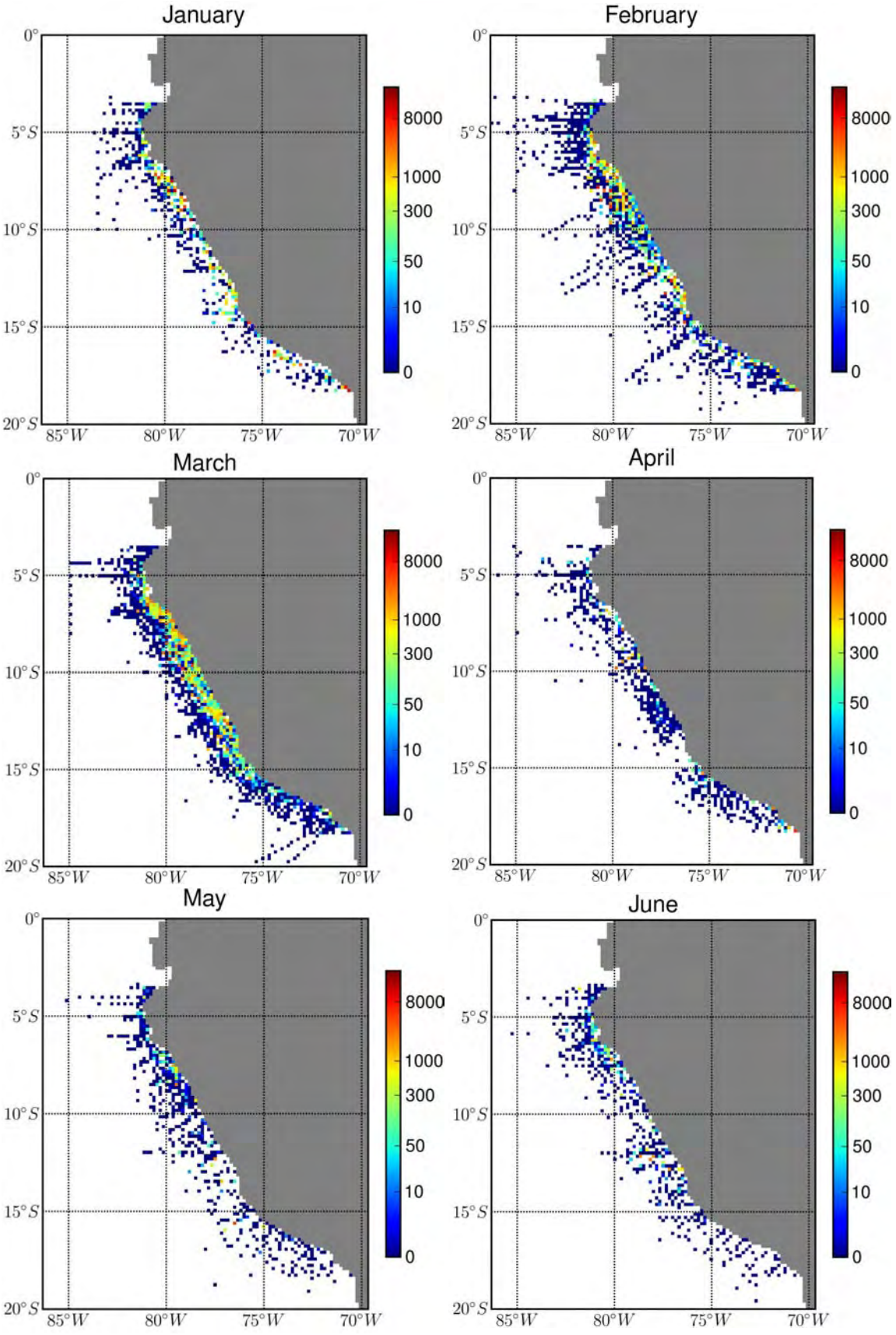

Figure A.1: Anchovy eggs monthly climatology made from observations (1961-2008) from January to June. Only Hensen nets were used to built this climatology. Data in number of Individual per square meters

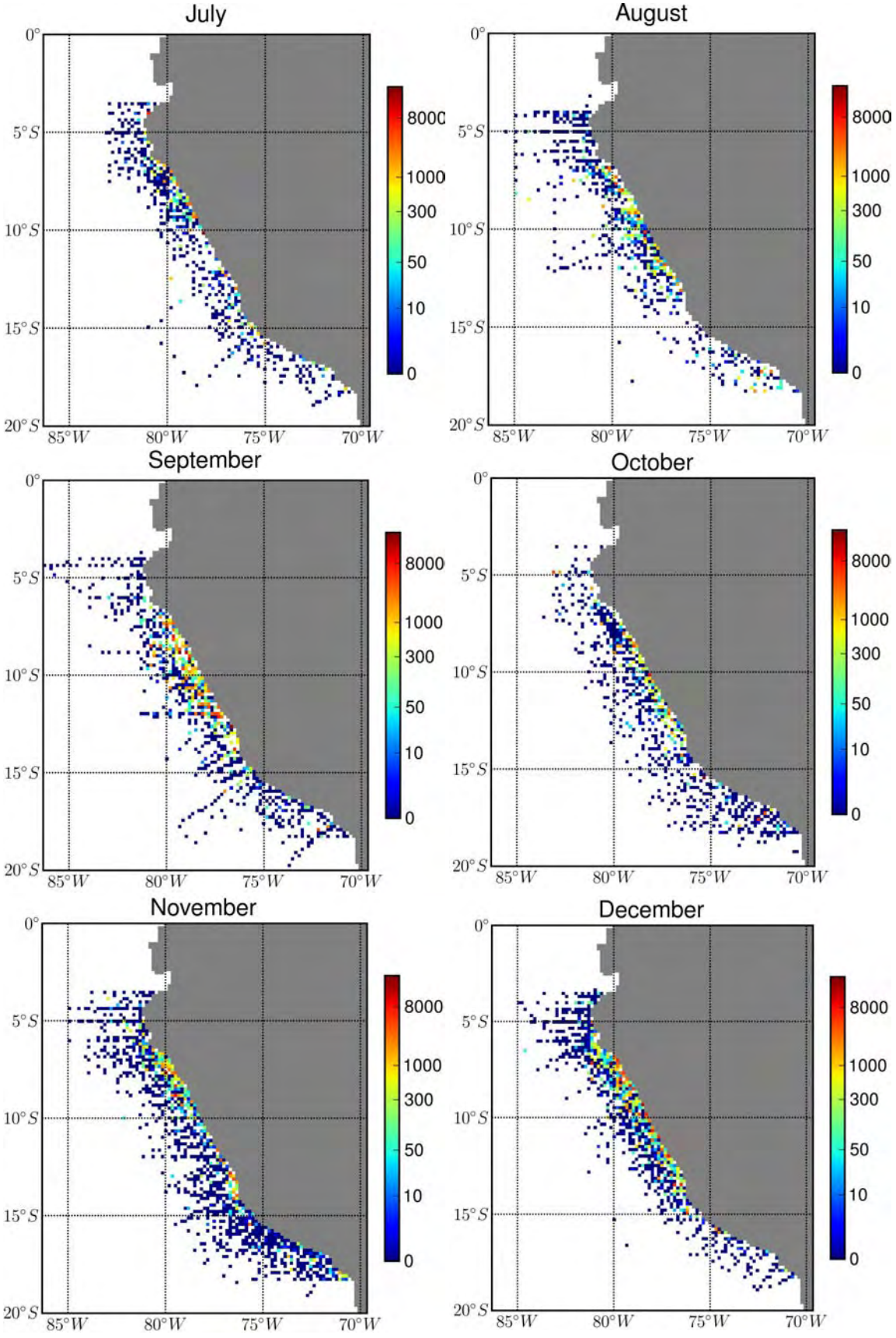

Figure A.2: Anchovy eggs monthly climatology made from observations (1961-2008) from July to December. Only Hensen nets were used to built this climatology. Data in number of Individual per square meters

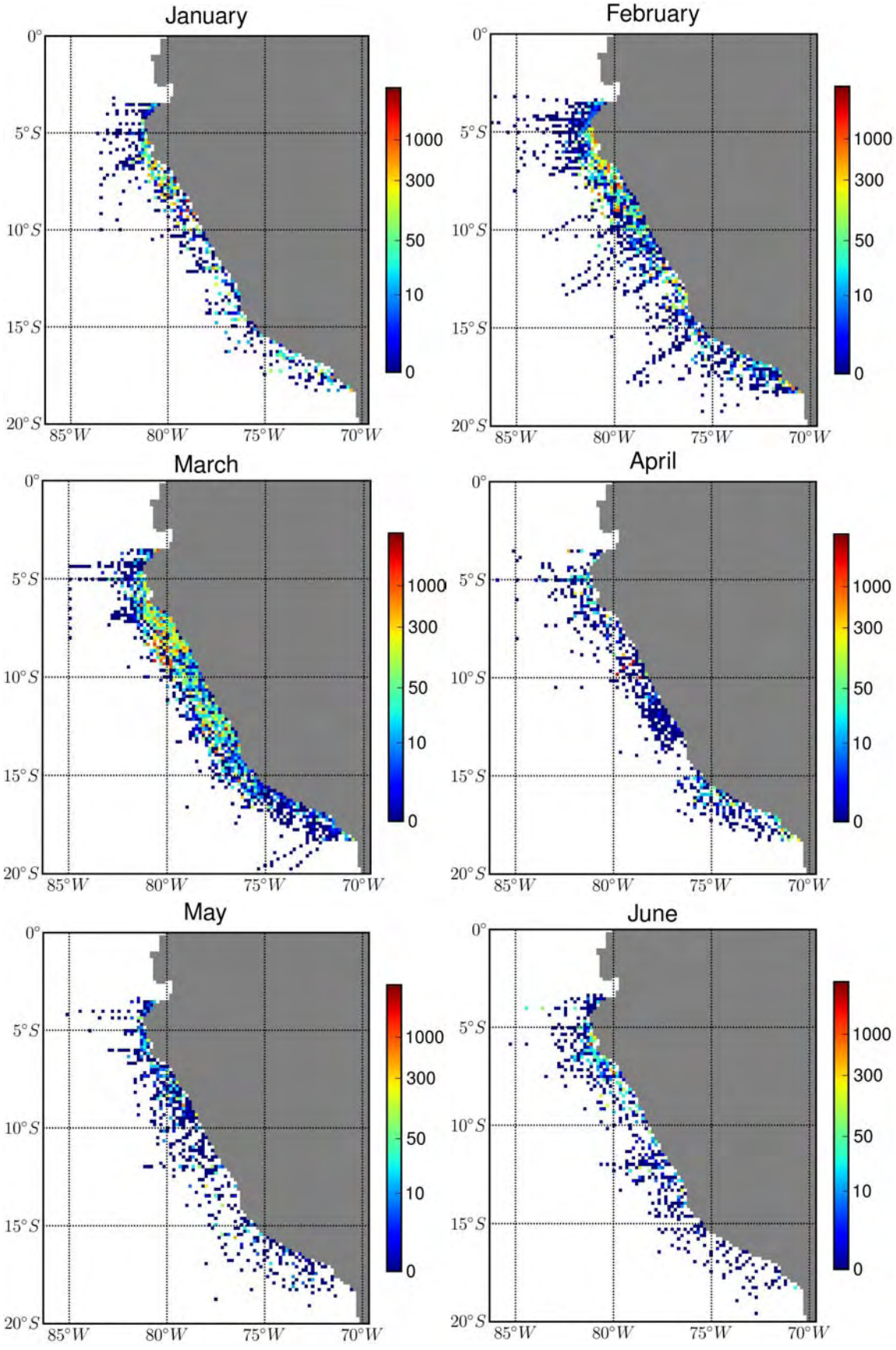

Figure A.3: Anchovy larvae monthly climatology made from observations (1961-2008) from January to June. Only Hensen nets were used to built this climatology. Data in number of Individual per square meters

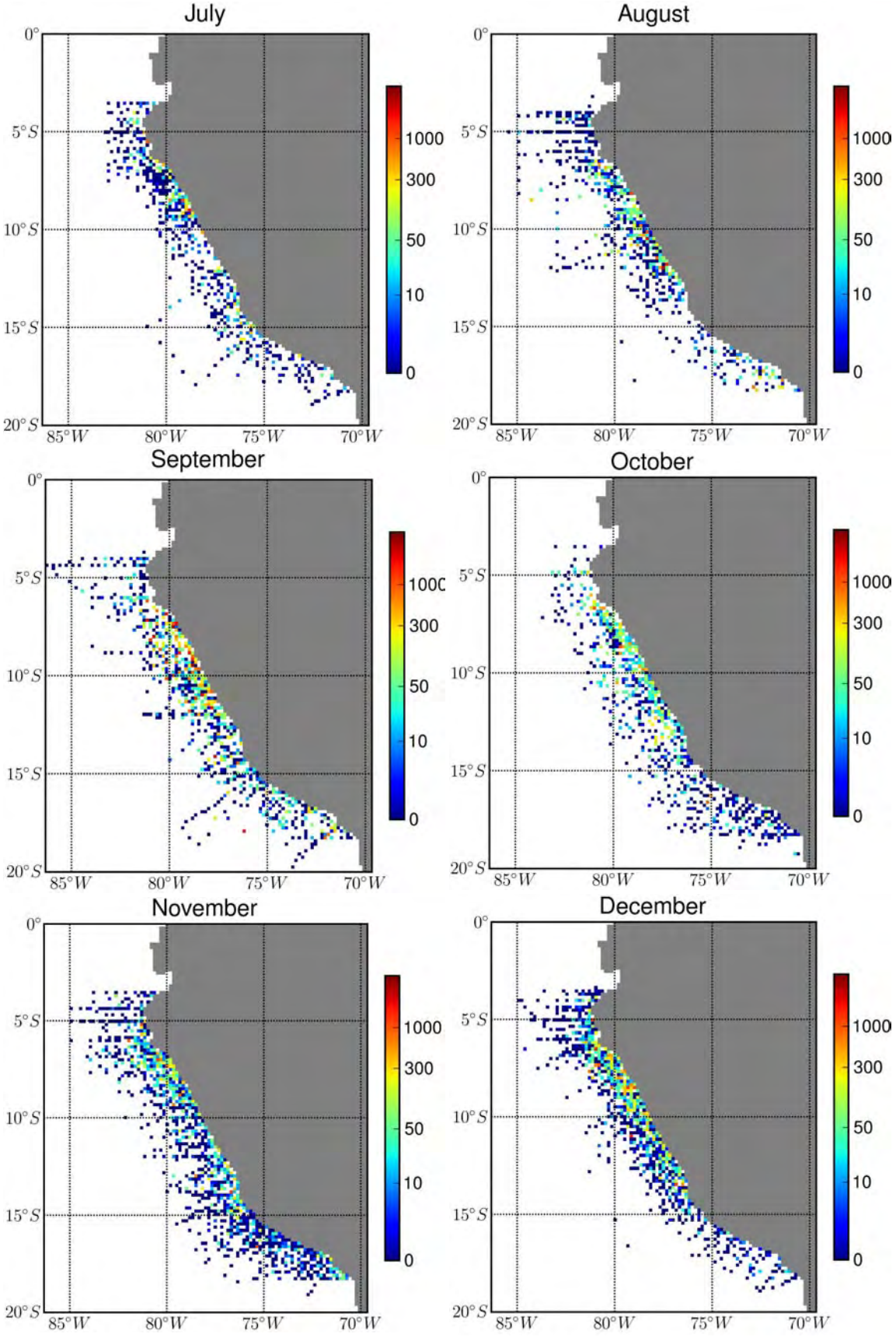

Figure A.4: Anchovy larvae monthly climatology made from observations (1961-2008) from July to December. Only Hensen nets were used to built this climatology. Data in number of Individual per square meters

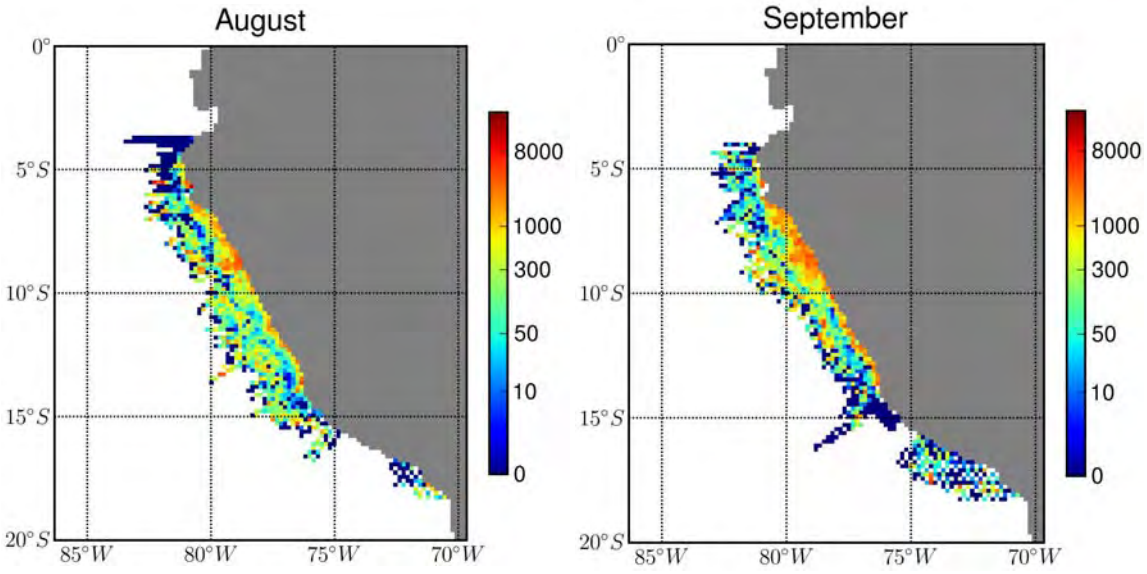

Figure A.5: Anchovy eggs monthly climatology made from observations (1961-2008) for August and September from climatolgy made with Calvet campaigns only.

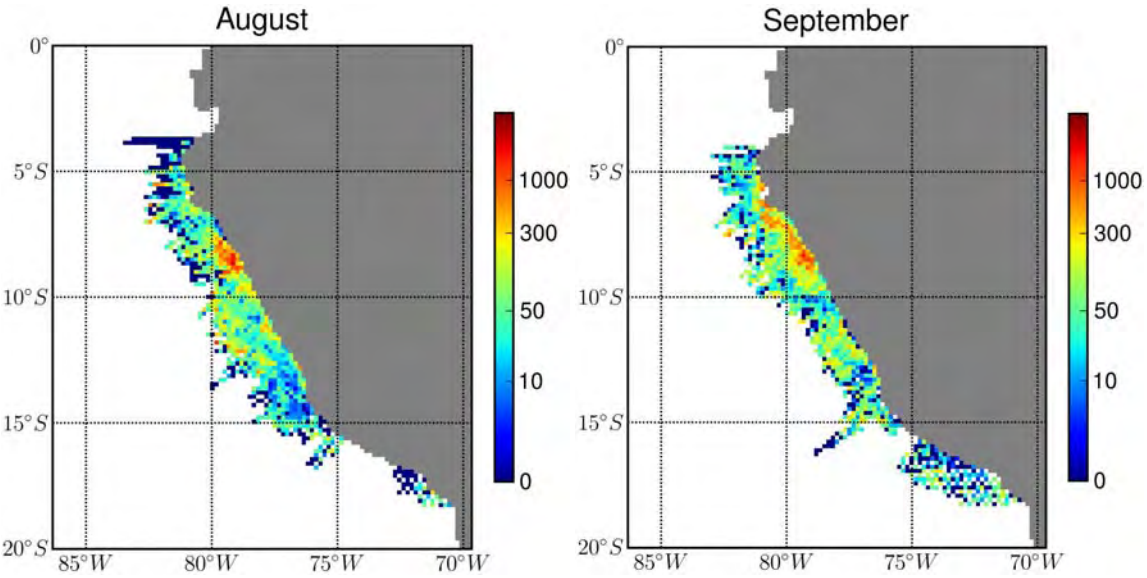

Figure A.6: Anchovy larvae monthly climatology made from observations (1961-2008) for August and September from climatolgy made with Calvet campaings only.

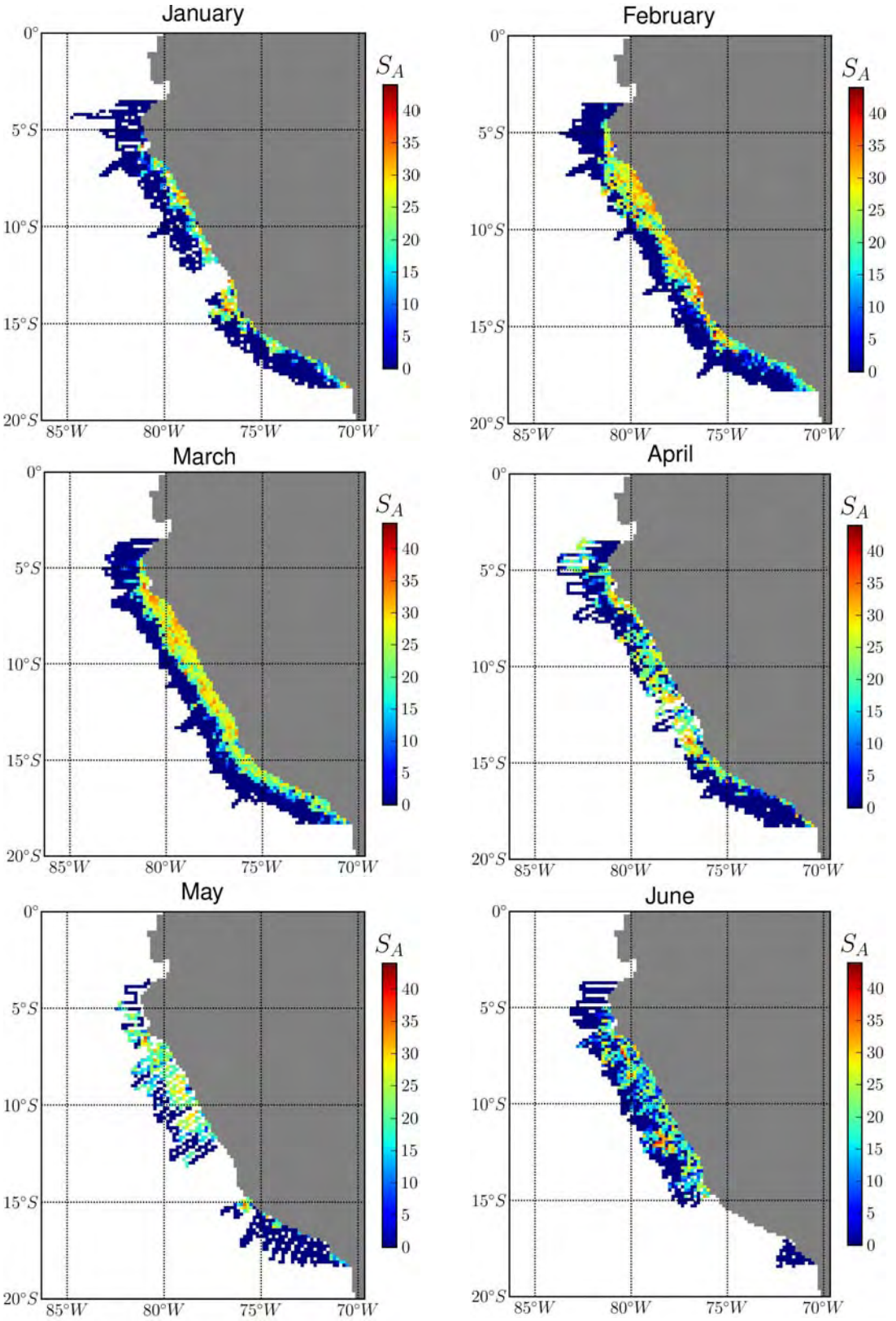

Figure A.7: Monthly climatology made from observations (1983-2009) for the anchovy acoustical abundance index (S_A: Nautical area scattering strength in dB re $1(m^2.nm^{-2})$,$S_A = 10 * log10(s_A)$ from January to June

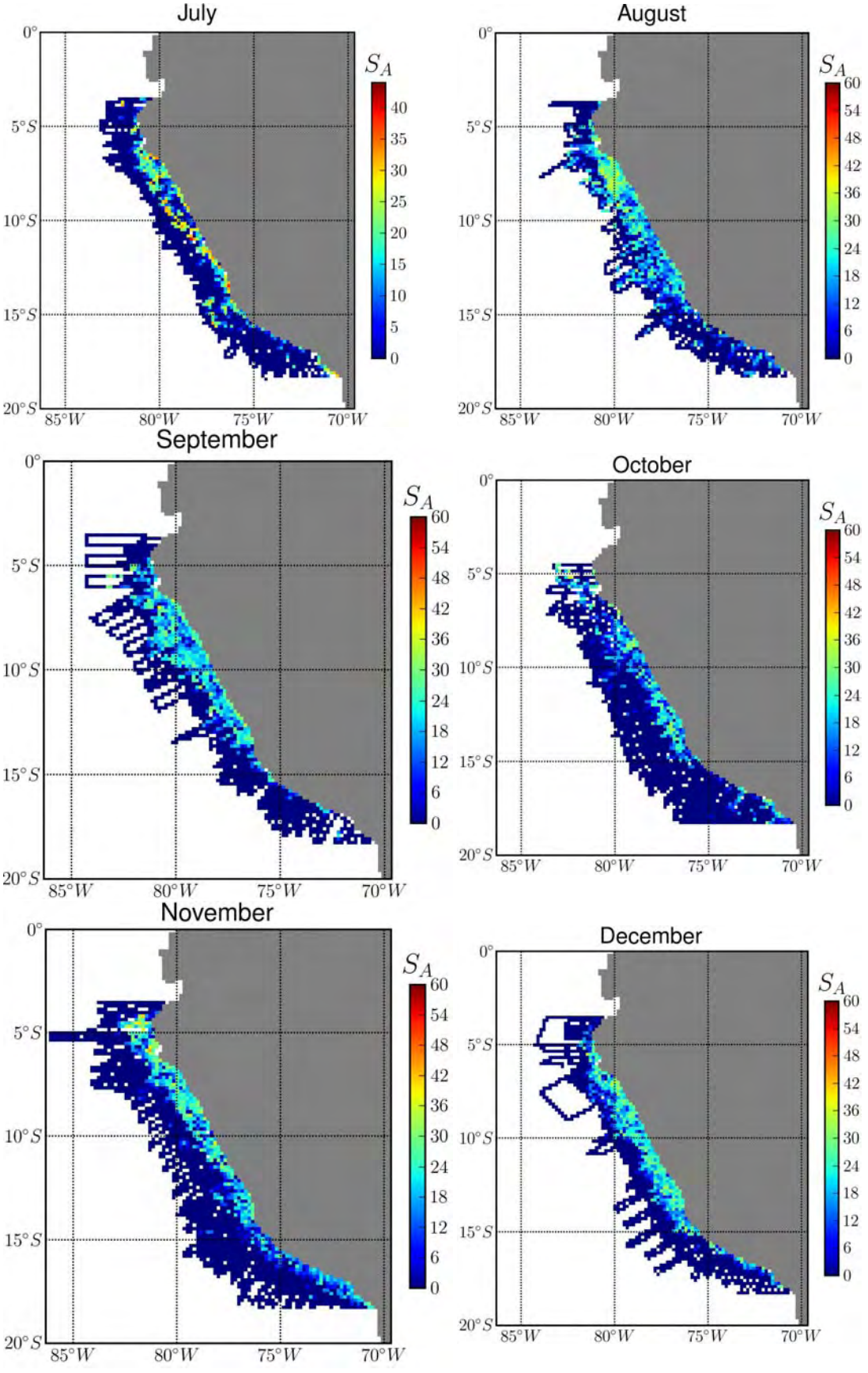

Figure A.8: Monthly climatology made from observations (1983-2009) for the anchovy acoustical abundance index (S_A: Nautical area scattering strength in dB re $1(m^2.nm^{-2})$,$S_A = 10 * log10(s_A)$ from July to December

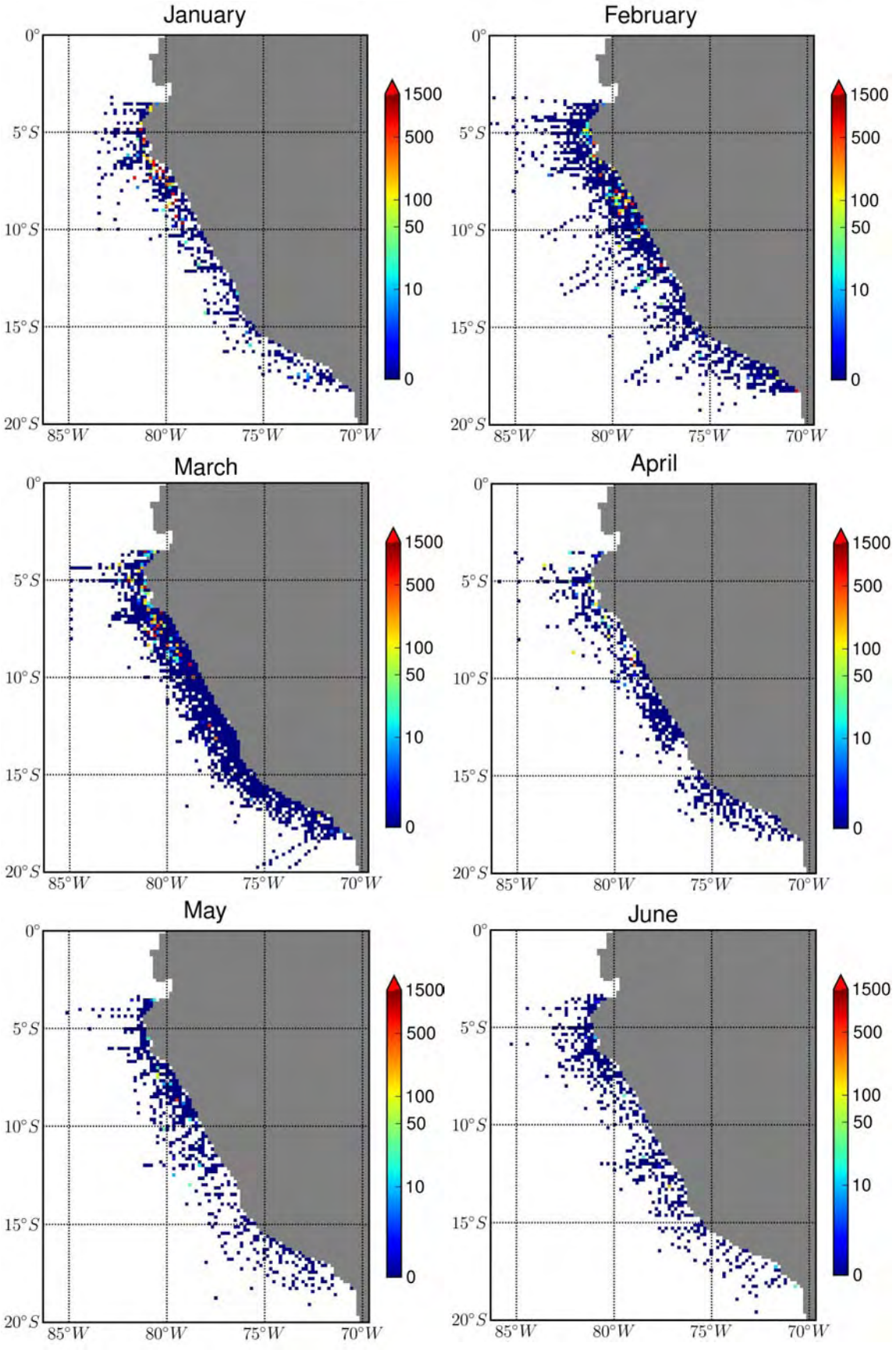

Figure A.9: Sardine eggs monthly climatology made from observations (1961-2008) from January to June. Only Hensen nets were used to built this climatology. Data in number of Individual per square meters. 10224 Ind/m2

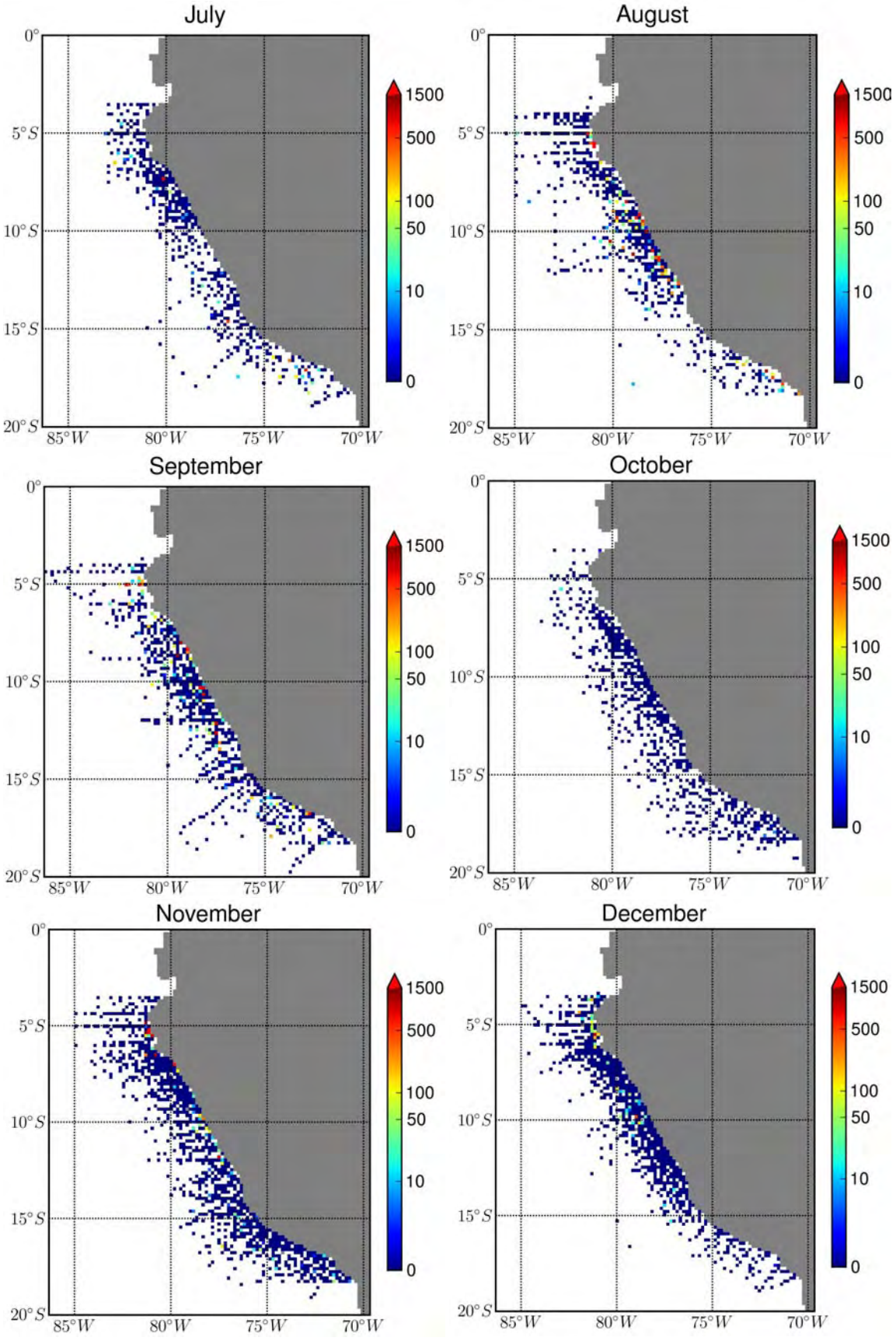

Figure A.10: Sardine eggs monthly climatology made from observations (1961-2008) from July to December. Only Hensen nets were used to built this climatology. Data in number of Individual per square meters 5904

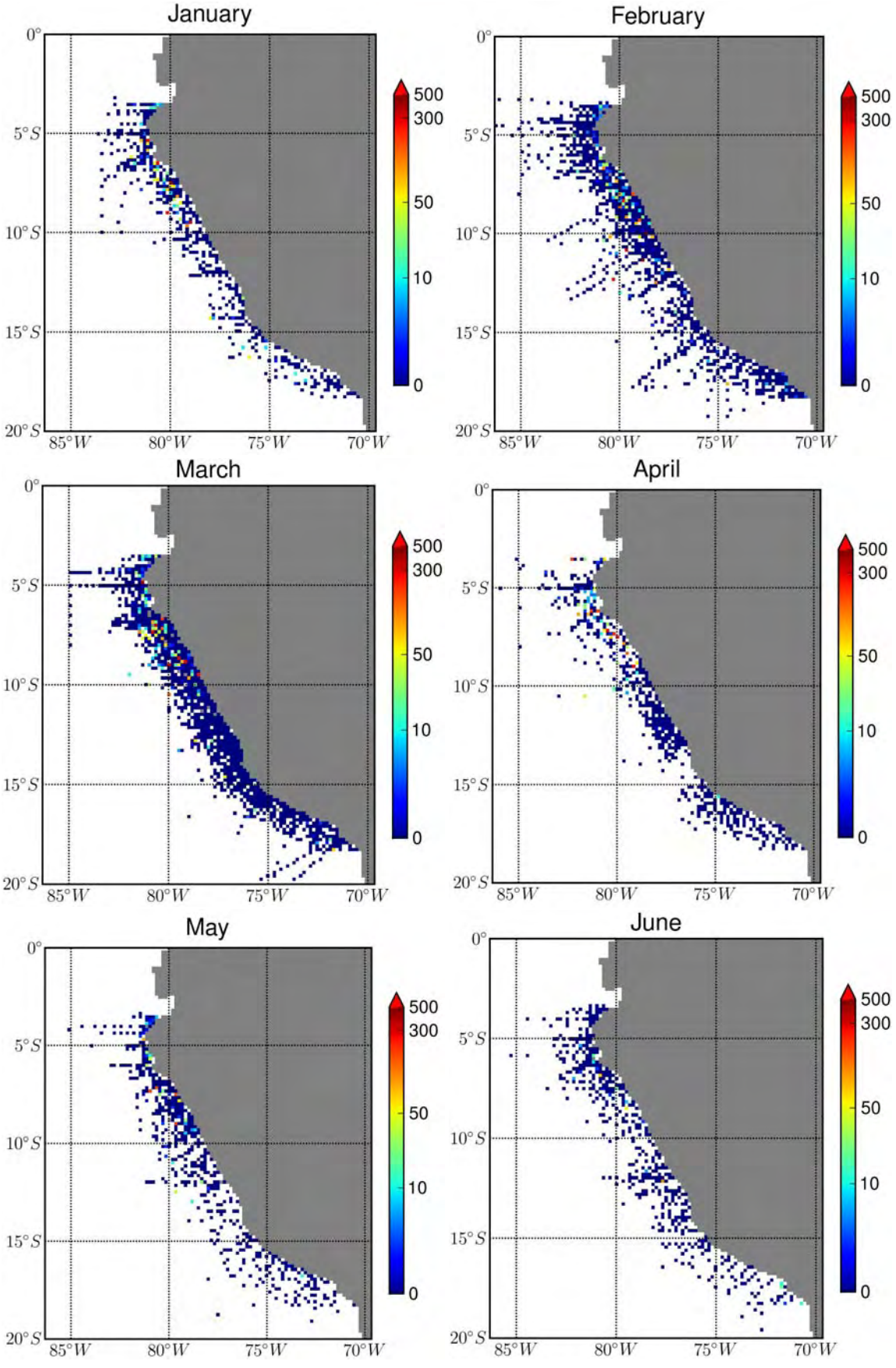

Figure A.11: Sardine larvae monthly climatology made from observations (1961-2008) from January to June. Only Hensen nets were used to built this climatology. Data in number of Individual per square meters

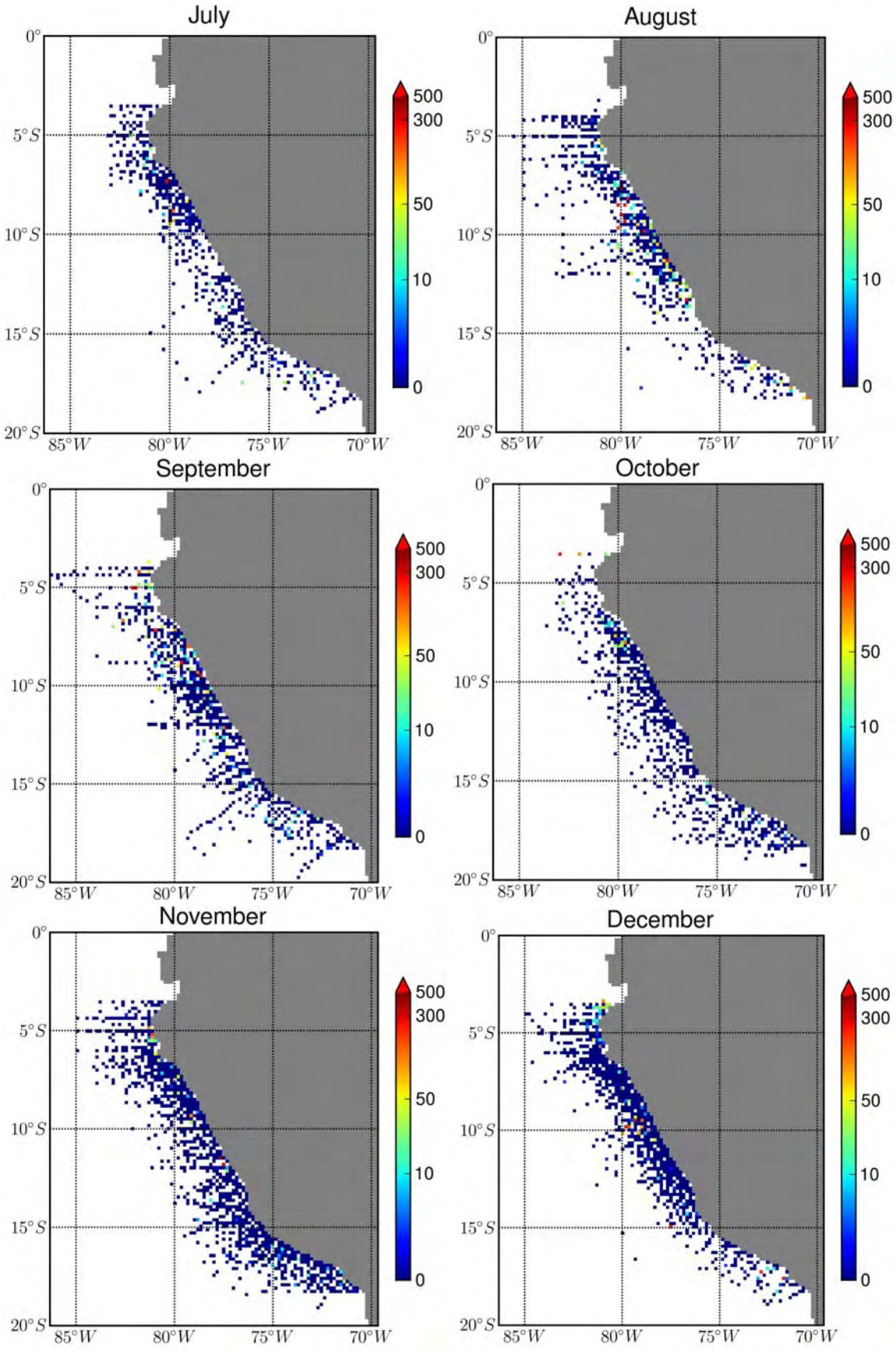

Figure A.12: Sardine larvae monthly climatology made from observations (1961-2008) from July to December. Only Hensen nets were used to built this climatology. Data in number of Individual per square meters

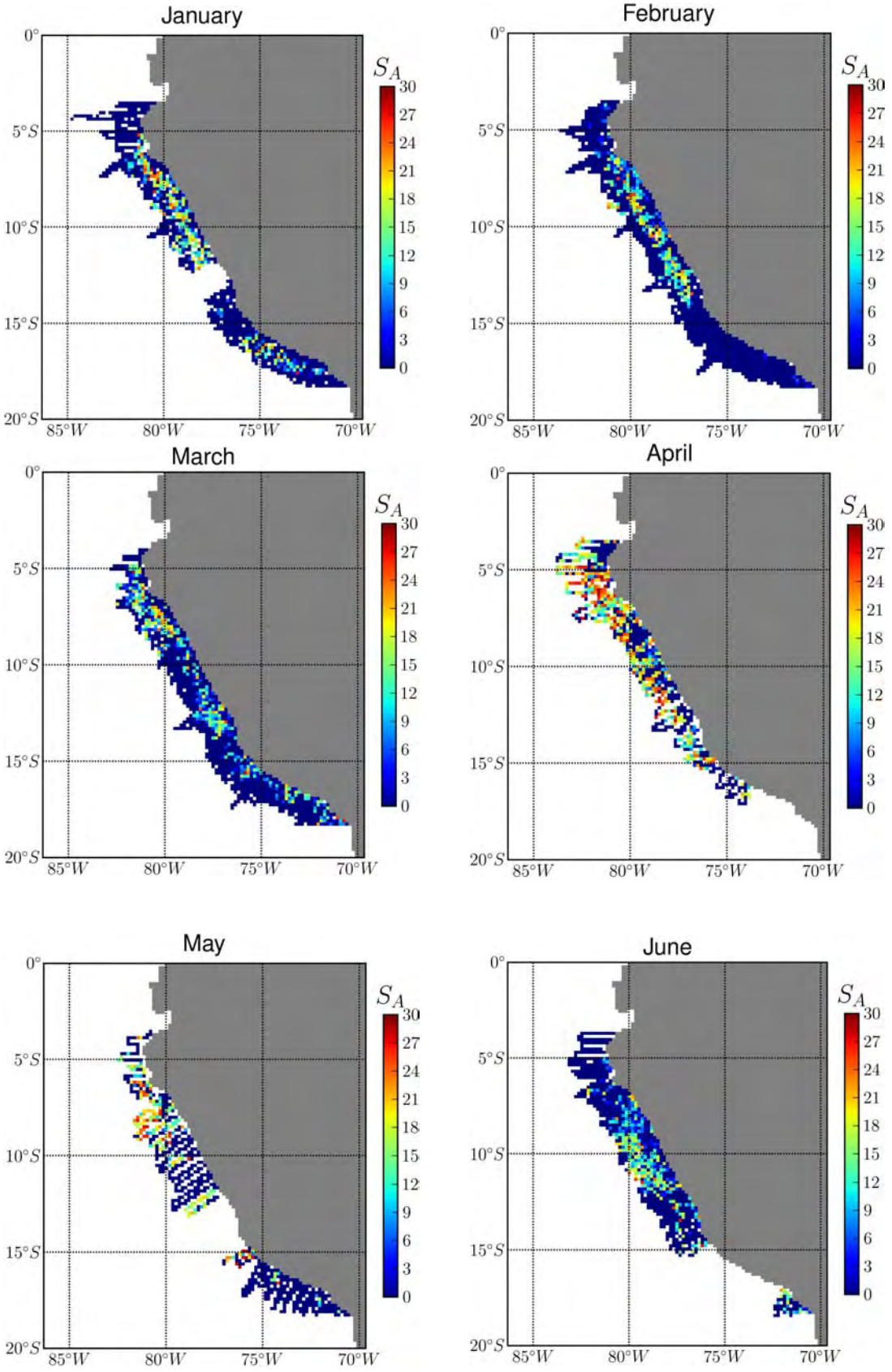

Figure A.13: Monthly climatology made from observations (1983-2009) for sardine acoustical abundance index (S_A: Nautical area scattering strength in dB re $1(m^2.nm^{-2})$,$S_A = 10 * log10(s_A$

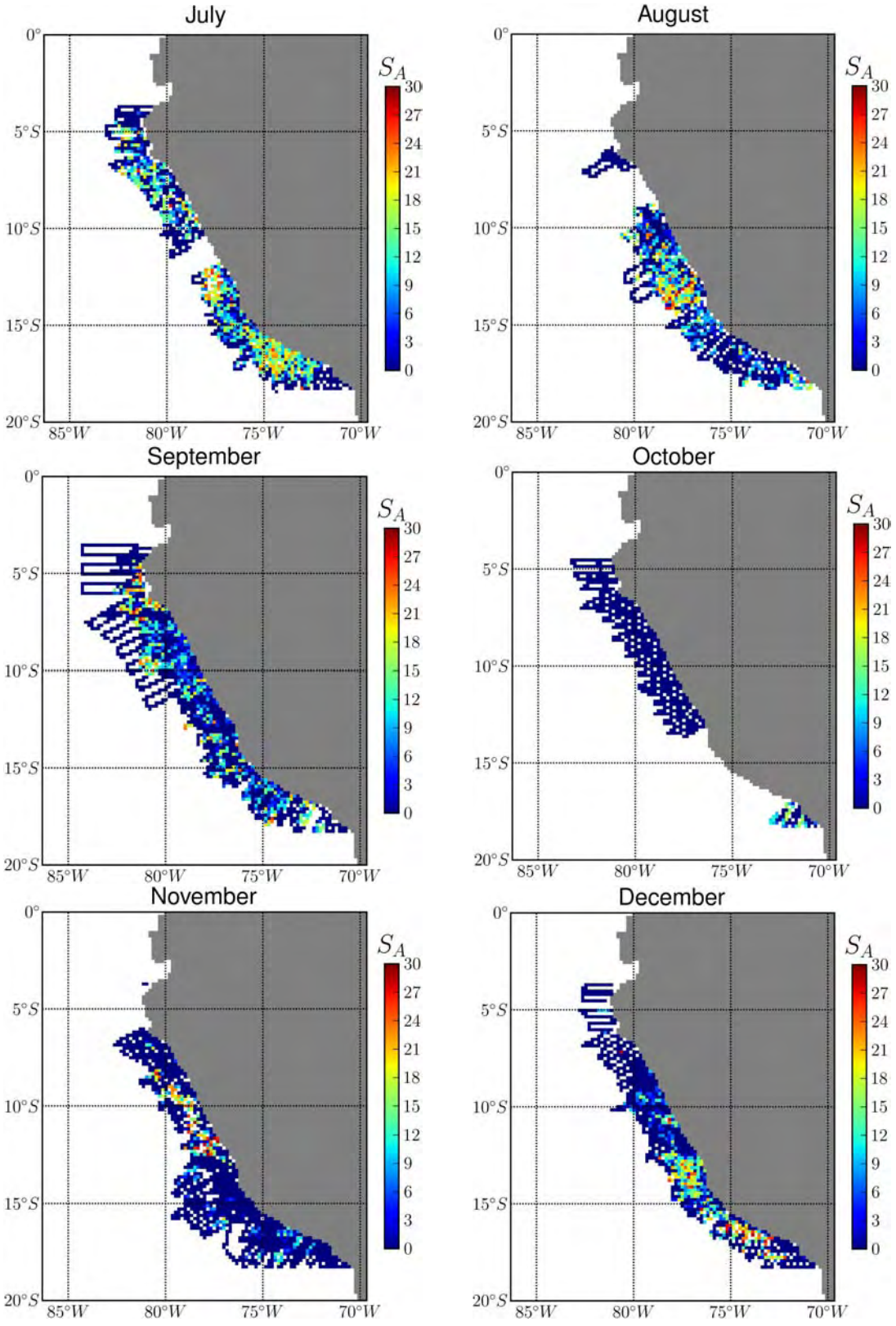

Figure A.14: Monthly climatology made from observations (1983-2009) for sardine acoustical abundance index (S_A: Nautical area scattering strength in dB re $1(m^2.nm^{-2})$,$S_A = 10 * log10(s_A$

Appendix B

Practical coding of the adjoint

To illustrate the concept of the adjoint model of a model as well as how to code adjoint model easily, we provide here an example using a simple case of the SEAPODYM code. Let focus on the definition of the spawning habitat. In the simplest case, we define H_s as a function of a gaussian of temperature. Therefore, the direct code of this function is:

$$H_s = f(T) = \exp \frac{-(SST - b)^2}{2a^2} \tag{B.1}$$

with a and b the parameters of the function to be optimized.
Consider the l-th step of the previous numerical algorithm.

$$(H_s)_l = \exp \frac{-(SST - b_{l-1})^2}{2(a_{l-1})^2}$$

The tangent linear code of H_s can be written as:

$$(\delta H_s)_l = (\frac{\partial H_s}{\partial a})_{l-1} \delta a_{l-1} + (\frac{\partial H_s}{\partial b})_{l-1} \delta b_{l-1}$$

The partial derivatives of H_s in respect to a and b are given by Eq. B.2 and B.3.

$$\frac{\partial H_s}{\partial a} = \frac{(SST - b)^2}{a^3} * \exp \frac{-(SST - b)^2}{2a^2} \tag{B.2}$$

$$\frac{\partial H_s}{\partial b} = \frac{(SST - b)}{a^2} * \exp \frac{-(SST - b)^2}{2a^2} \tag{B.3}$$

In a matrix form, the tangent linear code (the differential of H) is expressed as:

$$\begin{pmatrix} \delta a \\ \delta b \\ \delta H_s \end{pmatrix}_l = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ (\frac{\partial H_s}{\partial a})_{l-1} & (\frac{\partial H_s}{\partial b})_{l-1} & 0 \end{pmatrix} \begin{pmatrix} \delta a \\ \delta b \\ \delta H_s \end{pmatrix}_{l-1} \tag{B.4}$$

The index (l-1) denotes the values of the variables just before the execution of the assignment.

The corresponding adjoint code (noted with *) is obtained by doing the transpose of tangent linear code. It is calculated from a reverse mode (from time l to 0).

$$\begin{pmatrix} \delta a^* \\ \delta b^* \\ \delta H_s^* \end{pmatrix}_{l-1} = \begin{pmatrix} 1 & 0 & (\frac{\partial H_s}{\partial a})_{l-1} \\ 0 & 1 & (\frac{\partial H_s}{\partial b})_{l-1} \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} \delta a^* \\ \delta b^* \\ \delta H_s^* \end{pmatrix}_{l} \tag{B.5}$$

This matrix can be translated to the assignments:

$$\delta a^* = \frac{\partial H_s}{\partial a} \delta H_s^* \tag{B.6}$$

$$\delta b^* = \frac{\partial H_s}{\partial b} \delta H_s^* \tag{B.7}$$

$$\delta H_s^* = 0 \tag{B.8}$$

As observed from this equation, to calculate the tangent code, we need to calculate the gradient of L for the k parameters θ ($\frac{\partial H_s}{\partial \theta_k}$). Therefore, the computational cost will increase linearly with the number of parameters. Using the adjoint code, the computational cost is independent of the number of parameters, and therefore the optimization is more efficient when the number of parameters is high.

This is an example of the adjoint code of a function of SEAPODYM. I implemented the adjoint code for the different spawning habitat functions added for the anchovies and sardines case. More details about adjoint construction can be found in Giering and Kaminski, 1998. Different tests to validate the gradient code construction are detailed in section 6.5.2.

Bibliography

Giering, R. and T. Kaminski, 1998: Recipes for adjoint code construction. *ACM Trans. Math. Softw.*, **24**(4), 437–474.

Appendix C

Tracking blue whales in the eastern tropical Pacific with an ocean-bottom seismometer and hydrophone array

Tracking blue whales in the eastern tropical Pacific with an ocean-bottom seismometer and hydrophone array

Robert A. Dunn and Olga Hernandez[a)]
Department of Geology and Geophysics, University of Hawaii, Manoa 1680 East-West Road, Honolulu, Hawaii 96822

(Received 1 September 2008; revised 21 May 2009; accepted 29 May 2009)

Low frequency northeastern Pacific blue whale calls were recorded near the northern East Pacific Rise (9 °N latitude) on 25 ocean-bottom-mounted hydrophones and three-component seismometers during a 5-day period (November 22–26, 1997). Call types A, B, C, and D were identified; the most common pattern being ~130–135 s repetitions of the AB sequence that, for any individual whale, persisted for hours. Up to eight individual blue whales were recorded near enough to the instruments to determine their locations and were tracked call-by-call using the B components of the calls and a Bayesian inversion procedure. For four of these eight whales, the entire call sequences and swim tracks were determined for 20–26-h periods; the other whales were tracked for much shorter periods. The eight whales moved into the area during a period of airgun activity conducted by the academic seismic ship R/V *Maurice Ewing*. The authors examined the whales' locations and call characteristics with respect to the periods of airgun activity. Although the data do not permit a thorough investigation of behavioral responses, no correlation in vocalization or movement with airgun activity was observed. © *2009 Acoustical Society of America.* [DOI: 10.1121/1.3158929]

PACS number(s): 43.30.Sf, 43.80.Nd [RAS] Pages: 1084–1094

I. INTRODUCTION

The blue whale (Balaenoptera musculus) populates all of the world's oceans, forms vocally distinct groups, and has long-range seasonal migrations. Because of historic whaling pressure, it is considered endangered throughout its range and has been protected internationally since 1965 (Yochem and Leatherwood, 1985). Common to all blue whales is emission of high intensity, low frequency, and long duration acoustic calls in repetitive patterns, possibly used for communication (e.g., Stafford *et al.*, 1999, 2001; Thompson *et al.*, 1996; McDonald *et al.*, 2006). Owing to their high source levels (189 dB re 1 μPa at 1 m) and low frequencies (14–100 Hz), these calls can be detected at up to 200 km distance on bottom-moored hydrophones (Širović *et al.*, 2007). Blue whales are also highly vocal, producing distinct amplitude- and phase-modulated calls in repetitive patterns that allow tracking of individual animals, which is of prime importance to detailed behavioral studies. In an environment where individuals are often dispersed, passive acoustic methods are effective means to study their presence, movements, and calls.

From November 9–28, 1997, an array of ocean-bottom seismometers and hydrophones was deployed along the northern East Pacific Rise to collect seismic data during an active-source seismic study of the magma chambers beneath this volcanic system (Fig. 1). We recently examined the data for whale calls and found that blue whale calls were recorded during 5 of 20 days of recording. We present an analysis of calls from eight blue whales and track the whales using their calls and a localization algorithm based on a probabilistic grid search method. We obtained long, complete call sequences for four of the eight whales, and were able to track their motions for 20–26 h intervals. Reconstructions of continuous swim tracks of individual blue whales lasting more than just a few hours are rare (e.g., Watkins *et al.*, 2004).

The whales entered the area while the academic seismic ship R/V *Maurice Ewing* carried out a seismic experiment using a 20-gun, 139-l airgun source. The proximity of the whales and ship allowed us to examine their calls and swim tracks for any anomalous behavior with regard to airgun use. Sounds produced by airgun arrays have garnered increasing interest as there are concerns regarding the potential impact of airgun noise on marine mammals (Malakoff, 2001, 2002; National Research Council, 2003).

II. INSTRUMENTATION AND DATA

The study site is located in the eastern tropical Pacific Ocean, 750 km southwest of Mexico's coastline, in 2600–3200 m of water, and is centered on a section of the northern East Pacific Rise. An array of ocean-bottom seismometers and hydrophones was deployed over a 200-km-section of the mid-ocean ridge with a minimum station spacing of 12 km (Dunn *et al.*, 2001). Not all instruments recorded blue whale calls; those that did record calls are shown in Fig. 1. The instruments consisted of a mix of ocean-bottom receivers from the Woods Hole Oceanographic Institution: 9 ocean-bottom hydrophones (OBHs), 3 ocean Reftek in a ball (ORB) equipped with a hydrophone, and 13 Office of Naval Research three-component seismometers (OBSs) equipped with a hydrophone. We used both the hydrophone and vertical component seismometer data for this study. Recordings were made with a sampling rate of 200 Hz for the OBH and

[a)]Present address: MEMMS (Marine Ecosystem Modeling and Monitoring by Satellite), CLS, Satellite Oceanography Division, 8-10 rue Hermès, 31520 Ramonville, France.

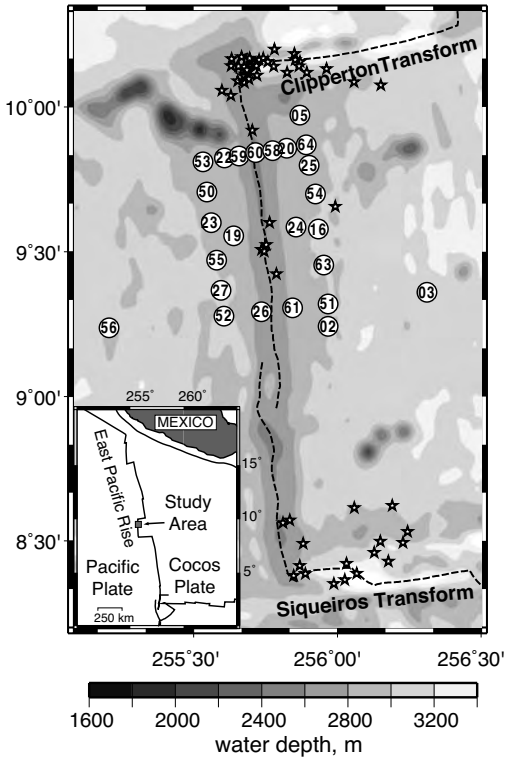

FIG. 1. Seafloor bathymetry map showing location of the East Pacific Rise (darker shading down the center of the figure; the dashed line indicates the axis of the ridge) and locations of the instruments that recorded blue whale calls (numbered circles). The instruments are a mix of ocean-bottom receivers from the Woods Hole Oceanographic Institution: 3 ORB (numbers 2–5) equipped with a hydrophone; 9 OBH (numbers 16–27), and 13 OBS (numbers ≥50) equipped with a hydrophone. Stars indicate epicenters of 58 locatable earthquakes, out of >580 total that occurred in the region during the study period.

ORB and 128 Hz for the OBS. For these instruments the useful band for acoustic detection of blue whales is between 5 and 60 Hz.

All 20 days of the seismic records were examined. Blue whale calls were detected from November 22–26, 1997 on subsets of the ocean-bottom stations, depending on the location of each whale with respect to the instrument array (i.e., only the closest instruments recorded the whale calls with large enough signal-to-noise ratio for analysis). Vocalizing blue whales entered the area during the latter part of the seismic experiment, after 13 days of intermittent airgun activity (Fig. 2). No other marine mammals were identified in the data. Apart from blue whale calls and the *Ewing* activities, the most common signals recorded were regional earthquakes, which tend to occur in high numbers along mid-ocean ridges and impart significant acoustic energy to the water column. Over the 20-day recording period the instruments detected more than 580 distinct earthquake events.

Throughout the seismic study, the position and speed of the R/V Maurice Ewing were digitally logged every minute and information on the status of the airgun array was logged at the time of each airgun pulse. The airgun array consisted of 20 bolt airguns that varied in volume from 145 to 875 in.^3 for a total discharge volume of 8503 in.^3 (~139 l); at the source, the airgun output was 237 dB (re 1 μPa P-P at 1 m). This is the effective output of the airgun array, as if the energy emanated from a point source. Because the array is spread over a large area, the actual output is much lower. Towed 40 m behind the ship and at 10 m depth, the array generated acoustic pulses every 210 s (150 s on November 25) as the ship traveled a pattern within the instrument network. The airgun array is designed to focus energy downward, rather than to the sides, and there is an azimuthal variation of the energy emission, with the highest levels emitted fore and aft of the ship and significantly less energy emitted to the sides.

Under current guidelines, the National Marine Fisheries Services defines the radii around airgun sources with received sound levels of 180 dB as a safety radii for cetaceans (NMFS, 2005); the radii with received levels of 160 dB are considered to be distances within which some cetaceans are likely to be subject to behavioral disturbance (NMFS, 2005). With regard to the Ewing's airgun array, theoretical calculations (Diebold, 2004) and field calibration studies (Tolstoy *et al.*, 2004) show that sound levels produced by the array depend on the depth of observation. Therefore, received levels at the whale will depend not only on the distance from the ship but the depth of the whale. Studies of dive characteristic of blue whales off the central California coast (Lagerquist *et al.* 2000) show that 72% of all dives are between 0 and 16 m and less than 1 min duration; the second most frequent dive interval is 97–152 m, accounting for 15% of all dives and <1.2% of the whale's total time underwater. Blue whales seldom dive to 150 m depth and even more rarely to greater depths. Theoretical calculations for the array used in this experiment (Diebold, 2004) indicate that peak sound levels of ≥180 dB (re 1 μPa rms) occur within 250 m of the array

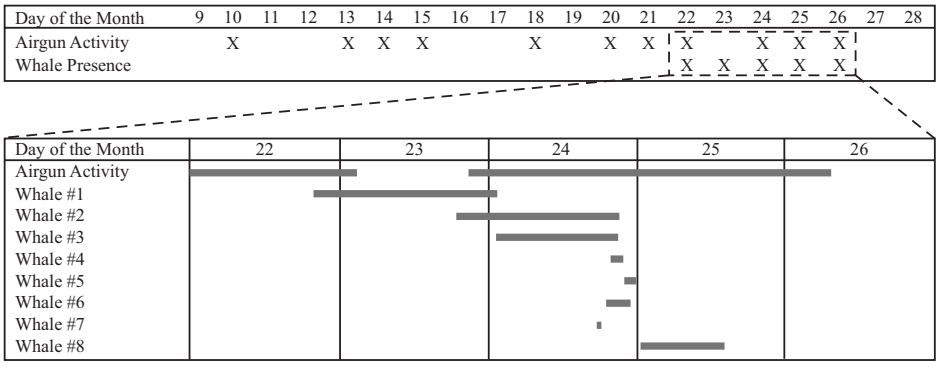

FIG. 2. Time line of events showing periods of airgun activity and periods when whale calls were detected and tracked. The numbering of the whales used here is consistent with a numbering scheme used throughout the text. The whales appeared in the study area after 13 days of on-and-off airgun activity.

at 10 m depth below the sea surface and within 500 m of the array at 125 m depth; received levels of ≥160 dB (re 1 μPa rms) within 600 m of the array at 10 m depth and within 1.75 km at 125 m depth. At 10 km distance, received levels are calculated to be 133 dB at 10 m depth and 136 dB at 125 m depth. A field calibration study (Tolstoy *et al.*, 2004) of the Ewing 20-gun array (with an ~8600 in.^{3} volume array rather than the 8503 in.^{3} array used here) found 180 dB (re 1 μPa rms) levels within ~1 km of the array and 160 dB levels within ~2–3 km of the array (recording depths were 18 and 500 m). At distances greater than 5 km, the received energy was <145 dB, with the majority of the energy in the 5–100 Hz range.

III. TRACKING METHOD

We adopt a Bayesian inversion method for locating the whales (Tarantola and Valette, 1982). The basic idea is to define the solution in terms of an *a posteriori* probability density that incorporates the data (onset times of vocalizations), the model parameters (vocalization position), and the theoretical link between model parameters and calculated synthetic data. From this, the most likely location of the whale is given as the position where the *a posteriori* probability is a maximum. About the maximum likelihood point, a full representation of the 95% probability region for the whale's location is easily extracted from the probability density.

The method involves a brute-force grid search comparing arrival time measurements of whale calls to predicted arrival times for all locations on a spatial grid. The synthetic arrival times are calculated to all grid points in advance and stored for later use, greatly improving the efficiency of the method. The benefits of this method are numerous: it allows for non-linear travel time calculations, it is consistent with respect to a change of variables, it allows for general error distributions in the data, it incorporates theoretical errors that arise from inaccurate parametrizations and theoretical simplifications, it allows for the formal incorporation of any *a priori* information concerning the location parameters (such as a probability distribution for the whale's location derived from an estimate of its position at an earlier time), and it provides a full representation of the probability of a whale's location. In short, the Bayesian inversion method is flexible and provides a mathematically robust location of a whale's location given noisy, sparsely recorded data.

The unknowns in the problem are the spatial coordinates of a whale (x,y), where x is longitude and y is latitude. Given the large station separation (>3 times the water depth), the data are incapable of resolving accurate whale depths. Given that vocalizing blue whales tend to spend the majority of their time near the ocean surface (Oleson *et al.*, 2007), we simply assume that the whale is located at the surface when calculating its lateral position. In practice, we define a grid over the model space $\mathbf{m}=(\mathbf{x},\mathbf{y})$, which is the area of the ocean for which we might find the whale. Because of transmission loss and detection thresholds, any recorded whale will be within ~50 km of the nearest station. Thus, we used a uniform 200×200 km^2 grid centered on the stations. The general solution (without the depth and time dependence) is

$$P(\mathbf{m}) = K\rho(\mathbf{m})\exp\left\{-\sum_{i=1}^{N}\frac{|t^i_{\mathrm{obs}} - t^i(\mathbf{m})|}{\sigma_i}\right\}. \quad (1)$$

The values t^i_{obs} are the N observed arrival times of a single whale call minus a weighted average of all such times [Tarantola and Valette, 1982; Eqs. 10–12]. Measurement of the arrival times can be made by several different methods and is a critical step in the location problem. The best results were obtained by band pass filtering the data to isolate the fundamental frequency of the B call (or the stronger 48 Hz overtone for the case of whale 3), lying the filtered time series over the spectrogram of the call, and then handpicking the onset of the B call from the joint time series and spectrogram plot. Picking onset times is generally less accurate than cross-correlation (e.g., Nosal and Frazer, 2006), but in our case the overlap in both the time and frequency domains of airgun pulses with the whale calls makes cross-correlation impractical. The 1-σ pick uncertainties, ranging from 0.1 to 6 s, are the largest source of error in the location problem but are included in the location method and weight their respective measurements.

The values $t^i(\mathbf{m})$ in Eq. (1) are the theoretical travel times from a grid location in $\mathbf{m}$ to each station for which exists a value t^i_{obs} minus a weighted average of these times [Tarantola and Valette, 1982; Eqs. 10–13]. The travel times were calculated using an algorithm that at distances far from a seismic station (greater than approximately three times the water depth) mimics T-phase propagation: acoustic energy travels along a direct path at a constant acoustic speed. At distances closer to a seismic station, it is important to account for the water depth of the seismic station and our algorithm models the acoustic propagation along a direct path from the whale (at the sea surface) to the station (on the seafloor). Given the experiment geometry (sparse station layout with large separation) and large pick errors, a more accurate acoustic propagation model would not result in appreciably better whale locations.

The values σ_i are a combination of observational and theoretical uncertainties: $\sigma^2=\sigma_t^2+\sigma_T^2$, where σ_t are the uncertainties corresponding to each measured time, t^i_{obs}, derived from the picking procedure, and σ_T are the uncertainties in the travel time calculations from a grid point in $\mathbf{m}$ to a recording instrument. The σ_T values include the uncertainty of the instrument positions, the uncertainty due to the method of travel time calculation, and the uncertainty of the acoustic medium. The instrument coordinates and uncertainties were determined via an inverse procedure using the travel times of the airgun pulses from the ship (whose position is accurately known via global positioning system) to the instruments. The 1-σ errors of the instrument positions are 3–150 m, depending on the amount and distribution of data available for each instrument. The error due to the acoustic path calculations and to unknown deviations in the acoustic velocity is <1 s.

By adding each of the variances of the different error sources, the total expected uncertainty in the theoretical calculation is ~1 s.

$\rho(\mathbf{m})$ consists of any *a priori* information that may exist on the whale's position (other than the travel time measurements) before we calculate an estimate of the position. If no such information exists, then initially the whale has equal probability of being anywhere on the grid and $\rho(\mathbf{m})=1/M$, for all values of $\mathbf{m}$, where M is the number of grid points. Thus a summation of $\rho(\mathbf{m})$ over the total model space yields a value of 1 (100% probability that the whale is somewhere on the grid, with equal probability at all locations). On the other hand, given the speed at which a whale can swim, we could state that the position of the whale at the (i+1)th call must be close to that at the *i*th call. In this case, we could write the *a priori* information function for the whale's position as

$$\rho(\mathbf{m}) \propto \exp\left\{-\frac{1}{2}[\mathbf{m}-\langle\mathbf{m}_{\mathrm{i}}\rangle]^T\left(\frac{1}{R^2}\right)[\mathbf{m}-\langle\mathbf{m}_{\mathrm{i}}\rangle]\right\},$$

where $\langle\mathbf{m}_{\mathrm{i}}\rangle$ is the estimated position at the previous call and R (units of distance) is the product of the average speed of the whale and the amount of time lapsed since the previous position was estimated. In other words, we establish *a priori* a Gaussian probability density such that there is a ~95% probability of the whale being within $2R$ of the previously estimated position. We used this approach because it tends to smooth the track of the whale, which is otherwise noisy due to the large pick uncertainties. We also post-processed the tracks with a three-point averaging filter to further reduce spurious call-to-call position noise.

Equation (1) is normalized by the constant K such that the probability of the whale being somewhere within the grid is 100%. Since the model parameters, $\mathbf{m}$, are discrete, K is defined as

$$K=\left(\sum\sum P'(x,y)\right)^{-1},$$

where P' are the un-normalized values from Eq. (1).

The probability density P($\mathbf{m}$) provides a full representation of the probability of a whale's coordinates. The maximum likelihood position of the whale is the position where P($\mathbf{m}$) is maximum and is thus the position where the weighted data misfit is a minimum [in the case of constant $\rho(\mathbf{m})$]. Our method uses the L_1 norm to quantify misfit length, because a solution is thereby less biased by outliers in the data. The shape of the distribution, which is not necessarily elliptical about the maximum likelihood position, provides a "map" of the uncertainty of the whale's position. Using a running algorithm over the time series of all identified calls, we calculated whale locations when at least four call observations were available on separate receivers; we rejected any locations when the misfit,

$$\Phi=\frac{1}{N}\sum_{i=1}^{N}\frac{|t^i_{\mathrm{obs}}-t^i(\mathbf{m})|}{\sigma_i},$$

exceeded a value of 1.5 or the location uncertainty exceeded 3 km.

IV. BLUE WHALE CALLS

Comparisons of the calls in our data with those of other studies reveals that our records are from a northeastern Pacific population of blue whales (e.g., Stafford *et al.*, 1999); a typical spectrogram, showing the ACB call components of northeastern Pacific blue whales, is shown in Fig. 3(a). A few D calls were also recorded [Fig. 3(b)] (e.g., Thompson *et al.*, 1996; Aroyan *et al.*, 2000; Mcdonald *et al.*, 2001). On ORB03, D calls are present November 23 around 2100 and 2230 GMT and again on November 25 around 2300 GMT. On ORB02, D calls are present from 1800 GMT November 24 to 0700 GMT November 25.

Of the whales studied here, the A calls have durations that last 20–30 s and the duration of the B calls is approximately 15–20 s (Fig. 4). The time between the onset of the A call and the onset of the B call is variable from whale-to-whale and between the calls emanated by any one whale and tends to be in the 50–60 s range. The duration of C calls is about 12 ± 1 s. We suggest caution when examining call durations in this and other data sets, since multipathing of the acoustic energy tends to elongate the apparent duration of calls in both the time series and spectrograms, and this elongation will be environmentally dependent. Furthermore, the rise time of the A call is very slow and A call duration measurements will be inaccurate for distant, noisy records. However, there are some anomalies in the calls particular to individual whales allowing them to be identified separately from other whales. One whale (whale 1) exhibits a brief discontinuity in its B calls and a pulse in the overtones at the end of its A calls [Fig. 3(a)]. Another whale (whale 3) exhibits a shortened A call rapidly followed by a 7–8 s un-modulated ~16 Hz tone (Fig. 4), which could be considered a separate call, but here we refer to both parts of this call as an "anomalous" A call. Some whales exhibit shorter A call durations and some longer, but this is not sufficient to identify individuals since multiple whales can exhibit one of the two durations, the durations are not necessarily constant across all calls of a single whale, and there is a bias toward measuring shorter times as the whale moves further from the recording instrument and signal-to-noise decreases.

Typically the calls appear in sequences of A and B combinations. The C call was nearly always recorded when a whale was close to an instrument, but generally undetectable at other times due to its low amplitude. Therefore, it is to be understood that, unless specific reference is made to the C call, a C call may have been present but is omitted from the discussion. The most common call pattern in the data is a sequence of AB calls (i.e., ABABAB $\cdots$). Only whale 3 deviated from this pattern, by forming repetitions of one A call followed by more than one B call, such as repetitions of ABB or ABBB (each B call is preceded by a C call in this case). Consecutive A calls were not recorded and each sequence starts with an A call. These patterns were repeated regularly, often for many hours. Using a 95% confidence interval ($\alpha=0.05$), the A-to-B and B-to-B spacings of calls were not statistically distinguishable between night and day. In a few rare instances, an A call was followed by a short silence, rather than a B call.

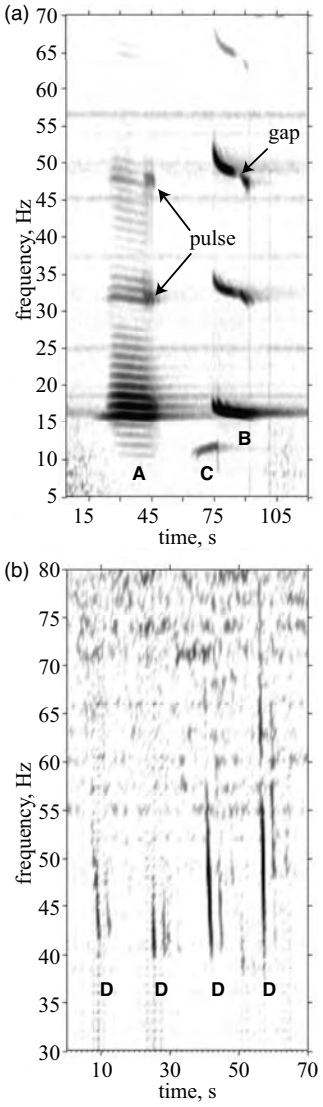

FIG. 3. (a) Spectrogram of an ACB sequence of calls constructed from a stack of 66 individual spectrograms of such calls that occurred after 0800 (GMT) on November 23 (whale 1). The A call begins with ~5–8 s of un-modulated 16 Hz signal that is not always readily apparent in distant noisy recordings and is followed by a train of amplitude modulated short pulses with a fundamental carrier frequency of 16 Hz and at least two harmonics at 32 and 48 Hz. Each pulse includes multiple frequency-offset non-harmonic components. The pulses are not obvious in the spectrogram, but they can be seen in the time series data (Fig. 4). The A call is slightly down-swept in frequency and the 32 and 48 Hz overtones of the A call often terminate with a short 2–3 s pulse, which is likewise smeared in time by the stacking. The low amplitude precursor to the B call, denoted a C call, consists of an upward sweeping call from ~10.5–11.5 Hz. The B call is characterized by a fundamental downward-swept (frequency-modulated) sound from ~17 to 15.5 Hz; a second harmonic that sweeps down from ~34 to 31 Hz; a strong third harmonic that sweeps down from ~52 to 46.5 Hz; and a fourth harmonic that sweeps down from ~68.5 to 62 Hz. The downward sweeping B tones exhibit faint, but persistent, "ghosts" that follow the main pulses by ~2–2.5 s. These ghosts are likely caused by acoustic energy that traveled secondary paths to the instruments (first and second water column multiples). Higher frequency components are expected for northeastern Pacific blue whales (e.g., Thompson *et al.*, 1996; McDonald *et al.*, 2001), but not recorded by our instruments. Narrow vertical lines on the spectrogram are internal instrument noise, the horizontal bands are ship traffic noise. (b) Example spectrogram of D calls (not stacked). These calls have a ~1 s duration, down sweeping from 80 Hz or less to about 40 Hz, and only occurred when more than one whale was present. Each direct arrival of the D call is followed by a fainter arrival at a time expected for the first water column multiple of the acoustic energy.

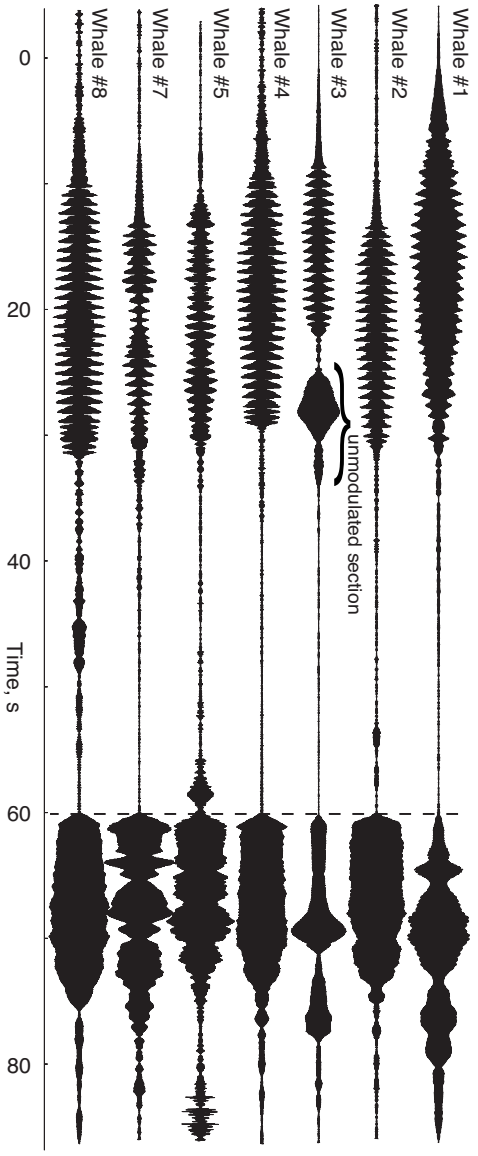

FIG. 4. Time series of the AB calls of blue whales located by this study. These records have been narrow band pass filtered (12–22 Hz) to isolate the fundamental component of the calls. The amplitude modulation is apparent in each of the A calls. The character of the calls changes considerably from call to call due to constructive and destructive interference of the direct path and one or more reflected paths, an effect that changes with the distance of the whale from the receiver and also presumably with the whale's depth (surface and bottom reflected wave interference) and orientation (radiation pattern). However, there are fundamental characteristics of the calls that persist across all calls of a single whale but differ between whales. Specifically, the duration of the A call tends to vary between whales (for example, compare whale 1 to whale 2) and whale 3 exhibits a short modulated A call followed by a separate 10 s un-modulated ~16 Hz call and then the B call. Other distinguishing characteristics may be seen in the overtones (not shown) and in the spectrograms of the calls (Fig. 3).

V. WHALE TRACKING

Over a period of 2.5 days, and after 13 days of on and off airgun activity, up to eight blue whales entered the area of the seismic experiment during continued airgun operations. Figure 2 and Table I summarize basic call detection and tracking information for these whales. While most of the time the whales were located just outside of the seismic array, making detection and tracking difficult, we were able to

TABLE I. Whale tracking summary.

Whale	Start date:time (GMT)	End date:time (GMT)	Number of calls located	Minimum distance traveled (km)	Mean speed over course (km/h)	Min/mean/max distance to airguns (km)
1	Nov 22 1997: 2005	Nov 24 1997: 0053	417	190	6.5	37/62/90
2	Nov 23 1997: 1916	Nov 24 1997: 2136	263	112	4	28/51/97
3	Nov 24 1997: 0200	Nov 24 1997: 1958	287	60	3	15/47/87
4	Nov 24 1997: 2012	Nov 24 1997: 2133	19	4	3	74/78/84
5	Nov 24 1997: 2201	Nov 24 1997: 2342	15	12	7	74/75/78
6	Nov 24 1997: 1910	Nov 24 1997: 2250	6	17	5	61/64/83
7	Nov 24 1997: 1816	⋯	1	⋯	⋯	33/33/33
8	Nov 25 1997: 0045	Nov 25 1997: 1410	68	53	4	72/76/84

identify calls and track some whales over multi-hour intervals. We suggest that individual whales were tracked, rather than multiple whales traveling together because the calls were regularly spaced (i.e., no out-of-sequence calls), not detectably dissimilar, and, perhaps most importantly, did not overlap with other calls emanating from the same location. Having said that, it cannot be ruled out that when one whale stopped vocalizing another whale, located near the same spot, took up where the first left off; or that non-vocalizing whales traveled together with the one vocalizing whale. In one case (whale 3), the A call of the whale is very anomalous as is the B call pattern, providing further support that in that particular case only one individual was tracked. In one or two cases, a whale that was tracked may have been a whale that had been previously identified and tracked over an earlier period of time (based on whale locations and detection times).

A. Blue whale 1

In the final hours of November 22, whale 1 was detected on western stations. By 2005 GMT, as the ship was finishing a seismic line to the north, this whale moved close enough to the seismic stations to be located [Fig. 5(a)]. The whale traveled southeast during the next day, crossing the array. In the final hours of November 23 the airgun activity recommenced southwest of the whale's position. At that point whale 1 continued its easterly heading until 0100 GMT on November 24, when it exited the area to the east and was no longer recorded on sufficient instruments to be located. The distance moved between any two calls is often similar to or smaller than the 1-σ uncertainty of the location, so is difficult to accurately measure the whale's detailed motions and instantaneous velocity. Examining the point-to-point path of the whale, over the 29-h period the whale traveled ~200 km at an average speed of ~6–7 km/h. This is only a rough approximation of the whale's true speed, since the calculated whale track tends to be noisy due to the picking uncertainties; nonetheless it is a typical speed and distance for cruising or migrating whales (Mate *et al.*, 1999). The distance between the whale and the ship during airgun operations was never less than ~37 km (sound levels <145 dB) and there are no detectable changes in the whale's heading nor speed upon the stopping and restarting of the ship's airguns.

We were able to monitor all calls [Fig. 6(a)] from whale 1 over a 24-h period beginning at approximately 0000 GMT on November 23. Throughout this 24-h period, AB calls were repeated semi-regularly; no other call sequence was formed. Repeated sequences of AB calls occur at 135 ± 5 s intervals (all call interval and gap times are measured from the onset of one call to the onset of the next call), with no statistically relevant change from day to night. There are often gaps of both small and large nature that interrupt the repeated AB sequences [Fig. 7(a)]. The majority of gaps are small, between ~160 and 340 s. While it has been suggested that small gaps may represent respiration times (Cummings and Thompson, 1971; Mcdonald *et al.*, 2001), the small gaps in the calls of whale 1 do not repeat at regular intervals and there are 45–60 min intervals when the spacing between calls does not exceed 150 s (or 15 s longer than the main repeat interval), suggesting that a larger call spacing is not required for breathing. Small gaps in the call sequences appear after some of the T-phases of regional earthquakes. Some of these gaps are real, but others may be due to masking of whale calls by the intense broadband earthquake energy. Furthermore, there are other small gaps of this nature in the call pattern that are not preceded by T-phases and several T-phase recordings not followed by gaps. Therefore, there is no obvious correlation between gaps in the call sequences and earthquake T-phases. The *largest* gap in the sequences is ~40 min and the two largest gaps containing no more than one or two AB calls correspond to the time intervals of approximately 1645–1745 and 1900–2015 GMT (0945–1045 and 1200–1315 local time, respectively) on November 23.

Near the end of the 24-h period, the airgun activity recommenced within the seismic array when the whale was located ~90 km from the airgun source. At that distance, sound pressure levels from the airguns are expected to be relatively low (Tolstoy *et al.*, 2004; Diebold, 2004), but seismic instruments near the whale did record the airgun pulses and it is conceivable that the whale detected them as well (airgun pulses occur within the vocalization band of blue whales). There is a small gap just after the first airgun pulses and a larger gap of ~20 min after the airguns had been powered up to full volume. After that, the AB pattern repeats as usual. While the correlation of the call gaps with the airgun activity is of interest, it is not possible to make any causative judgments about these gaps, since similar gaps are

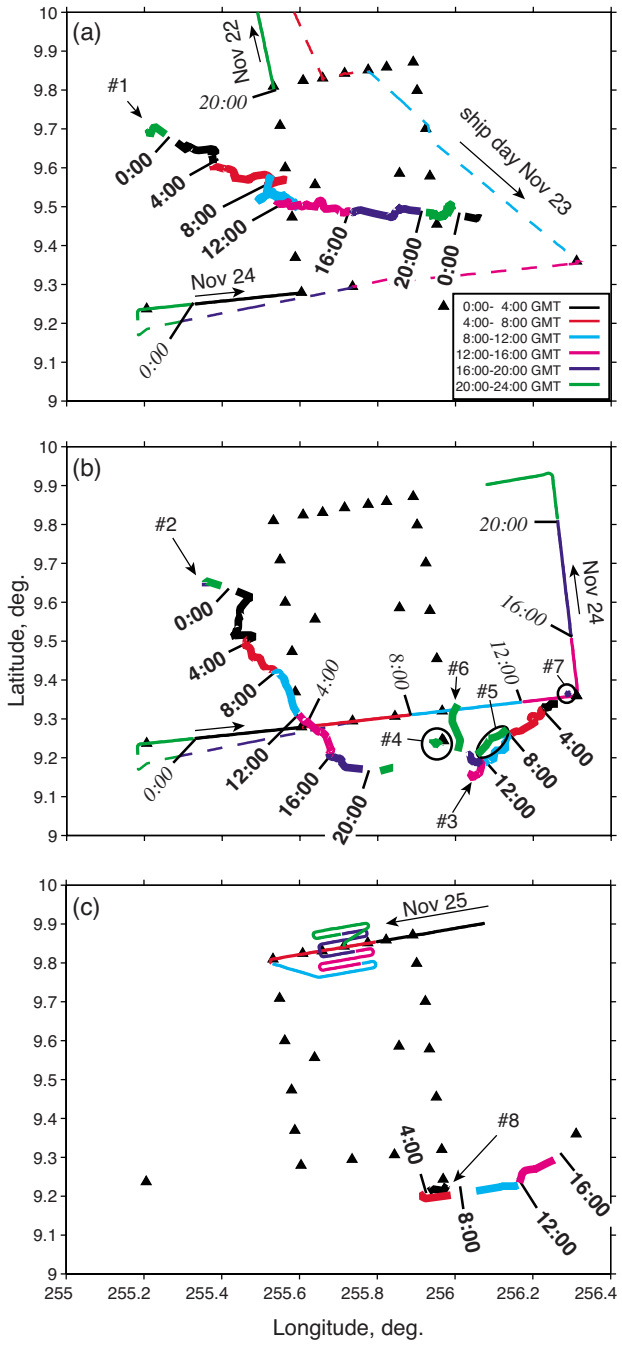

FIG. 5. (Color online) Whale swim tracks for (a) November 23, (b) November 24, and (c) November 25. The number of the whale given on the figure corresponds with the numbers used in the text. Thin dashed (no airgun activity) and solid (airgun activity) lines indicate the ship tracks. Time marks (GMT) are indicated on the tracks in bold (whale times) and italics (ship time). The whale and ship tracks are also color coded, as indicated in panel (a), by 4-h periods.

found throughout the 24-h period. Using high signal-to-noise ratio calls recorded on the station closest to the whale, we compared mean amplitudes of whale calls that occurred 3 h before and 3 h after the airguns restarted and found no statically meaningful difference (i.e., using a 95% confidence interval, we cannot reject the null hypothesis that the amplitudes before and after airgun startup are the same). A similar analysis for when the airguns shutdown earlier in the day has poor resolution because whale 1 was located far from any stations at that point and the calls all have low signal-to-noise ratios; in any case, there was no obvious change in call amplitude. We also compared the mean time intervals of the calls both before and after airgun startup and likewise found no meaningful difference.

Over its entire swim track, whale 1 may have traveled alone, as no overlapping or out-of-sequence calls were recorded that would indicate an accompanying whale or whales. Some of the outlying stations did record more distant whales. The most prominent examples being whale 2 who moved into the area of the study from the west at the end of the 24-h period as whale 1 exited the area to the east, and at least two whales in the vicinity of ORB03 (far eastern region) after 1300 GMT. On ORB03, repeated patterns of AB, ABB, ABBB, ABBBB, and even one clear ABBBBB were recorded, many of them overlapping in time with each other and/or the AB calls of whale 1. We were able to track one of these whales, whale 3, as it moved nearer to the main array on November 24.

B. Blue whale 2

On November 23 at 1900 GMT blue whale 2 approached close enough to the array from the west to be recorded on multiple instruments. At that time its location was determined to be near the location where whale 1 had entered the area 24-h previously [Fig. 5(b)]. Thereafter, whale 2 traveled southeast at ~4 km/h on average until ~2000 GMT at which time it turned east and subsequently stopped vocalizing (last known call occurs at 2136 GMT). The first few locatable calls from this whale occurred just before airgun startup on November 23, thereafter all monitoring of this whale occurred during airgun activities. An analysis of call amplitude changes, upon startup of the airguns, could not be made because the whale was located far from any stations at that point and the calls have low signal-to-noise ratios (scatter in the amplitudes are too large to allow for a meaningful test). Later in the day, the closest distance between the whale and ship was 28 km. There is no indication that the whale tended to avoid the ship, but rather it assumed a heading that crossed the ship's path (aft of the ship at 65 km distance).

We were able to monitor all calls from whale 2 over a 20-h period beginning at approximately 0000 GMT on November 24. Throughout the 20-h period, AB calls [Fig. 6(b)] were repeated semi-regularly. Repeated sequences of AB calls are predominantly spaced at 129 ± 5 s time intervals [Fig. 7(b)], with irregularly occurring gaps that are twice (260 s) and three times (390s) the fundamental spacing. The character of the gaps is thus much different from that of whale 1 (and other whales in this study). During this 20-h period, whale 2 may have traveled alone, as no other calls (overlapping or out of sequence) were recorded that would indicate an accompanying whale or whales. Near the end of the day of November 24, when vocalizations ceased, whale 2 was last detected approaching the positions of at least three other whales in the southeast corner of the study area.

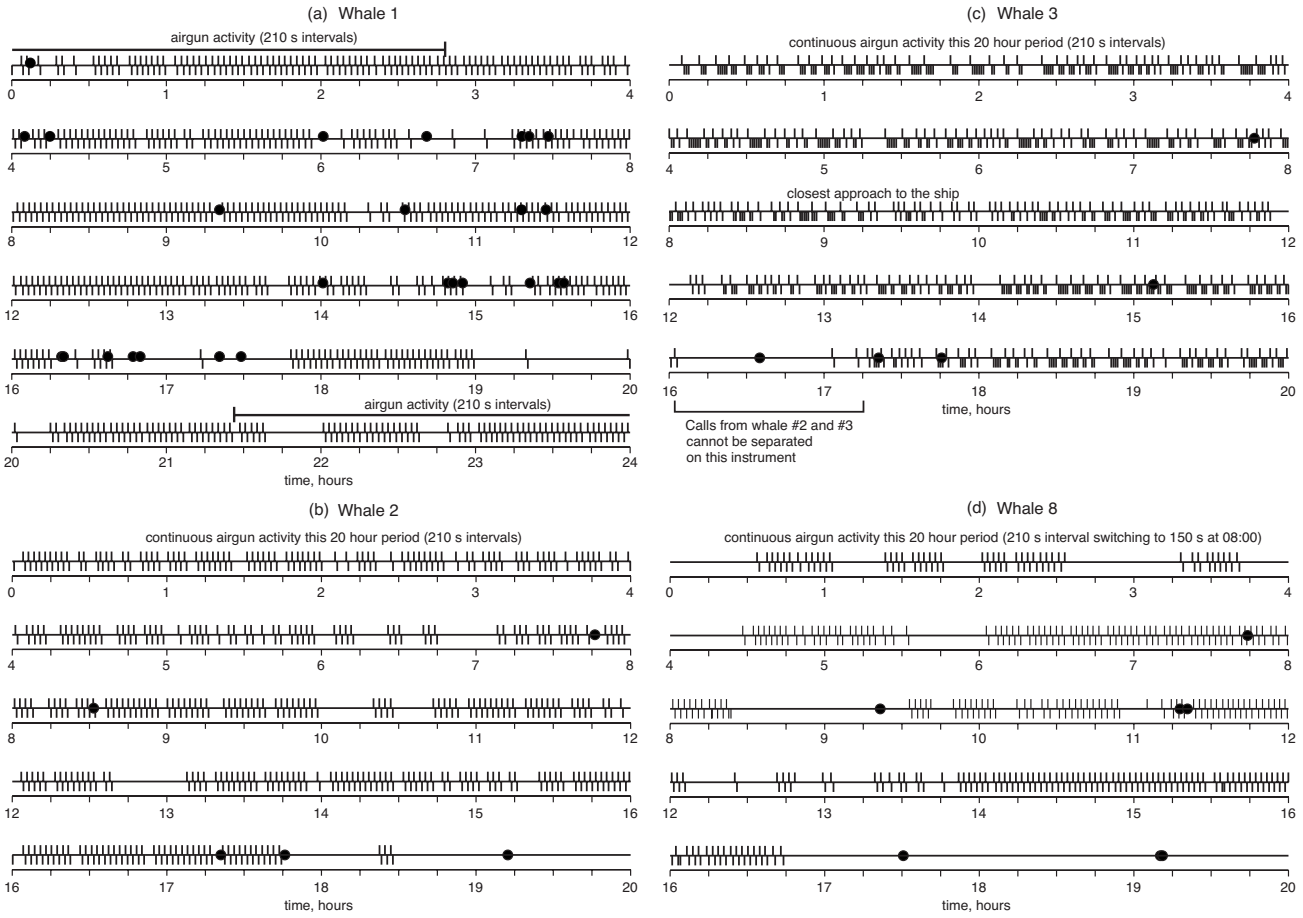

FIG. 6. Call sequences of whales 1, 2, 3, and 8 [subplots (a)–(d), respectively] for 20- to 24-h periods. Ticks above the centerline indicate A calls, and ticks below the center line indicate B calls. C and D calls are not shown. Calls were determined by using seafloor instruments that were located closest to the whale (as indicated on the plots) as well as additional instruments when calls from other whales or airgun pulses masked calls. The black dots indicate earthquake T-phase events. Whales 1, 2, and 8 vocalizations consist almost exclusively of repeated AB sequences. Whale 3 vocalizations consist of AB and A multiple-B sequences in seemingly random order. For whale 1, during the first 30 min of this 24-h period a few AB calls may have been missed, due to a low signal-to-noise ratio. Periods of airgun activity are indicated on each plot.

C. Blue whale 3

On November 24 at 0100 GMT whale 3 approached close enough to the array from the east to be recorded on multiple instruments and its location was determined [Fig. 5(b)]. This whale entered the area near where whale 1 exited the area. We know that whales 1 and 3 are distinct, because the calls of the two whales overlapped during the early hours of November 24. Furthermore, whale 3 exhibits an anomalous A call (Fig. 4) with an A-multiple-B pattern, in contrast to the generic AB calls of whales 1 and 2. The anomalous A call and distinct call sequences also help identify this whale later on November 24 when other AB whale calls (whales 2, 4, 5, and 7) were present in the data.

We reconstructed the entire call sequence for whale 3 over the period 0000 GMT to 2000 GMT on November 24 [Fig. 6(c)]. No obvious groupings or patterns occur within the sequences of calls. For example, there is no apparent pattern to the number of B calls that follow the A calls. Over the 20-h of recordings, the largest break in the sequences is only ~16 min [the large gap at 1600 GMT in Fig. 6(c) is due to masking of calls by another whale] and there are no obvious diurnal changes in the calls.

During the entire period that we tracked whale 3, the ship carried out airgun activities. Whale 3 initially traveled southwestward, passing within ~15 km of the ship, which was heading in the opposite direction; several hours later it turned north and then ceased vocalization. As it passed the ship, there was no obvious heading change that could be construed as an attempt to avoid the ship, such as reported by McDonald *et al.* (1995) for a blue whale approaching a seismic ship to within ~10 km. At 15 km distance from the ship the received sound levels are <145 dB (re 1 μPa rms) (Tolstoy *et al.* 2004; Diebold, 2004), less than what is expected to elicit a behavior response in some marine mammals (NMFS, 2005).

Beginning at about 1000 GMT the calls of whale 2 begin to overlap with whale 3's calls on stations near whale 3; whale 2 was located ~60 km to the northeast at that time. Later, around 1800 GMT, a third whale (whale 6) is detected by its repeated D calls that overlap the AB calls of whale 3; that whale's position was initially determined to be ~20 km to the north of whale 3. At ~2000 GMT whale 4 was suddenly vocally active at a position ~9 km to the north of

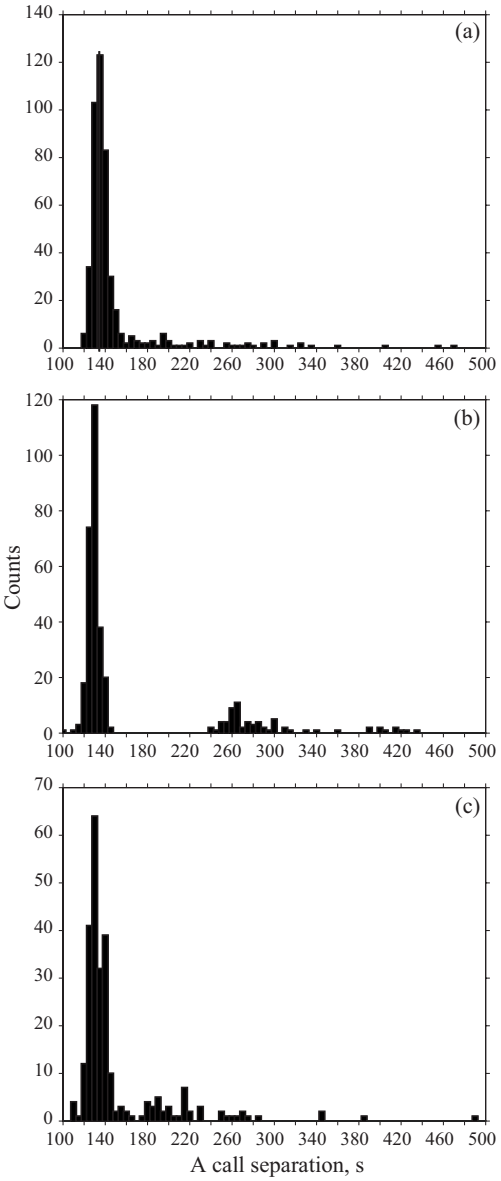

FIG. 7. Histograms (5 s bins) showing the time gap distribution between successive A calls for whales 1, 2, and 8 [subplots (a)–(c), respectively]. Data are for the calls shown in Fig. 6; a time difference was determined as the time between the onset of each A call. The histograms show that the most frequent spacing between calls is 130–135 s. While larger time gaps tend to be random, whale 2 (b) exhibits time gaps of ~260 and ~390 s, which are two and three times the size of the fundamental gap at 130 s. For clarity, time gaps greater than 500 s are not shown. Some of the scatter in the histograms is due to errors in identifying the onset of the A call.

whale 3. Once whale 4 became vocally active, whale 3 was suddenly quiet and its anomalous-A-multiple-B pattern was not detected again.

D. Blues whales 4, 5, 6, and 7

On November 24 whale 4 became vocally active at ~2000 GMT and whale 5 became vocally active after 2145 GMT. Both whales had moved silently into the southeast region of the study, but one of them may have been whale 1 whose last known position was just to the north. We were able to track their calls for a few hours, during which time whale 4 remained in one area and whale 5 moved southwestward [Fig. 5(b)]. Whales 4 and 5 exhibited AB call sequences; neither whale exhibited the anomalous A call of whale 3. During the time we were able to locate them, the ship (with airguns in use) operated ~60 km north of the whales' positions. Calls of whale 2 overlap in time with those of whale 4, indicating that they are distinct whales. Calls of whale 2 do not overlap with those of whale 5, which began vocalizing identifiable calls only 10 min after the last recognizable call of whale 2, but there may have been some earlier calls of whale 5 that were masked by those of whales 2 and 4. 36 km separates whale 2 from whale 5, indicating that these were also distinct whales.

D calls (whale 6) were recorded from 1800 GMT November 24 to 0700 GMT November 25 on stations in the southeast quadrant of the study. These calls overlap with those of whales 2, 3, 4, 5, and 8 on those stations, indicating the presence of another distinct whale. Owing to a poor signal-to-noise ratio, the swim track could only be determined for a short period between 1900 and 2300 GMT on November 24 as the whale moved southward.

For a short time earlier in the day (1500–2000 GMT) another blue whale, whale 7, was recorded near ORB3. The ship may have passed closely by whale 7, but since we detected and located it several hours after the ship had left the area, the proximity between the ship and whale cannot be established. The difficultly in locating this whale was caused by a low signal-to-noise ratio due to its position well outside of the main array, noise from the airgun source, and a persistent masking of its calls by those of whale 3.

E. Blue whale 8

In the early hours of November 25, at least two and possibly three blue whales were recorded on stations near ORB02 in the southeast quadrant of the study area. These calls likely were made by whales 2, 4, and 5, who were last detected in this general vicinity, but probably not whale 3, since the calls were purely of AB combinations and did not exhibit the anomalous A call of whale 3. Due to the overlapping nature of the calls and weak signals, it was not possible to locate each of these whales, but one whale did have a strong signal-to-noise ratio and we were able to track it for many hours as it moved from ORB02 eastward out of the array at a speed of ~5 km/h [Fig. 5(c)]. Although we designated it whale 8, it may be one of the whales previously identified. Throughout the day, the ship performed airgun work 72–84 km to the north of this whale.

We were able to monitor all calls from whale 8 over a 20-h period beginning at approximately 0000 GMT on November 25. Throughout the 20-h period, AB calls [Fig. 6(d)] were repeated semi-regularly; except for the rare ABB call or A-only call, no other call sequence was formed. Repeated sequences of AB calls are spaced at 130 ± 6 s time intervals, but like other whales there are larger gaps of random size [Fig. 7(c)]. During the 20-h period, whale 8 may have traveled alone, as no overlapping or out-of-sequence calls from the same location were recorded that would indicate an accompanying whale or whales. Near the end of the day of November 25, whale 8 approached ORB03 at the eastern-

most extreme of the study area. At least two other whales (one of which may have been whale 7) were vocalizing in the area (probably within 10 km to the east of ORB03 based on call amplitudes). We lost track of whale 8 as its calls overlapped with those of the other whales and as it moved far enough away from the main array to no longer be recorded on enough stations to determine its location. For the remainder of the day and throughout the next day, its calls and the calls of at least one other whale were recorded (faintly) on ORB03, indicating that these whales remained within ~40 km of the eastern edge of the experiment area.

VI. DISCUSSION AND CONCLUSIONS

The blue whales detected by this study are members of a vocally distinct population of blue whales that inhabit the northeast Pacific, ranging from the Gulf of Alaska to a region off Central America (e.g., Stafford *et al.*, 2001; Stafford, 2003; McDonald *et al.*, 2006). Individuals found in the Gulf of California are also thought to be part of this group (Calambokidis *et al.*, 1990; Thompson *et al.*, 1996). Although details of the migration routes and numbers of whales that migrate are poor, a hydrophone study detected members of the northeastern Pacific blue whale population year round in the eastern tropical Pacific (Stafford *et al.*, 1999, 2001). Therefore, it is not unexpected that the blue whales detected by our study in the month of November are members of the northeastern Pacific population.

Six out of eight whales formed closely spaced, repeated AB call sequences; the exceptions are whale 3 who formed sequences of A calls followed by up to six B calls and whale 6 who was tracked by its D calls. While we cannot be assured that an individual whale's call behavior will not change over time, some whales in this study did exhibit anomalous call components that repeated with each call: for example, the gap in the B call of whale 1 [Fig. 3(a)] and the anomalous A call of whale 3 (Fig. 4). Thode *et al.* (2000) also detected distinguishing characteristics of blue whale B calls that allowed them to identify individual whales. This suggests that in the future it may be possible to track some individual whales via fixed hydrophone arrays call-by-call for extended periods of time, even in the presence of other blue whales. Other than the AB call patterns that tended to be spaced every 130–135 s, we found no other regular call patterns or repetitions other than a tendency for whale 2 to exhibit a gap between AB calls that were two and three times its fundamental call spacing [Fig. 7(b)]. While there appears to be longer gaps in the latter parts of the 24-h periods of observation, this may be due to the presence of other whales at those times or foraging behavior (see below) rather than diurnal behavioral variations.

Four of the whales exhibited long (20–26 h or more) repetitive AB vocalization sequences (whales 1, 2, 3, and 8). A recent study by Oleson *et al.* (2007) indicates that repetitive AB calling sequences are characteristic of lone, migrating males, rather than foraging whales or whales in groups. We were able to reconstruct the swim tracks and time series of almost every call made by these whales as they passed through the area. Although the determination of detailed whale movements and dive depths was not possible, we were able to obtain general locations with an uncertainty of 1–2 km. The whales tended to travel alone or at least without other vocalizing whales; for hours on end there were no overlapping or out-of-sequence calls from closely spaced whales. The calls that are present tend to be evenly spaced, with an interval time typical of an individual whale. The whales also tended to travel long distances, 100 km or more over a 24-h period. Average swim speeds were ~3–7 km/h over the course of monitoring. These distances and swim speeds may be typical of blue whales that are migrating or cruising, but not foraging (Mate *et al.*, 1999). In summary, the long swim tracks of these four whales and their AB calling behavior indicate that these individuals were lone, migrating males.

On November 24, while airgun activity continued ~60–80 km away, several whales congregated in the southeast corner of the study area near ORB02 (whales 2, 3, 4, 5, 6, and 8). Such clustering behavior is indicative of foraging (Mate *et al.*, 1999). The presence of other whales also had an obvious correlation with changes in calling behavior: mainly a cessation of calling or long pauses between calls. For example, (1) as whales 2 and 3 moved into this area they ceased vocalizations, (2) whales 4 and 5 were vocally active for only short periods of time in this area, and (3) whale 8 became vocally active only as it left this area. Also notable is whale 6 (possibly the previously identified whale 1), who passed through this area while producing only D calls. In the study of Oleson *et al.* (2007), D calls were heard from both sexes during foraging, commonly from individuals within groups. Our observations, taken together with those from the study of Oleson *et al.* (2007), suggest that lone traveling males moved into this area, subsequently ceased most AB call sequences, and perhaps spent some time foraging, whether or not females were present is unknown.

For whales 1 and 2, the instruments recorded calls both during airgun activity and between airgun activity (Fig. 2). At times of starting or stopping airgun activity, these whales were located tens of kilometers from the airgun source (whale 1: 69 km at airgun shutdown and 90 km at airgun startup; whale 2: 42 km at airgun startup) and we did not detect corresponding changes in swim tracks or call behavior. For whale 1, a 20-min gap in calls occurred after the airguns became active, but many gaps occurred in the call sequences throughout the day—both during and not during airgun activity—so no causative relationship is supported. There is no indication that the whales attempted to time calls to fall between airgun pulses. The AB calls are generally spaced every 130–135 s, while the airgun pulses mainly occurred every 210 s so that the calls moved in and out of the spaces between airgun pulses. We also examined the call sequences for any anomalous behavior due to the presence of earthquake acoustic energy. Earthquakes produce significant water column energy in the frequency band used by blue whales and can mask whale calls for tens of seconds, but we found no obvious correlation between the many earthquake events that occurred during the monitoring and changes in calling behavior (changes in duration and timing). Presum-

ably, blue whales are accustomed to such high-amplitude sounds, as they occur frequently along mid-ocean ridges.

During airgun operations, airgun pulses were recorded across the entire seismic array and were thus presumably detectable by all eight whales. Overall we found no anomalous behavior that could be directly ascribed to the use of the airguns, though it should be reemphasized that the average distance from airgun source to the whales was tens of kilometers (Table I). For whale 3, who approached the ship to within about 15 ± 2 km (the closest of any whale), the call patterns and the whale's heading exhibit no detectable changes. Since the whales were not closer than ~ 15 km to the ship, and usually much farther away, sound levels produced by the Ewing's airguns and experienced by the whales are expected to be less than 145 dB (re 1 μPa). Under current guidelines, the National Marine Fisheries Services defines the radius about the ship with received sound levels of 160 dB as distances within which some cetaceans are likely to be subject to behavioral disturbance (NMFS, 2005). While this study found no behavioral response to the airgun activity, and hence supports these guidelines, further studies with more detailed observations are warranted.

ACKNOWLEDGMENTS

This research was partially supported by the National Science Foundation, Ocean Sciences Division, under Grant No. OCE0224903. We thank John Diebold for the theoretical estimates of the Ewing airgun array output levels and M. Carolina Anchieta for the earthquake locations shown in Fig. 1; Carolina located these events during an undergraduate summer internship at UH. Olga Hernandez contributed to this work during an internship at the University of Hawaii (that formed part of her Prédoctorat Programme at the Ecole Normale Supérieure de Paris). We also thank two anonymous reviewers for their careful consideration of the manuscript.

Aroyan, J. L., McDonald, M. A., Webb, S. C., Hildebrand, J. A., Clark, D., Laitman, J. T., and Reidenberg, J. S. (**2000**). "Acoustic models of sound production and propagation," in *Hearing by Whales and Dolphins*, edited by W. W. L. Au, A. N. Popper, and R. R. Fay (Springer-Verlag, New York), Vol. **12**, pp. 409–469.

Calambokidis, J., Steiger, G. H., Cubbage, J. C., Balcomb, K. C., Ewald, C., Kruse, S., Wells, R., and Sears, R. (**1990**). "Sightings and movements of blue whales off central California 1986–1988 from photo-identification of individuals," Rep. Int. Whal. Comm. **12**, 343–348.

Cummings, W. C., and Thompson, P. O. (**1971**). "Underwater sounds from the blue whale, *Balaenoptera musculus*," J. Acoust. Soc. Am. **50**, 1193–1198.

Diebold, J. (**2004**). "Modeling marine seismic source arrays," Lamont Doherty Earth Observatory, http://www.ldeo.columbia.edu/res/fac/oma/sss/NMFS_Lamont_airgun+modeling.pdf (Last viewed January 2009).

Dunn, R. A., Toomey, D. R., Detrick, R. S., and Wilcock, W. S. D. (**2001**). "Continuous mantle melt supply beneath an overlapping spreading center on the East Pacific Rise," Science **291**, 1955–1958.

Lagerquist, B. A., Stafford, K. M., and Mate, B. R. (**2000**). "Dive characteristics of satellite-monitored blue whales (*Balaenoptera musculus*) off the central California Coast," Marine Mammal Sci. **16**, 375–391.

Malakoff, D. A. (**2001**). "Roaring debate over ocean noise," Science **291**, 576–578.

Malakoff, D. A. (**2002**). "Suit ties whale death to research cruise," Science **298**, 722–723.

Mate, B. R., Lagerquist, B. A., and Calambokidis, J. (**1999**). "Movements of North Pacific blue whales during the feeding season off Southern California and their southern fall migration," Marine Mammal Sci. **15**, 1246–1257.

McDonald, M. A., Calambokidis, J., Teranishi, A. M., and Hildebrand, J. A. (**2001**). "The acoustic calls of blue whales off California with gender data," J. Acoust. Soc. Am. **109**, 1728–1735.

McDonald, M. A., Hildebrand, J. A., and Webb, S. C. (**1995**). "Blue and fin whales observed on a seafloor array in the Northeast Pacific," J. Acoust. Soc. Am. **98**, 712–721.

McDonald, M. A., Mesnick, S. L., and Hildebrand, J. A. (**2006**). "Biogeographic characterization of blue whale song worldwide: Using song to identify populations," J. Cetacean Res. Manage. **8**, 55–65.

National Research Council (**2003**). *Ocean Noise and Marine Mammals* (National Academy Press, Washington, DC).

NMFS (**2005**). "Small takes of marine mammals incidental to specified activities: Marine seismic survey of the Aleutian Islands in the North Pacific Ocean/notice of issuance of an incidental take authorization," Fed. Regist. **70**, 901–913.

Nosal, E.-M., and Frazer, L. N. (**2006**). "Track of a sperm whale from delays between direct and surface-reflected clicks," Appl. Acoust. **67**, 1187–1201.

Oleson, E. M., Calambokidis, J., Burgess, W. C., McDonald, M. A., LeDuc, C. A., and Hildebrand, J. A. (**2007**). "Behavioral context of call production by eastern North Pacific blue whales," Mar. Ecol.: Prog. Ser. **330**, 269–284.

Širović, A., Hildebrand, J. A., and Wiggins, S. M. (**2007**). "Blue and fin whale call source levels and propagation range in the Southern Ocean," J. Acoust. Soc. Am. **122**, 1208–1215.

Stafford, K. M. (**2003**). "Two types of blue whale calls recorded in the gulf of Alaska," Marine Mammal Sci. **19**, 682–693.

Stafford, K. M., Nieukirk, S. L., and Fox, C. G. (**1999**). "An acoustic link between blue whales in the Eastern Tropical Pacific and the Northeast Pacific," Marine Mammal Sci. **15**, 1258–1268.

Stafford, K. M., Nieukirk, S. L., and Fox, C. G. (**2001**). "Geographic and seasonal variation of blue whale calls in the North Pacific," J. Cetacean Res. Manage. **3**, 65–76.

Tarantola, A., and Valette, B. (**1982**). "Inverse problems=quest for information," J. Geophys. **50**, 159–170.

Thode, A. M., D'Spain, G. L., and Kuperman, W. A. (**2000**). "Matched-field processing, geoacoustic inversion, and source signature recovery of blue whale vocalizations," J. Acoust. Soc. Am. **107**, 1286–1300.

Thompson, P. O., Findley, L. T., Vidal, O., and Cummings, W. C. (**1996**). "Underwater sounds of blue whales, *Balaenoptera musculus*, in the Gulf of California, Mexico," Marine Mammal Sci. **12**, 288–293.

Tolstoy, M., Diebold, J. B., Webb, S. C., Bohnenstiehl, D. R., Chapp, E., Holmes, R. C., and Rawson, M. (**2004**). "Broadband calibration of R/V *Ewing* seismic sources," Geophys. Res. Lett. **31**, L14310.

Watkins, W. W., Daher, M. A., George, J. E., and Rodriguez, D. (**2004**). "Twelve years of tracking 52-Hz whale calls from a unique source in the North Pacific," Deep Sea Res. Part I **51**, 1889–1901.

Yochem, P. K., and Leatherwood, S. (**1985**). "Blue whale—Balaenoptera musculus (Linnaeus, 1758)," in *Handbook of Marine Mammals*, edited by S. H. Ridgeway and R. Harrison (Academic, London), Vol. **3**, pp. 193–240.

Appendix D

Correlation coefficient and standard deviation uncertainties

Standard deviation uncertainties of estimated parameters and correlation coefficient between optimal parameters obtained by using the best spawning habitat definition and the climatological run, are presented in Tables D.1 and D.2 for anchovy and Tables D.3 and D.4 for sardine, respectively.

	μ	ϵ	a	b	c	d	R	$\beta_{e,1}$	$\beta_{l,1}$
μ	1	**-0.91**	-0.078	0.098	0.763	**0.839**	-0.732	-0.023	0.003
ϵ	**-0.91**	1	0.027	-0.399	-0.655	-0.596	0.458	-0.004	0.015
a	-0.078	0.027	1	0.236	0.102	-0.007	0.081	0.037	-0.032
b	0.098	-0.399	0.236	1	-0.126	-0.222	0.337	0.082	-0.066
c	0.763	-0.655	0.102	-0.126	1	0.797	-0.673	-0.043	0.021
d	**0.839**	-0.596	-0.007	-0.222	0.797	1	**-0.971**	-0.046	0.02
R	-0.732	0.458	0.081	0.337	-0.673	**-0.971**	1	0.052	-0.026
$\beta_{e,1}$	-0.023	-0.004	0.037	0.082	-0.043	-0.046	0.052	1	-0.006
$\beta_{l,1}$	0.003	0.015	-0.032	-0.066	0.021	0.02	-0.026	-0.006	1

Table D.1: Correlation coefficient between optimal parameters obtained for Hs3 experiment in the anchovy case.

	μ	ϵ	a	b	c	d	R	$\beta_{e,1}$	$\beta_{l,1}$
std	0.058	0.839	0.029	0.044	0.019	0.213	0.074	0.003	0.004

Table D.2: Standard deviation uncertainties of estimated parameters obtained for Hs3 experiment in the anchovy case.

	μ	ϵ	σ_0	T_0	a	b	c	d	R	$\beta_{e,1}$	$\beta_{l,1}$
μ	1	-0.452	0.094	-0.295	-0.26	-0.058	0.287	**0.747**	-0.524	-0.024	0.004
ϵ	-0.452	1	0.091	0.134	-0.112	-0.081	-0.341	-0.125	0.043	-0.003	0.044
σ_0	0.094	0.091	1	0.562	-0.018	-0.047	0.049	0.114	-0.12	0.018	0.025
T_0	-0.295	0.134	0.562	1	0.18	-0.182	0.256	-0.293	0.449	-0.02	0.045
a	-0.26	-0.112	-0.018	0.18	1	0.266	0.209	-0.102	0.127	0.063	-0.059
b	-0.058	-0.081	-0.047	-0.182	0.266	1	-0.301	0.045	-0.022	0.034	-0.036
c	0.287	-0.341	0.049	0.256	0.209	-0.301	1	0.385	-0.214	-0.018	0.007
d	**0.747**	-0.125	0.114	-0.293	-0.102	0.045	0.385	1	**-0.886**	-0.007	0.004
R	-0.524	0.043	-0.12	0.449	0.127	-0.022	-0.214	**-0.886**	1	-0.016	0.013
$\beta_{e,1}$	-0.024	-0.003	0.018	-0.02	0.063	0.034	-0.018	-0.007	-0.016	1	-0.007
$\beta_{l,1}$	0.004	0.044	0.025	0.045	-0.059	-0.036	0.007	0.004	0.013	-0.007	1

Table D.3: Correlation coefficient between optimal parameters obtained for Hs5 experiment in the sardine case.

	μ	ϵ	σ_0	T_0	a	b	c	d	R	$\beta_{e,1}$	$\beta_{l,1}$
std	0.041	0.356	0.236	0.093	0.073	0.03	0.032	0.098	0.046	0.003	0.004

Table D.4: Standard deviation uncertainties of estimated parameters obtained for Hs5 experiment in the sardine case.

Correlation coefficient between optimal parameters obtained by using the Hs5 spawning habitat definition and the interannual run 1992-2000 are presented in Table D.5 for anchovy and D.6 for sardine.

	μ	ϵ	σ_0	T_0	a	b	c	d	R	$q_{l,1}$	$q_{l,2}$	$\beta_{e,1}$	$\beta_{l,1}$	$\beta_{e,2}$	$\beta_{l,2}$
μ	1	-0.138	**0.625**	-0.048	0.036	**-0.772**	0.002	0	-0.374	-0.198	0.363	0.003	-0.029	0.018	-0.005
ϵ	-0.138	1	0.221	0.198	0.206	-0.087	0.001	0	-0.096	0.008	-0.101	0.004	0.001	0.011	0.042
σ_0	**0.625**	0.221	1	0.463	0.242	-0.496	0.001	0	-0.443	-0.013	0.236	-0.015	0.015	0.032	-0.021
T_0	-0.048	0.198	0.463	1	-0.051	0.003	0	0.001	0.248	-0.196	-0.352	-0.028	0.006	0.019	0.011
a	0.036	0.206	0.242	-0.051	1	0.255	0	0	0.185	-0.162	0.03	0.036	-0.009	0.017	-0.033
b	**-0.772**	-0.087	-0.496	0.003	0.255	1	-0.001	0.002	**0.629**	0.031	-0.194	-0.002	0.029	-0.046	0.021
c	0.002	0.001	0.001	0	0	-0.001	1	-0.019	-0.001	0.001	0	0	0	0	0
d	0	0	0	0.001	0	0.002	-0.019	1	0.002	-0.004	-0.003	0	0	0	0
R	-0.374	-0.096	-0.443	0.248	0.185	**0.629**	-0.001	0.002	1	-0.488	-0.53	0.006	-0.03	-0.034	0.052
$q_{l,1}$	-0.198	0.008	-0.013	-0.196	-0.162	0.031	0.001	-0.004	-0.488	1	0.298	-0.01	0.042	0.001	-0.03
$q_{l,2}$	0.363	-0.101	0.236	-0.352	0.03	-0.194	0	-0.003	-0.53	0.298	1	-0.001	0.025	-0.01	-0.018
$\beta_{e,1}$	0.003	0.004	-0.015	-0.028	0.036	-0.002	0	0	0.006	-0.01	-0.001	1	-0.002	0.001	-0.002
$\beta_{l,1}$	-0.029	0.001	0.015	0.006	-0.009	0.029	0	0	-0.03	0.042	0.025	-0.002	1	-0.001	-0.001
$\beta_{e,2}$	0.018	0.011	0.032	0.019	0.017	-0.046	0	0	-0.034	0.001	-0.01	0.001	-0.001	1	-0.006
$\beta_{l,2}$	-0.005	0.042	-0.021	0.011	-0.033	0.021	0	0	0.052	-0.03	-0.018	-0.002	-0.001	-0.006	1

Table D.5: Correlation coefficient between optimal parameters obtained for anchovy for the interannual run 1992-2000.

	μ	ϵ	σ_0	T_0	a	b	c	d	R	$\beta_{e,1}$	$\beta_{l,1}$
μ	1	-0.016	-0.003	0.001	-0.002	-0.002	0.006	-0.004	0.009	-0.004	0
ϵ	-0.016	1	0	0	0	0	0	0	0	0	0
σ_0	-0.003	0	1	**0.682**	0.033	0.143	-0.141	0.143	-0.22	0.011	-0.016
T_0	0.001	0	**0.682**	1	0.014	-0.035	0.314	-0.303	0.336	-0.044	0.065
a	-0.002	0	0.033	0.014	1	0.61	-0.094	0.079	-0.021	0.012	-0.015
b	-0.002	0	0.143	-0.035	0.61	1	-0.419	0.352	-0.298	0.021	-0.042
c	0.006	0	-0.141	0.314	-0.094	-0.419	1	-0.385	0.518	-0.098	0.106
d	-0.004	0	0.143	-0.303	0.079	0.352	-0.385	1	**-0.969**	0.041	-0.119
R	0.009	0	-0.22	0.336	-0.021	-0.298	0.518	**-0.969**	1	-0.066	0.135
$\beta_{e,1}$	-0.004	0	0.011	-0.044	0.012	0.021	-0.098	0.041	-0.066	1	-0.016
$\beta_{l,1}$	0	0	-0.016	0.065	-0.015	-0.042	0.106	-0.119	0.135	-0.016	1

Table D.6: Correlation coefficient between optimal parameters obtained for sardine for the interannual run 1992-2000.

Correlation coefficient between optimal parameters obtained by using the Hs5 spawning habitat definition and the interannual run 2000-2006 are presented in Table D.7 for anchovy.

	μ	ϵ	σ_0	T_0	a	b	c	d	R	$q_{l,1}$	$q_{l,2}$	$\beta_{e,1}$	$\beta_{l,1}$	$\beta_{e,2}$	$\beta_{l,2}$
μ	1	**-0.993**	0.164	-0.249	-0.216	0.398	0.586	**0.842**	**-0.803**	-0.105	0.31	0.056	0.026	-0.043	-0.009
ϵ	**-0.993**	1	-0.172	0.251	0.21	-0.408	-0.589	**-0.849**	**0.811**	0.062	-0.279	-0.056	-0.028	0.044	0.009
σ_0	0.164	-0.172	1	0.525	-0.007	0.114	0.037	0.204	-0.212	0.2	0.264	0.006	0.038	-0.018	-0.018
T_0	-0.249	0.251	0.525	1	0.113	-0.037	-0.123	-0.228	0.258	-0.162	-0.3	-0.013	-0.008	0.034	-0.007
a	-0.216	0.21	-0.007	0.113	1	**0.769**	-0.072	-0.218	0.381	0.102	-0.069	-0.057	0.048	-0.024	0.015
b	0.398	-0.408	0.114	-0.037	**0.769**	1	0.262	0.33	-0.15	0.033	0.08	-0.012	0.062	-0.048	0
c	0.586	-0.589	0.037	-0.123	-0.072	0.262	1	**0.828**	**-0.792**	-0.045	0.142	0.028	0.013	-0.022	-0.004
d	**0.842**	**-0.849**	0.204	-0.228	-0.218	0.33	**0.828**	1	**-0.981**	-0.038	0.26	0.049	0.027	-0.044	-0.01
R	**-0.803**	**0.811**	-0.212	0.258	0.381	-0.15	**-0.792**	**-0.981**	1	0.014	-0.283	-0.053	-0.019	0.039	0.011
$q_{l,1}$	-0.105	0.062	0.2	-0.162	0.102	0.033	-0.045	-0.038	0.014	1	0.226	-0.012	0.021	-0.016	0
$q_{l,2}$	0.31	-0.279	0.264	-0.3	-0.069	0.08	0.142	0.26	-0.283	0.226	1	0.011	0.027	-0.031	-0.002
$\beta_{e,1}$	0.056	-0.056	0.006	-0.013	-0.057	-0.012	0.028	0.049	-0.053	-0.012	0.011	1	-0.001	-0.001	-0.002
$\beta_{l,1}$	0.026	-0.028	0.038	-0.008	0.048	0.062	0.013	0.027	-0.019	0.021	0.027	-0.001	1	-0.005	0
$\beta_{e,2}$	-0.043	0.044	-0.018	0.034	-0.024	-0.048	-0.022	-0.044	0.039	-0.016	-0.031	-0.001	-0.005	1	0
$\beta_{l,2}$	-0.009	0.009	-0.018	-0.007	0.015	0	-0.004	-0.01	0.011	0	-0.002	-0.002	0	0	1

Table D.7: Correlation coefficient between optimal parameters obtained for anchovy for the interannual run 2000-2006.

Eulerian modelling of anchovy and sardine spawning habitat and larvae dynamics in the Peruvian upwelling system

The Humboldt Current System is the most productive oceanic system in terms of fish biomass. Its main species, the Peruvian anchovy (*Engraulis ringens*) and sardine (*Sardinops sagax*), have known important historical variations of their stocks, a particularity which has questioned oceanographers for many years. The goal of the present work is to implement an eulerian approach focusing on the early life stages of peruvian anchovy and sardine, in order to study the influence of environmental parameters and climate variability on population dynamics and stocks variability. The model is adapted from the "Spatial Eulerian Ecosystem and Population Dynamic Model" SEAPODYM, initially developed for large pelagic fishes at ocean basin scale. It uses the outputs of a regional ROMS-PISCES coupled physical-biogeochemical model as environmental forcing. In order to constrain model parameters, we developed a data assimilation framework using a unique dataset of eggs and larvae abundances, collected by the Instituto del Mar del Peru (IMARPE) in the HCS waters over the last 40 years. This framework allows to test different mechanisms proposed to control fish spawning habitat and larval recruitment: optimal temperature, prey abundance, trade-off between preys and predators, and influence of currents on retention and dispersion. We show that the combination of a few simple concepts can reasonably explain the overall spatial distribution of eggs and larvae and the interannual variability, but the current description is still not sufficient to explain the observed abundance seasonality in coastal areas.

AUTEUR : Olga HERNANDEZ
TITRE : Modélisation eulérienne de l'habitat de ponte et de la dynamique des larves des anchois et sardines dans le système d'upwelling du Pérou
DIRECTEURS DE THESE : Philippe GASPAR (CLS, Toulouse), Patrick LEHODEY (CLS, Toulouse) et Arnaud BERTRAND (IRD, Sète)
LIEU ET DATE DE SOUTENANCE : Jeudi 21 juin 2012 à Toulouse

RESUME:
Le Système du Courant de Humboldt est l'écosystème le plus productif au monde en termes de biomasse de poisson. Ses espèces principales, l'anchois *Engraulis ringens* et la sardine *Sardinops sagax*, ont connu d'importantes variations historiques de leurs stocks, une particularité dont l'origine est encore débattue. L'objectif de ce travail est d'implémenter pour la première fois une approche eulérienne se focalisant sur les premiers stades de vie des anchois et sardines du Pérou, et visant à étudier l'influence des paramètres environnementaux et de la variabilité climatique sur la dynamique des populations et les variations de leurs stocks. Le modèle est adapté du modèle eulérien de dynamique spatiale des écosystèmes SEAPODYM, initialement développé pour les grands poissons pélagiques à l'échelle du bassin océanique, et utilise le modèle couplé physico-biogéochimique ROMS-PISCES en entrée. Afin de contraindre les paramètres du modèle, nous avons développé une approche par assimilation de données basée sur un jeu unique de données de densités d'oeufs et de larves, acquis par l'Institut de la Mer du Pérou au cours des 40 dernières années. Le modèle permet de tester différents mécanismes susceptibles de contrôler l'habitat de ponte et le recrutement larvaire : température optimale, abondance des proies, présence/absence de prédateurs, et influence des courants sur la rétention et la dispersion. Nous montrons que la combinaison de quelques concepts simples permet d'expliquer raisonnablement la répartition spatiale des oeufs et larves ainsi que la variabilité interannuelle, mais la description actuelle ne permet pas d'expliquer la saisonnalité observée dans les zones côtières.

MOTS-CLES: Anchois, Sardine, *Engraulis ringens*, *Sardinops sagax*, SEAPODYM, modèle de dynamique spatiale des populations, assimilation de données, Estimation du Maximum de Vraisemblance, méthode adjointe, upwelling, Humboldt Current system, Pérou, habitat de ponte

---

DISCIPLINE ADMINISTRATIVE : Océanographie

LABORATOIRE : CLS (Collecte Localisation Satellite)
8-10, rue Hermès, Parc Technologique du Canal; 31520 Ramonville Saint-Agne