THÈSE DE DOCTORAT

Spécialité : Écosystèmes (École Doctorale SIBAGHE - Systèmes intégrés en Biologie, Agronomie, Géosciences, Hydrosciences et Environnement)

Présenté par : Rocío de Abril JOO ARAKAWA

Pour obtenir le grade de : DOCTEUR de l'UNIVERSITÉ DE MONTPELLIER 2

Sujet de la thèse :

A behavioral ecology of fishermen: hidden stories from trajectory data in the Northern Humboldt Current System

Une écologie du comportement des pêcheurs : histoires cachées à partir des données de trajectoires dans le système de Courant de Humboldt

Soutenu le **19 décembre 2013** devant le jury composé de :

Mme. Stéphanie Mahévas	Rapporteur
M. Jan-Jaap Poos	Rapporteur
M. Roger Pradel	Examinateur
M. David Nerini	Examinateur
M. Francisco Chavez	Examinateur (invité)
Mme. Sophie Bertrand	Directeur de thèse
M. Ronan Fablet	Co-directeur de thèse

A los pollitos.
Los que fueron, los que son, los que serán.

Abstract

This work proposes an original contribution to the understanding of fishermen spatial behavior, based on the behavioral ecology and movement ecology paradigms. Through the analysis of Vessel Monitoring System (VMS) data, we characterized the spatial behavior of Peruvian anchovy fishermen at different scales: (1) the behavioral modes within fishing trips (i.e., searching, fishing and cruising); (2) the behavioral patterns among fishing trips; (3) the behavioral patterns by fishing season conditioned by ecosystem scenarios; and (4) the computation of maps of anchovy presence proxy from the spatial patterns of behavioral mode positions. At the first scale considered, we compared several Markovian (hidden Markov and semi-Markov models) and discriminative models (random forests, support vector machines and artificial neural networks) for inferring the behavioral modes associated with VMS tracks. The models were trained under a supervised setting and validated using tracks for which behavioral modes were known (from on-board observers records). Hidden semi-Markov models performed better, and were retained for inferring the behavioral modes on the entire VMS dataset. At the second scale considered, each fishing trip was characterized by several features, including the time spent within each behavioral mode. Using a clustering analysis, fishing trip patterns were classified into groups associated to management zones, fleet segments and skippers' personalities. At the third scale considered, we analyzed how ecological conditions shaped fishermen behavior. By means of co-inertia analyses, we found significant associations between fishermen, anchovy and environmental spatial dynamics, and fishermen behavioral responses were characterized according to contrasted environmental scenarios. At the fourth scale considered, we investigated whether the spatial behavior of fishermen reflected to some extent the spatial distribution of anchovy. Finally, this work provides a wider view of fishermen behavior: fishermen are not only economic agents, but they are also foragers, constrained by ecosystem variability. To conclude, we discuss how these findings may be of importance for fisheries management, collective behavior analyses and end-to-end models.

Key words: vessel monitoring system; tracking data; foraging movement; hidden semi-Markov models; model validation; anchoveta *Engraulis ringens*

Résumé

Ce travail propose une contribution originale à la compréhension du comportement spatial des pêcheurs, basée sur les paradigmes de lécologie comportementale et de l'écologie du mouvement. En s'appuyant sur des données du 'Vessel Monitoring System', nous étudions le comportement des pêcheurs d'anchois du Pérou à des échelles différentes: (1) les modes comportementaux (c-à-d. recherche, pêche et route) au sein des marées (sorties de pêche), (2) les patrons comportementaux parmi les marées, (3) les patrons comportementaux par saison de pêche conditionnés par des scénarios écosystémiques et (4) les patrons spatiaux des positions de modes comportementaux, que nous utilisons pour la création de cartes de probabilité de présence d'anchois. Pour la première échelle, nous comparons plusieurs modèles Markoviens (modèles de Markov et semi-Markov cachés) et discriminatifs (forêts aléatoires, machines à vecteurs de support et réseaux de neurones artificiels) pour inférer les modes comportementaux associés aux trajectoires VMS. L'utilisation d'un ensemble de données pour lesquelles les modes comportementaux sont connus (grâce aux données collectées par des observateurs embarqués), nous permet d'entraîner les modèles dans un cadre supervisé et de les valider. Les modèles de semi-Markov cachés sont les plus performants, et sont retenus pour inférer les modes comportementaux sur l'ensemble de données VMS. Pour la deuxième échelle, nous caractérisons chaque marée par plusieurs descripteurs, y compris le temps passé dans chaque mode comportemental. En utilisant une analyse de classification hiérarchique, les patrons des marées sont classées en groupes associés à des zones de gestion, aux segments de la flottille et aux personnalités des capitaines. Pour la troisième échelle, nous analysons comment les conditions écologiques donnent forme au comportement des pêcheurs à l'échelle d'une saison de pêche. Via des analyses de coinertie, nous trouvons des associations significatives entre les dynamiques spatiales des pêcheurs, des anchois et de l'environnement, et nous caractérisons la réponse comportementale des pêcheurs selon des scénarios environnementaux contrastés. Pour la quatrième échelle, nous étudions si le comportement spatial des pêcheurs reflète dans une certaine mesure la répartition spatiale de l'anchois. Nous construisons un indicateur de la présence d'anchois à l'aide des modes comportementaux géo-référencés inférés à partir des données VMS. Ce travail propose enfin une vision plus large du comportement de pêcheurs: les pêcheurs ne sont pas seulement des agents économiques, ils sont aussi des fourrageurs, conditionnés par la variabilité dans l'écosystème. Pour conclure, nous discutons de la façon dont ces résultats peuvent avoir de l'importance pour la gestion de la pêche, des analyses de comportement collectif et des modèles end-to-end.

Mots clés: système de suivi des bateaux; données de suivi; mouvement des forageurs; modèles de semi-Markov cachés; validation des modelès; anchoveta *Engraulis ringens*

Acknowledgements

It doesn't interest me
where or what or with whom
you have studied.
I want to know
what sustains you from the inside
when all else falls away.

Oriah Mountain Dreamer (The Invitation)

I remember one day I was returning home from the CRH[1]. It was late and I was tired. I was preparing a presentation for a close date. Back in those days, I was feeling uncomfortable with moving (from cities and residences) too often. Because I was continuously looking for a place to stay in, I didn't have a place that actually felt like home and where I actually felt like hanging pictures or putting up posters. That night, walking back to the place where I was living, I saw a beautiful full moon, one of the biggest images of the moon I had ever seen. I then recalled that a friend of mine posted a comment on facebook about the beautiful full moon she saw the night before in Cuzco. I thought that later that night (local time in Peru), she would see that moon again. By contemplating the moon, I felt connected. And I suddenly realized that, just as the moon shines with a light that is not hers [2], we also shine with a light that is not ours. It's the light of the people that we find in our way the one that makes us shine. Then, it hit me. I already had a home. Home was in my heart and in the heart of the people that love me. All the pictures that I have, I hang them in my soul and I carry them wherever I go.

Now that this PhD journey is coming to an end, I'd like to take some time to thank the people that have given me light and provided me home in these three years.

[1]CRH: Centre de Recherche Halieutique

[2]The moon is feminine in Spanish (la luna). In the Andean cosmology, the moon or Mama Quilla is a goddess, both sister and wife of the god Inti (the Sun). She is the goddess that protects women.

I'll start with my advisors, who have significantly supported me (p-value is not shown) through this journey. Sophie, I think this has been a rich experience for both of us, for learning how to work with each other and for knowing each other in a personal level. Although I'm still learning to put you in copy of the e-mails, some progress has been made. The two things for which I am more grateful to you is your availability (deeply appreciated during the final race) and your honesty. Thank you for always encouraging me to do more, for always challenging me to do things better and to not forget the questions when I was having fun and sometimes losing myself with the ways to find answers. Ronan, it has been great meeting you and working with you. I have learned a lot and had great Markovian fun; so much fun that during a long period, I didn't want it to end. I'm always amazed by the speed of your thinking process and your practicality for tackling methodological issues. It has been really difficult to keep up with your pace, but that has been a joyful challenge. Thank you for always motivating me to do more, in a methodological perspective. Although you have been pretty busy (pretty much all the time since you returned to Brest), I really appreciated the time that you dedicated to working with me, talking with me about my academical progress and how I was feeling. I also treasure our conversations about the children education systems in Peru and in Bretagne, and how you put great efforts to make it work as you think it should.

I also want to thank my official advisor over most of the PhD, Arnaud Bertrand, a.k.a. papá pollito. Thank you for being always available, despite the zillion things you always have to do (I wonder how many gallons of coffee I need to drink in order to follow your rhythm). It impresses me how you always have a minute or two for your students, either for academic purposes or just for asking how we feel. You will probably never understand the excitement that an algorithm or an equation can cause. Likewise, it still costs me to recognize an ecological finding and therefore, to get excited by it. But it is something I've started to learn from you, as well as to tell 'sexy stories' in papers. Moreover, it has been really fun and great for my stomach (thanks for the cookies!) to share an office with you in IMARPE (you know, with you and the other four to eight fellows).

I want to thank great colleagues and friends I worked with as well, who were part of my thesis committees and/or coauthors. Alexis, Jorge, Mariano, Monique, Hervé, Marilú, Nicolas, thank you very much for your disposal and your advice.
Likewise, I thank the members of the jury at the PhD defense: Jan Jaap, Stéphanie, Roger, David, Francisco and of course, Sophie and Ronan. I never thought I would

say this, but I really enjoyed and had a lot of fun during the defense. Thank you for reading the PhD manuscript and for your questions. It would be a pleasure to continue the discussions that we started that day.

I also want to thank the students I got the pleasure to meet at Télécom-Bretagne, particularly Hakim, Jia, Said, Malek, Omid, Souhaila, Brahim, and Sileye and Pierre (postdocs). Even though I spent very little time there, I felt really welcomed and enjoyed the stay. Jia, I hope you can go visit Peru soon and meet the vicuñas you seem in love with. I also want to thank Kai, my dorm neighbor, for the welcome and the earplugs. The journeys in Brest would not have been the same without Eric, Eric, Gégé, Thierry, Andrea, Claude, Kim, Belia, Claudia and Arturo. I am very happy I met such great friends. I also want to thank Pierre and Irène. It has been a great pleasure to meet you – your love for each other just blows my mind. I am really grateful for the way you opened up your heart and your home to me, as you do for every other person you receive.

Through my journeys in Sète I met really good friends with whom I had awesome parties and very interesting conversations. Thanks to Ainhoa, Edgar, Jeanne, Liliana, Rigo, Marino, Iker, Maitane, Daniel, Ghislaine, Andréa, Camille, Emily, Maria Grazia, Justin, Isabelle, Alexandra, Florence, Marion, Elisabeth, Mariana, Tarek, Bastien, Jérémie, Lysel, Blandine. It would make a very long paragraph to describe all the reasons why I am grateful to each one of you (trust me, I've tried). Either for the political, historical and cultural conversations, the sharing or listening about personal processes, the delicious food, the amazing parties, the rides to CRH or the airport, the help in the last days before and right after the defense, the great company, the PhD soundtrack, the shiny smiles or even a hug, I truly thank you from the bottom of my heart. All the people at CRH has been really nice, and I want to give a particular acknowledgement to Nicolas and Pierre who welcomed me since I first arrived – actually in Montpellier, for my master studies. Thank you and your families for everything. I also want to thank the temporary members of the 'secta peruana' (A. Mousseigne, pers. comm.): Carlos, Vicky, Pepe, Ricardo, Ana, Daniel, Giannina, and of course, the honorary member, Alexander (but only because you had a Peruvian nanny, lol). I had a blast with all of you and it was also great to count on Peruvian products to cook quite often!

And speaking of Peruvian, the pollitos (little chickens) hold a special place in my heart. I thank Zaida, Daniel, José, Gary, Rosmery, Vilma and Erick. You've made

the office a really fun place to work in, and it is always inspiring to see more people constructing a better future for themselves and the community. Lula and Marie are also great friends whose work, laughs and smiles contributed to a great environment that made it easier to stand the traffic in Lima for arriving to work. It was also pretty great to go out eating and dancing with you and the pollitos. Sharing meals at the 'casita verde' (I don't remember the real name of the restaurant) with Dante, Katixa, Marie, Danny, Criscely, Luis, Jorge, Carlos, Jonathan, was always a pleasure (except for the almost absence of speed for going to the restaurant and walking back to IMARPE). I also want to thank Michael, Johanna, Ainhoa, Marie, Claude for the great times at the office and the dinners. Michael, I miss the bread you prepared. And of course, Yann, it was pleasant and fun to share and transform the office with you, and to decorate it with pictures of Tux. Special thanks to the department of pelágicos in IMARPE, particularly to Marilú, Erich, Julio and Andrés for helping in everything and collaborating with me, as well as inviting me to collaborate with you. Rodrigo was a key actor for my thesis. Without his dedication and efficiency for typing great amounts of data, I would have done a lot less than what I've been able to. I also thank Giannina and Omar; I've learned a great deal from both – completely different – short and intense internship experiences.

This journey was also marked by other fundamental aspects going on in my life. For that, I thank the members of the PIE. Inspired in El País de las Mujeres and infected by the enthusiasm awaken in several places in Latin America, we dared to dream together about a different world, a different way to make politics where happiness was above all, and that, is something I'll always treasure. I don't know how things will go on, but I did start to believe that if a bunch of people share a crazy fantastic dream and are ready to fight for it, a revolution can happen. Besides, fighting fear with hope is something I learned from the women of Mujeres Dignidad, when campaigning on the second round of presidential elections where the menace of dark times coming back was strong. Adapting another paragraph of The Invitation: 'It didn't matter where you lived, what you did for living, or how much money you had. What you showed me is that you could get up after a day of grief and despair, weary and bruised to the bone, and do what needed to be done for our country and our people'. I will never forget our meetings several times a week until late at night, the discussions, the fights, then the demonstrations once a week, the flowers we handed out to the people in the streets, the pickets in the parks next to the families of the disappeared (what brave hearts they do have!), the discussions with the police, the anonymous threads that most of you received,

the fear we experienced when it seemed our hopes and efforts were not going to be enough, and the joy of a battle won. Thank you Luz, Lena, Myc, Liz, Lucía, Gony, Lourdes, Zaria, Marita, Eliana, Maggie, Vilma and many others. I've been honored to meet you, to work at your side and to be your friend.

I also want to thank friends and family that have always been there to listen to me and have allowed me listening to them as well. I particularly thank my parents and my grandmother, for believing in me and trusting me even when they did not understand nor share my choices, and my brother, who always had a smile and a hug to warm the heart. My dear friends from the UNEC, some of them whom I saw in Lima and some of them in Europe; great thanks to Paula, Margot, Lichi, Mariel, Luisfer, Consuelo, Bea, Paola, MZ, Adam, Huaché, Guille, Miguel, Karina, Martín for accompanying my process, questioning me and motivating me with their lives.

Most of all, I need to thank Daniel for being such a great partner. No words can describe how much I owe you. I can only try an approximation – instead of a 'numerical approximation method', it could be like a 'poetical approximation' – by using the words of Otto René Castillo:
Si me preguntaras / qué es lo que más quiero / sobre la anchura de la tierra, / yo te contestaría: / a ti, amor mío, y a la gente / sencilla de mi pueblo. / (...) / A ti te quiero, porque eres el mío. / El compañero que la vida me dio, / para ir luchando por el mundo.

Because food, drinks and shelter have been key to survive these years, I want to thank, first of all, Daniel, for feeding me during the last months of the thesis. I also want to thank the guy that sells quinua and pan con palta (avocado) in front of IMARPE because that's the best breakfast ever! He and his wife wake up everyday at about four o' clock in the morning to prepare the quinua and the sandwiches. I've been amazed by the dedication and strength of this couple for guaranteeing a better future for their kids and beating the obstacles in their path. I also want to thank Steph for the laughs and IRD for the low prices at the Patelle. I would like to thank every person that brought me a Pisco bottle from Peru; I hope I'm not forgetting any name: François, Daniel, Ana, Marie, Mariano, Jeremie, Arnaud, Pepe. I also want to thank Edgar for the tequila and Daniel for the rum. I must give enormous thanks to all of you who have given me shelter and shared apartments and dorms with me: Vicky, Manuel, Luz, Luna, Mariam, Justin, Jeanne, Emily, Thomas, Marie, Daniel, Ainhoa, Flo, Angélica, Ricardo, Ana, Maria Grazia, Angelo

and Sombra. And during my two-week break, Salvatore, Miguel, Karina, Margarita, Guille and Joana. For everyone I lived with, it has been really great to get to know you, and for everyone who received me in their homes, I will never be able to thank you enough.

Last, but not least, I want to make a sincere recognition to the administrative staffs of IRD in Lima and Sète, and Télécom-Bretagne: Lula, Liliana, Alain, Jeannette, Ghislaine, Laurence, Isabelle, Anne-Catherine, Monique. I really suck at doing paperwork, I get easily lost and very much stressed in the proceedings. So your help and guidance, which sometimes was just part of your work and sometimes was way further, felt like a rescue.

For every gesture that made me feel at home, for every heart I found shelter in, and for every person that made me get up in the morning with a hopeful heart and carry a backpack full of books and dreams, I genuinely thank you. Merci beaucoup, muchas gracias, eskerrik asko, milesker anitz, trugarez braz, muito obrigada, tante grazie.

Agradecimientos

No me interesa dónde has estudiado, ni qué has estudiado,
ni con quién lo has hecho.
Quiero saber qué es lo que te sostiene desde adentro
cuando todo lo demás falla.

Oriah Soñadora de la Montaña (La Invitación)

Recuerdo un día en que volvía a casa desde el CRH[3]. Era tarde y estaba cansada. Estaba preparando una presentación para una fecha cercana. En aquellos días, me estaba sintiendo un poco incómoda mudándome (de ciudades y residencias) demasiado seguido. Al estar buscando continuamente algún lugar donde quedarme, no tenía ningún lugar que de verdad se sintiera un hogar y donde tuviera ganas de colgar cuadros o pegar posters. Aquella noche, volviendo al lugar donde estaba viviendo, vi una bella luna llena, una de las imágenes más grandes de la luna que he visto. Luego recordé que una amiga mía publicó en su muro de facebook que había visto una luna llena hermosa la noche anterior en Cuzco. Pensé que más tarde aquella noche (hora local de Perú), ella vería esa luna nuevamente. Contemplando la luna, me sentí conectada. Y luego me di cuenta de que, así como la luna brilla con una luz que no es la suya, nosotros y nosotras también brillamos con una luz que no es la nuestra. Es la luz de la gente que encontramos en nuestro camino la que nos hace brillar. De repente, me di cuenta de que ya tenía un hogar. Mi hogar estaba en mi corazón y en los corazones de la gente que me quiere. Todos los cuadros que tengo, los cuelgo en mi alma y los llevo por donde quiera que vaya.

Ahora que este recorrido se termina, quisiera tomar un tiempo para agradecer a la gente que me ha brindado luz y proporcionado un hogar en estos tres años.

Voy a comenzar por mis directores de tesis, quienes me han apoyado significantivamente (el p-valor no es mostrado) a través de este camino. Sophie, creo que ésta

[3]CRH: Centre de Recherche Halieutique

ha sido una experiencia rica para ambas, para aprender como trabajar con la otra y para conocernos en un nivel personal. Aunque aún estoy aprendiendo a ponerte en copia de los e-mails, hay algún progreso. Las dos cosas por las que te estoy más agradecida son tu disponibilidad (profundamente apreciada durante el último tramo) y tu honestidad. Gracias por siempre motivarme a hacer más, por retarme a hacer mejor las cosas y a no olvidar las preguntas, cuando de pronto me estaba divirtiendo y a veces perdiendo con las formas de encontrar respuestas. Ronan, ha sido genial conocerte y trabajar contigo. He aprendido mucho y tenido mucha diversión Markoviana, tanta que durante un largo período, no quise que terminara. Siempre estaré asombrada por la velocidad de tu proceso de reflexión y tu practicidad para abordar asuntos metodológicos. Tratar de ir a tu paso ha sido difícil y al mismo tiempo, un reto placentero. Gracias por motivarme siempre a hacer más, desde una perspectiva metodológica. Aunque has estado muy ocupado (prácticamente todo el tiempo desde que volviste a Brest), realmente valoro el tiempo que dedicaste a trabajar conmigo, conversar conmigo sobre mi progreso académico y sobre cómo me iba sintiendo. Valoro mucho también nuestras conversaciones sobre los sistemas educativos en Perú y Bretagne, y cómo haces grandes esfuerzos para hacer que el sistema camine como crees que debe ser.

También quiero agradecer a mi director de tesis oficial durante gran parte de mi PhD, Arnaud Bertrand, alias papá pollito. Gracias por estar siempre disponible, a pesar de las millones de cosas que siempre tienes que hacer (me pregunto cuántos galones de café tengo que tomar para ir a tu ritmo). Me impresiona cómo siempre tienes un minuto o dos para tus estudiantes, sea para propósitos académicos o para preguntar cómo están. Probablemente nunca entenderás el entusiasmo que surge de un algoritmo o una ecuación. De la misma manera, aún me cuesta reconocer un hallazgo ecológico y por lo tanto entusiasmarme por ello. Pero es algo que he comenzado a aprender de ti, así como también contar historias sexys en los papers. Además, ha sido muy divertido y genial para mi estómago (¡gracias por las galletas!) compartir una oficina contigo en el IMARPE (ya sabes, contigo y las otras 4 a 8 personas).

Quiero agradecer a colegas y amigos con quienes también trabajé, y que fueron parte de mis comités de tesis y/o coautores. Alexis, Jorge, Mariano, Monique, Hervé, Marilú, Nicolas, muchas gracias por su disposición y consejos. Asimismo, quiero agradecer a las y los miembros del jurado de la sustentación de tesis: Jan Jaap, Stéphanie, Roger, David, Francisco y por supuesto, Sophie y Ronan. Nunca

pensé que diría esto, pero realmente disfruté y me divertí mucho durante la defensa. Gracias por leer el manuscrito y por sus preguntas. Me daría mucho placer poder continuar las discusiones que empezamos aquel día.

También quiero decir gracias a los estudiantes que tuve el placer de conocer en Télécom-Bretagne, particularmente Hakim, Jia, Said, Malek, Omid, Souhaila, Brahim, y Sileye y Pierre (postdocs). A pesar de haber pasado poco tiempo ahí, me sentí realmente acogida y disfruté la estadía. Jia, espero que puedas visitar Perú pronto y conocer a las vicuñas de las que al parecer te has enamorado. También quiero agradecer a Kai, mi vecino de habitación, por la acogida y los recuerdos. Las estadías en Brest no habrían sido lo mismo sin Eric, Eric, Gégé, Thierry, Andrea, Claude, Kim, Belia, Claudia y Arturo. Estoy muy feliz de haber conocido tan buenos amigos. También quiero agradecer a Pierre e Irène. Ha sido un gran placer conocerles – el amor que se tienen me conmueve profundamente. Estoy muy agradecida por la forma en que me abrieron las puertas de sus corazones y su hogar, así como a toda persona que han recibido.

En mis estadías en Sète encontré muy buenos amigos con quienes tuve fiestas estupendas y conversaciones muy interesantes. Gracias a Ainhoa, Edgar, Jeanne, Liliana, Rigo, Marino, Iker, Maitane, Daniel, Ghislaine, Andréa, Camille, Emily, Maria Grazia, Justin, Isabelle, Alexandra, Florence, Marion, Elisabeth, Mariana, Tarek, Bastien, Jérémie, Lysel, Blandine. Necesitaría hacer un párrafo demasiado extenso para describir todas las razones por las cuales me siento agradecida con ustedes (créanme, lo he intentado). Sea por las conversaciones sobre política, historia, cultura, por compartir o escuchar sobre procesos personales, la comida deliciosa, las fiestas increíbles, los transportes al CRH o al aeropuerto, la ayuda en los días previos o justo después de la defensa, la grata compañía, la banda sonora del PhD, las sonrisas relucientes o incluso un abrazo, verdaderamente les agradezco desde el fondo de mi corazón. Toda la gente del CRH ha sido muy gentil, y quiero dar un reconocimiento particular a Nicolas y Pierre quienes me recibieron desde el primer día en que llegué – en verdad a Montpellier, para mis estudios de master. Gracias a ustedes y sus familias por todo. También quiero agradecer a las y los miembros temporales de la 'secta peruana' (A. Mousseigne, com. pers.): Carlos, Vicky, Pepe, Ricardo, Ana, Daniel, Giannina, y por supuesto, el miembro honorario, Alexander (pero solo porque tuviste una nana peruana...!). Me divertí mucho con ustedes y fue genial contar con productos peruanos para cocinar frecuentemente.

Y hablando de peruano, los pollitos ocupan un lugar especial en mi corazón. Gracias Zaida, Daniel, José, Gary, Rosmery, Vilma y Erick. Hicieron de la oficina un lugar divertido para trabajar, y es siempre inspirador ver más gente apostar por construir un mejor futuro para ellos y la comunidad. Lula y Marie son también grandes amigas cuyo trabajo, risas y sonrisas contribuyeron a un buen clima de trabajo que hizo más fácil soportar el tráfico de Lima para llegar a trabajar. Además fue genial salir a comer y bailar con ellas y los pollitos. Compartir almuerzos en la 'casita verde' (no recuerdo el nombre verdadero del restaurante) con Dante, Katixa, Marie, Danny, Criscely, Luis, Jorge, Carlos, Jonathan, fue un placer. También quiero agradecer a Michael, Johanna, Ainhoa, Marie, Claude por los buenos momentos en la oficina y las cenas. Michael, extraño el pan que hacías. Y claro, Yann, fue muy divertido compartir y transformar la oficina contigo, y decorarla con imágenes de Tux. Muchas gracias al área de pelágicos en IMARPE, particularmente a Marilú, Erich, Julio y Andrés por ayudar en todo cuanto podían y colaborar conmigo, además de invitarme a colaborar con ustedes también. Rodrigo fue un actor fundamental en mi tesis. Sin su dedicación y eficiencia para tipear grandes volúmenes de datos, habría hecho mucho menos de lo que finalmente pude hacer. Gracias también a Giannina y Omar, he aprendido mucho de ambas experiencias de prácticas cortas e intensas aunque totalmente diferentes.

Este camino también estuvo marcado por otros aspectos fundamentales que ocurrían en mi vida. Por ello, gracias a los y las miembros del PIE. Inspiradas en El País de las Mujeres y contagiadas del entusiasmo despertado en varios lugares de Latinoamérica, nos atrevimos a soñar juntas (y juntos) sobre un mundo distinto, una manera distinta de hacer política donde la felicidad estaba por sobre todo, y eso, es algo que siempre valoraré. No sé cómo continuarán las cosas, pero sí comencé a creer que si un grupo de gente comparte un sueño loco y fantástico y están listos para luchar por ello, puede haber una revolución. Combatir el miedo con esperanza es algo que aprendí de las Mujeres Dignidad, cuando hicimos campaña en la segunda vuelta de las elecciones presidenciales y la amenaza del regreso de los tiempos oscuros era fuerte. Adaptando otro párrafo de La Invitación: 'No importó donde vivían, lo que hacían para vivir, o cuánto dinero tenían. Lo que me mostraron es que podían levantarse después de una jornada de dolor y desesperanza, agotadas y golpeadas hasta los huesos, y hacer lo que había que hacer por nuestro país y nuestra gente'. Nunca olvidaré nuestras reuniones varias veces por semana hasta tarde en la noche, las discusiones, las peleas, luego las marchas semanales, las flores que entregamos a la gente en la calle, los piquetes en los parques junto a los familiares de desaparecidos

(¡qué corazones más valientes tienen!), las discusiones con la policía, las amenazas anónimas que muchas de ustedes recibieron, el miedo experimentado cuando parecía que nuestras esperanzas y esfuerzos no serían suficientes, y la alegría de una batalla ganada. Gracias Luz, Lena, Myc, Liz, Lucía, Gony, Lourdes, Zaria, Marita, Eliana, Maggie, Vilma y muchas otras. Me siento honrada por haberles conocido, trabajado a su lado y ser su amiga.

También quiero agradecer a amigos, amigas y familiares que siempre estuvieron para escucharme, y que me han permitido escucharles también. Particularmente gracias a mis padres y mi Oba, por creer y confiar en mí aún cuando no entendían ni compartían mis opciones, y a mi hermano, que siempre tuvo una sonrisa y un abrazo para calentar el corazón. Mis queridos amigos y amigas de la UNEC, algunas que reencontré en Lima y otras en Europa; muchas gracias a Paula, Margot, Lichi, Mariel, Luisfer, Consuelo, Bea, Paola, MZ, Adam, Huaché, Guille, Miguel, Karina, Martín por acompañar mi proceso, cuestionarme y motivarme con sus vidas.

Mais que nada, necesito agradecer a Daniel por ser tan buen compañero. No hay palabras para describir cuánto te debo. Sólo puedo intentar una aproximación – en lugar de usar un método clásico de aproximación numérica será uno de aproximación poética – apoyándome en palabras de Otto René Castillo:
Si me preguntaras / qué es lo que más quiero / sobre la anchura de la tierra, / yo te contestaría: / a ti, amor mío, y a la gente / sencilla de mi pueblo. / (. . .) / A ti te quiero, porque eres el mío. / El compañero que la vida me dio, / para ir luchando por el mundo.

Dado que comida, bebida y refugio han sido claves para sobrevivir estos años, quiero agradecer primeramente a Daniel, por alimentarme durante los últimos meses de la tesis. También quiero agradecer al tío que vende quinua y pan con plata frente al IMARPE porque eso es para mí el mejor desayuno del mundo. Él y su esposa se levantan cada día como a las 4 de la mañana para preparar la quinua y los sandwiches. Me han impresionado la dedicación y fuerza de esta pareja para garantizar un mejor futuro para sus hijos y superar los obstáculos del camino. También quiero agradecer a Steph por las risas y al IRD por los bajos precios a la Patelle. Asimismo, quiero agradecer a cada persona que me trajo Pisco desde Perú; espero no olvidar ningún nombre: François, Daniel, Ana, Marie, Mariano, Jérémie, Arnaud, Pepe. También quiero agradecer a Edgar por el tequila y a Daniel por el ron. Debo agradecer enormemente a todos y todas quienes me proporcionaron alo-

jamiento y compartieron conmigo departamentos y habitaciones: Vicky, Manuel, Luz, Luna, Mariam, Justin, Jeanne, Emily, Thomas, Marie, Daniel, Ainhoa, Flo, Angélica, Ricardo, Ana, Maria Grazia, Angelo y Sombra. Y durante mis dos semanas de vacaciones, Salvatore, Miguel, Karina, Guille y Joana. Para todos/as con quienes he vivido, ha sido un gran placer poder conocerles, y para cada persona que me ha recibido en sus hogares, infinitas gracias.

Por último, pero no menos importante, quiero hacer un reconocimiento sincero al personal administrativo del IRD en Lima y Sète, y Télécom-Bretagne: Lula, Liliana, Alain, Jeannette, Ghislaine, Laurence, Isabelle, Anne-Catherine, Monique. Soy malísima para hacer papeleo, me pierdo muy fácilmente en los procedimientos y eso me estresa mucho. Por ello, su ayuda y guía, que a veces era sólo parte de su trabajo y a veces iba mucho más allá, se sintió como un rescate.

Por cada gesto que me hizo sentir como en casa, por cada corazón en el que encontré abrigo, y a los portadores de sueños, que me hicieron levantarme en la mañana con un corazón esperanzado y cargar una mochila llena de libros y sueños, gracias totales, thank you very much, merci beaucoup, eskerrik asko, milesker anitz, trugarez braz, muito obrigada, tante grazie.

En todas las profecías
está escrita la destrucción del mundo.

Todas las profecías cuentan
que el hombre creará su propia destrucción.

Pero los siglos y la vida
que siempre se renueva
engendraron también una generación
de amadores y soñadores,
hombres y mujeres que no soñaron
con la destrucción del mundo,
sino con la construcción del mundo
de las mariposas y los ruiseñores.

Desde pequeños venían marcados por el amor.
Detrás de su apariencia cotidiana
Guardaban la ternura y el sol de medianoche.

Las madres los encontraban llorando
por un pájaro muerto
y más tarde también los encontraron a muchos
muertos como pájaros.
Estos seres cohabitaron con mujeres traslúcidas
y las dejaron preñadas de miel y de hijos verdecidos
por un invierno de caricias.
Así fue como proliferaron en el mundo los portadores de sueños,
atacados ferozmente por los portadores de profecías
habladoras de catástrofes.
Los llamaron ilusos, románticos, pensadores de utopías
dijeron que sus palabras eran viejas
y, en efecto, lo eran porque la memoria del paraíso es antigua
en el corazón del hombre.
Los acumuladores de riquezas les temían
lanzaban sus ejércitos contra ellos,
pero los portadores de sueños todas las noches
hacían el amor
y seguía brotando su semilla del vientre de ellas
que no sólo portaban sueños sino que los
multiplicaban
y los hacían correr y hablar.
De esta forma el mundo engendró de nuevo su vida
como también había engendrado
a los que inventaron la manera
de apagar el sol.

Los portadores de sueños sobrevivieron a los climas gélidos
pero en los climas cálidos casi parecían brotar por
generación espontánea.
Quizá las palmeras, los cielos azules, las lluvias torrenciales
tuvieron algo que ver con esto.
La verdad es que como laboriosas hormiguitas
estos especímenes no dejaban de soñar y de construir
hermosos mundos,
mundos de hermanos, de hombres y mujeres que se
llamaban compañeros,
que se enseñaban unos a otros a leer, se consolaban

en las muertes,
se curaban y cuidaban entre ellos, se querían, se ayudaban
en el arte de querer y en la defensa de la felicidad.

Eran felices en su mundo de azúcar y de viento
de todas partes venían a impregnarse de su aliento
de sus claras miradas,
hacia todas partes salían los que habían conocido
portando sueños
soñando con profecías nuevas
que hablaban de tiempos de mariposas y ruiseñores
y de que el mundo no tendría que terminar en la hecatombe.
Por el contrario, los científicos disearían
puentes, jardines, juguetes sorprendentes,
para hacer más gozosa la felicidad del hombre.

Son peligrosos
– imprimían las grandes rotativas
Son peligrosos
– decían los presidentes en sus discursos
Son peligrosos
– murmuraban los artífices de la guerra.

Hay que destruirlos
– imprimían las grandes rotativas
Hay que destruirlos
– decían los presidentes en sus discursos
Hay que destruirlos
– murmuraban los artífices de la guerra.

Los portadores de sueños conocían su poder
por eso no se extrañaban
también sabían que la vida los había engendrado,
para protegerse de la muerte que anuncian las profecías
y por eso defendían su vida aún con la muerte.
Por eso cultivaban jardines de sueños
y los exportaban con grandes lazos de colores.
Los profetas de la oscuridad se pasaban noches
y días enteros

vigilando los pasajes y los caminos
buscando estos peligrosos cargamentos
que nunca lograban atrapar
porque el que no tiene ojos para soñar
no ve los sueños ni de día, ni de noche.

Y en el mundo se ha desatado un gran tráfico de sueños
que no pueden detener los traficantes de la muerte;
por doquier hay paquetes con grandes lazos
que sólo esta nueva raza de hombres puede ver
la semilla de estos sueños no se puede detectar
porque va envuelta en rojos corazones
en amplios vestidos de maternidad
donde piesecitos soñadores alborotan los vientres
que los albergan.

Dicen que la tierra después de parirlos desencadenó un cielo de arcoiris
y sopló de fecundidad las raíces de los árboles.
Nosotros sólo sabemos que los hemos visto
sabemos que la vida los engendró
para protegerse de la muerte que anuncian las profecías.

Gioconda Belli (Los portadores de sueños)

Contents

List of Figures

List of Tables

Résumé Exécutif

Introduction

La pêche est probablemment la dernière activité de fourragement des êtres humains dans plusieurs endroits du monde (Cury and Miserey, 2008). Elle a fait l'objet d'études en écologie comportementale surtout dans des perspectives anthropologiques et archéologiques (centrées sur des populations anciennes et des techniques de pêche primitives; par exemple dans Bird and O'Connell, 2006; Cronk, 1991). La question du comportement de fourragement des pêcheurs modernes et de ses réponses adaptatives aux conditions écologiques n'a été abordée que dans très peu d'études, surtout à cause des difficultés d'accès aux données simultanées du comportement des pêcheurs et des conditions écologiques, mais aussi, dans la plupart des cas, à cause de l'hypothèse implicite que les pêcheurs sont au-dessus de tous les autres predateurs, et donc à l'abri de modifications des conditions abiotiques et biotiques sur le court terme.

Dans le cadre de l'écologie du mouvement (Nathan *et al.*, 2008), nous abordons la question de l'écologie du comportement des pêcheurs par l'analyse des données de ses trajectoires dans un contexte écologique. Ces analyses nécessitent un laboratoire naturel idéal, où des trajectoires des pêcheurs et des observations sur leur comportement seraient accessibles, et où la variabilité des composantes clés dans l'écosystème serait évaluée soigneusement. L'écosystème du Courant de Humboldt au large du Pérou offre ces conditions, pour trois raisons principales (Chavez *et al.*, 2008): (1) il soutient la pêcherie monospécifique la plus large du monde (anchois péruvien ou *Engraulis ringens*), (2) il est soumis à une variabilité climatique régionale intense à diverses échelles spatio-temporelles, et (3) il existe une forte surveillance de l'écosystème et la pêcherie. Cette surveillance comprend des informations par satellite sur les conditions environnementales, des indices sur la biomasse et la distribution des population de poissons, d'un système de suivi des bateaux de pêche (VMS, acronyme en anglais pour Vessel Monitoring System) et d'un programme

d'observateurs en mer pour une petite fraction de la flottille. En utilisant toutes ces sources d'information, nous visons à caractériser le comportement des pêcheurs d'anchois péruviens et à analyser comment leur comportement s'adapte à des conditions externes diverses.

Les dynamiques du comportement des pêcheurs sont explorées à quatre échelles différentes: (1) les modes comportementaux (c-à-d., recherche, pêche et route) au sein des marées (sorties de pêche) ; (2) les patrons comportementaux parmi les marées; (3) les patrons comportementaux par saison de pêche conditionnés par des scénarios écosystémiques; et (4) les patrons spatiaux des positions de modes comportementaux, que nous utilisons pour la création de cartes de probabilité de la présence d'anchois.

Données

Données sur la pêche

Pour ce travail, nous avons utilisé une base de données couvrant 100% de la flotille industrielle péruvienne de pêche d'anchois (> 1000 bateaux) de 2000 à 2009. L'utilisation du VMS par toute la flotille est imposée par la loi depuis 2000. Les positions des bateaux (± 100 m de précision) sont émises chaque ∼ 1 heure; pourtant il peut y avoir quelques irrégularités (par exemple, 0.17, 0.99, 1.2 heures) dans la durée d'un pas (constitué de deux émissions consécutives). Puisqu'il n'existe pas de méthode d'interpolation optimale standard pour ces cas (Langrock *et al.*, 2012), nous avons travaillé avec les émissions telles qu'elles sont. Par conséquent, nous avons considéré que les données VMS comprennent des trajectoires (c-à-d., des séries de positions) avec des pas non réguliers. Pour chaque trajectoire VMS, plusieurs variables observées ont été calculées à chaque pas: vitesse (sp), cap (θ), changement de vitesse et de cap entre le pas précedent et le pas présent (Δsp_{-1} et $\Delta\theta_{-1}$) et entre le pas présent et le pas suivant (Δsp_{+1} et $\Delta\theta_{+1}$). Pour chaque marée, plusieurs descripteurs ont été calculés aussi: durée totale (Dur), distance parcourue (Dist), distance maximale à la côte (Max.DC), latitudes et longitudes maximales et minimales atteintes durant la marée (Lat.Max, Lat.Min, Lon.Max et Lon.Min).

De plus, l'Institut de la Mer du Pérou (IMARPE), dispose d'un programme d'observateurs en mer qui couvre un échantillon de ∼ 1% des marées. Ils enregistrent la position et le temps associés à chaque mode comportemental réalisé au

cours des marées; à savoir: pêche, recherche et route. Pour les autres 99% des marées, les modes comportementaux sont unconnus.

En utilisant les critères décrits par Bertrand *et al.* (2007, 2005); Joo *et al.* (2011) et dans l'annèxe A, nous avons construit une base de données croisée (VMS-observateurs), pour laquelle le mode comportemental associé à chaque émission VMS est connu. Cette base de données couvre l'année 2008 et est composée de 242 marées.

Données sur les anchois

Depuis 1983, IMARPE conduit, en moyenne, deux campagnes scientifiques acoustiques par an, pour évaluer la biomasse et la distribution des populations de poissons. Ces campagnes comprennent des transects de $\sim$ 100 nm parallèles entre eux (séparés par $\sim$ 15 nm) et perpendiculaires à la côte. Des échosondeurs scientifiques Simrad (kongsberg Maritime AS, Norvège) travaillant à des fréquences différentes sont utilisés pour estimer les biomasses (voir Castillo *et al.*, 2009; Gutiérrez *et al.*, 2007; Simmonds *et al.*, 2009). Un échantillonnage extensif au chalut pélagique complète les évaluations acoustiques pour l'identification des espèces. Le 'nautical-area-backscattering coefficient' (NASC, en $m^2.mn^{-2}$), un indice de biomasse de poisson (Simmonds and MacLennan, 2005), est enregistré à chaque unité élémentaire d'échantillonnage geo-référencée (ESDU, acronyme en anglais) de 1 nm.

À partir de chaque campagne acoustique, cinq descripteurs ont été extraits: (i) le NASC d'anchois moyen, utilisé comme un indice de biomasse d'anchois (s_A); (ii) un indice de biomasse local (s_A^+), c-à-d., le NASC d'anchois moyen uniquemment pour les ESDU avec présence d'anchois; (iii) un indice d'occupation spatiale (ISO), c-à-d., le pourcentage d'ESDU avec de l'anchois; et (iv et v) le centre de gravité et l'inertie de la distance à la côte du NASC de l'anchois (DC et I, respectivement).

Données environnementales

Nous avons utilisé des données de température superficielle de la mer (SST) obtenues avec le capteur AVHRR des satellites de la NOAA entre 2000 et 2009. Les données satellite de chlorophylle-a (CHL) entre 2000 et 2007 ont été obtenues du capteur SeaWIFS et celles entre 2008 et 2009, du capteur MODIS. Les données CHL de MODIS ont été corrigées pour correspondre aux données SeaWIFS en utilisant la période commun pour les deux capteurs (2002-2008). Toutes les données satellite, initialement à une résolution spatiale de 4 km et temporelle hebdomadaire, ont été

moyennées par mois et sur toute la superficie étudiée au large du Pérou. Ensuite, pour obtenir une valeur représentative par période de temps étudiée, une valeur moyenne est calculée pour chacune de ces variables.

Pour prendre en compte la distribution verticale de la zone de minimum d'oxygène, un paramètre critique dans l'écosystème d'Humboldt (Bertrand *et al.*, 2011), nous avons calculé la profondeur de l'oxycline (OXY), c-à-d., la profondeur à laquelle l'oxygène dissolu est égal à 2 ml.l^{-1}. Toutes les mesures ont été faites à partir des bouteilles Niskin et des données CTD échantillonnées pendant les campagnes scientifiques conduites par IMARPE. Pour chaque période de temps étudiée, la profondeur de l'oxycline moyenne a été calculée comme une moyenne des valeurs mensuelles pondérée par le nombre d'observations par mois.

Méthodes

Nous présentons ici les méthodes statistiques utilisées dans ce travail: les modèles Markoviens, que nous utilisons pour inférer les modes comportementaux au sein des marées; les méthodes multivariées, que nous utilisons ensuite pour caractériser les marées et leurs associations avec des variables liées à la biomasse et distribution des poissons, et aux conditions environnementales; et les outils géostatistiques, que nous utilisons pour produire des cartes de probabilité de présence d'anchois et pour évaluer la co-variation spatiale entre ces probabilités de présence et la biomasse acoustique de l'anchois.

Modèles Markoviens

Les modèles de Markov cachés (HMM; Rabiner, 1989) sont les modèles les plus utilisés pour inférer des modes comportementaux en utilisant la série des données de trajectoires observées. Un HMM combine deux processus: un processus de Markov de premier ordre sous-jacent pour la séquence d'états (c-à-d., la séquence de modes comportementaux), où la probabilité d'être dans l'état s_t au temps t dépend uniquement de l'état précedent s_{t-1}; et un processus d'observation conditionnel aux états, où la probabilité de l'observation $X_t = x_t$ depend uniquement de l'état actuel s_t.

Dans un processus de Markov de premier ordre, il est supposé que le temps passé dans un état donné suit une loi géométrique. La loi géométrique est sans mémoire; cela veut dire que, au temps t donné, le temps d'attente pour changer d'état est

indépendant du temps passé dans l'état précedent. Cependant, en pratique, le comportement d'un pêcheur n'est pas sans mémoire. Alors, un processus de semi-Markov peut être plus adéquat. Il modélise les durées des états explicitement et peut considérer n'importe quelle loi de probabilité. Un modèle de semi-Markov caché (HSMM; Guédon, 2003) est donc un généralisation d'un modèle de Markov caché. Un HSMM combine deux processus: un processus d'observation conditionnel aux états comme dans les HMM et un processus de semi-Markov pour les états.

Modèles discriminatifs

Les modèles discriminatifs sont des approches alternatives pour inférer les modes comportementaux associés à des trajectoires. L'inférence de la séquence des modes comportementaux est défini comme un problème de classification, c-à-d., de la détermination de la classe (mode comportemental) associée à chaque position dans la trajectoire. Dans un cadre supervisé, les modèles discriminatifs apprennent des règles de classification pour prédire une classe à partir d'un vecteur observé x_t. Les forêts aléatoires (RF; Breiman, 2001), les machines à vecteurs de support (SVM; Burges, 1998) et les réseaux de neurones artificiels (ANN; Warner and Misra, 1996) sont parmi les techniques d'apprentissage automatique les plus utilisées (Hastie *et al.*, 2009). Pour les SVM, l'objectif est de maximiser la distance marginale de l'hyperplane qui sépare les classes. Pour les ANN, l'objectif est de minimiser l'erreur de classification. Enfin pour les RF, la discrimination est achevée par la minimisation des variances à l'intérieur des groupes et la maximisation de la variance entre groupes. Les performances relatives de ces méthodes varient selon la structure de l'espace d'observation de chaque cas d'étude particulier (Meyer *et al.*, 2003). Une propriété importante des modèles discriminatifs est qu'ils n'ont besoin d'aucune hypothèse sur la nature des variables observées, leurs distributions ou leurs covariances.

Méthodes multivariées exploratoires et descriptives

- Les analyses en composantes principales (PCA; Pearson, 1901) sont utilisées pour étudier la relation entre variables et entre individus, en réduisant la dimensionalité de la base de données.

- Les analyses de cluster sont utilisées pour grouper des individus de telle façon que les individus d'un même cluster soient plus similaires entre eux qu'avec ceux des autres clusters. Ce sont des méthodes de classification non supervisée

puisque les clusters ne sont pas définis à priori. L'analyse de classification hiérarchique organise les partitions dans un dendrogramme (Johnson *et al.*, 1992) ce qui permet d'étudier de niveaux de classification différents.

- Le coefficient RV (Escoufier, 1973) est une généralisation multivariée du coefficient de corrélation de Pearson au carré. Ses valeurs vont de 0 à 1, où 0 indique une association nulle et 1, une association parfaite.

- L'analyse de co-inertie (Dolédec and Chessel, 1994; Dray *et al.*, 2003) est une méthode multivariée utilisée pour établir le couplage de deux tableaux ayant le même nombre de lignes (qui peuvent être des individus ou des variables). Cette analyse vise à maximiser la co-inertie entre les lignes des deux tableaux. L'analyse de co-inertie multiple (Chessel and Hanafi, 1996) est une généralisation de l'analyse de co-inertie pour k tableaux ($k > 2$). Elle vise à maximiser la co-inertie entre les lignes des tableaux, et calcule un tableau synthétique représentant la structure en commun des tableaux.

Géostatistiques

Les méthodes géostatistiques donnent des outils pour capturer et modéliser la variabilité spatiale d'une variable quelconque distribuée dans l'espace ou simultanément dans le temps et l'espace (Chilès and Delfiner, 2012). Pour faire de l'interpolation spatiale, nous utilisons le krigeage, la méthode d'interpolation géostatistique. Celui-ci nécessite un modèle de variogramme théorique, qui modélise l'increment dans la variance entre deux points eloignés entre eux par une distance h, quand h augmente.

Pour évaluer les échelles spatiales auxquelles deux processus (comme la probabilité de présence et la biomasse acoustique) co-varient, nous nous appuyons sur un co-variogramme empirique. Un co-variogramme est une extension d'un variogramme pour deux variables; il évalue, à une distance h, dans quelle mesure l'incrément dans une des variables est associé à un increment ou une réduction dans l'autre variable, quand h augmente (Rivoirard *et al.*, 2000).

Résultats

Inférence des modes comportementaux

Nous avons comparé plusieurs modèles discriminatifs (RF, SVM et ANN) et Markoviens (HMM et HSMM) pour inférer des modes comportementaux associés aux trajec-

toires des pêcheurs, pour un échantillon de $\sim$ 300 marées pour lesquels les modes comportementaux étaient connus. Tous les modèles ont obtenu une précision globale supérieure à 75%. Le HSMM a été le plus performant, suivi par le HMM. La précision du HSMM à été supérieure pour l'inférence globale des modes (80%) et aussi pour chaque mode comportemental (pêche, recherche et route; 77%, 67% et 89%, respectivement). Une meilleure performance des modèles Markoviens par rapport aux modèles discriminatifs met en évidence l'importance de la modélisation des dynamiques des états pour pouvoir inférer de façon précise les séquences de modes comportementaux. Parmi les modèles Markoviens, le HSMM s'est montré supérieur pour représenter les séquences des modes comportementaux car il modélise explicitement la durée des états et considère des transitions à l'échelle du segment (c-à-d., une séquence consecutive des pas associés à un seul mode comportemental). Grâce à une expérience de simulation, nous avons montré qu'une augmentation de la résolution temporelle (au minimum une émission VMS toutes les 30 minutes) cause un incrément significatif dans la précision de la inférence du HSMM. Par conséquent, nous avons utilisé le HSMM pour inférer les modes comportementaux pour tous les trajectoires entre 2000 et 2009 qui n'ont pas été échantillonnées par les observateurs en mer.

Classification de patrons de marées

Les marées, caractérisés par les descripteurs présentés précédemment, plus le temps de la marée associé à chaque mode comportemental, ont été groupés en utilisant des analyses de classification hiérarchique. Les marées ont été groupées en 4 clusters, ce qui explique 61% de la variance. Un premier cluster, composé de 17% des marées, est associé aux pêcheurs dits 'stochastiques' (Allen and McGlade, 1986), qui prennent des risques importants et se lancent à la découverte. Les marées dans ce cluster sont les plus longues en durée et en distance, et les plus éloignées de la côte. Un deuxième cluster, composé de 15% des marées, est associé à la zone sud. Les politiques de gestion de la pêche dans cette zone sont différentes de celles de la zone nord-centre (par exemple, les dates de fermeture de la pêche, les quotas de capture, et les restrictions côtières pour pêcher). Un troisième cluster, composé de 32% des marées, est surtout lié aux marées des bateaux à coque en bois (les autres clusters sont plutôt lié aux marées des bateaux à coque en acier). Ces marées sont courtes et près de la côte, et une grande partie des marées est dédiée à la pêche et à la recherche, plutôt qu'à la route. Un dernier cluster est composé de 36% des marées. Celles-ci, ont parcouru des distances plus grandes, duré plus longtemps et se sont

plus éloignées de la côte que celles des clusters 2 et 3. Pourtant, ces distances et durées ne sont pas extraordinaires. Nous appelons les marées du quatrième cluster des marées ordinaires. Ainsi, nous avons trouvé des groupes des marées associés aux zones de gestion, aux segments de la flottille et à la personnalité des pêcheurs.

Les scénarios écosystémiques donnent forme aux patrons comportementaux des pêcheurs

En utilisant des données simultanées des campagnes acoustiques, information satellite sur l'environnement et VMS (décrits précédemment), nous avons étudié comment les conditions environnementales et de biomasse et de distribution de l'anchois donnent forme au comportement des pêcheurs à l'échelle d'une saison de pêche (c-à-d., les périodes de deux/trois mois dans une saison de pêche où toutes ces données sont accessibles). À l'aide des analyses en composantes principales et des coefficients RV, nous avons quantifié l'association entre les dynamiques de comportement spatial des pêcheurs (*pêcheurs*), les conditions environnementales (*environnement*) et la biomasse et la distribution spatiale de l'anchois (*poissons*). Les associations étaient plus fortes pour des relations plus directes (0.63 de coefficient RV pour *pêcheurs* et *poissons*, et 0.61 pour *poissons* et *environnement*) que pour la moins directe (0.56 pour *pêcheurs* et *environnement*). Toutes ces associations ont été statistiquement significatives. L'analyse de co-inertie multiple nous a permis de caractériser comment les *pêcheurs* sont conditionnés par *poissons* et *environnement*. Nous avons trouvé, par exemple, qu'en période d'été où la SST et la CHL prennent des valeurs élévées, et l'anchois est abondant localement et près de la côte, plus de temps est dédié à la pêche qu'à la route, et les pêcheurs restent près de la côte. En outre, quand l'oxycline était profonde, l'abondance locale et la superficie occupée par l'anchois était faible; dans ces conditions, les pêcheurs allaient loin de la côte pour chercher du poisson. Par ailleurs, quand l'anchois était trés concentré dans l'espace, les mouvements des pêcheurs étaient très diffusifs, c-à-d., il y avait beaucoup plus de pas courts (dans les aggrégations d'anchois) que de pas longs (entre les aggrégations). Enfin, nous avons montré que le comportement spatial des pêcheurs est conditionné fortement par des scénarios environnementaux. Ceci ouvre des possibilités pour l'utilisation des pêcheurs comme des indicateurs de l'écosystème.

Le comportement spatial des pêcheurs et la biomasse acoustique des poissons: les deux faces d'une même pièce?

Pour évaluer si le comportement spatial des pêcheurs reflète celui de la distribution spatiale de l'anchois, nous avons: (1) construit un proxy de la présence d'anchois en utilisant les modes comportementaux géo-référencés au cours de trois saisons de pêche, (2) fait des cartes du proxy de présence par krigeage, (3) comparé ces patrons spatiaux à ceux de la biomasse acoustique de l'anchois (aussi krigée). Au cours de cette dernière étape, nous avons trouvé que les co-variations en distance à la côte et superficie occupée varient selon la saison de pêche étudiée; pour 2001 et 2009 il y a des fortes corrélations, mais pour 2008 ce n'est pas le cas. Par ailleurs, les analyses des variogrammes croisés indiquent que la co-variation entre le proxy de présence d'anchois et la biomasse acoustique reste positive pour toutes les distances et les trois saisons, et atteint ses valeurs les plus élévés à des très grandes échelles. Cependant, les variogrammes de proxy de présence et de biomasse acoustique on montré des structures à petite échelle qui n'ont pas été trouvées dans les variogrammes croisés, ce qui indique que ces structures n'ont pas été co-occurrentes. Les deux sources d'information, pêcheurs et campagnes scientifiques acoustiques, semblent donc se complémenter mutuellement; et les cartes de proxy de présence peuvent être utilisées plutôt comme des cartes d'effort spatialisé et/ou des cartes d'anchois détecté par les pêcheurs.

Conclusions

Les analyses effectuées dans ce travail montrent que le comportement du pêcheur n'est pas invariant d'échelle, et permettent de mieux comprendre ce comportement à des échelles différentes. Dans la figure 0.1, un schéma de Stommel du comportement des pêcheurs est présenté. Nous montrons les unités de comportement étudiées (le mode comportemental, la marée, la saison de pêche); les facteurs qui, à l'égard de nos résultats, conditionnent le comportement à chaque échelle (flèche noire), et les structures des aggrégations de poisson et de l'environnement correspondantes aux échelles étudiées. Dans ce travail, nous avons montré que, à l'échelle du mode comportemental, ce sont les états internes qui comptent principalement pour inférer les modes comportementaux. Ces états internes se manifestent à travers la trajectoire observée (c-à-d., à travers les vitesses et les changements de cap) et les séquences comportementales (par exemple, le comportement au cours du segment précédent de la séquence conditionne le comportement au segment suivant). Ces composantes ont

permis l'obtention de 80% de précision lors de l'inférence des modes comportementaux par des modèles de semi-Markov cachés. À l'échelle de la marée, les facteurs principaux sont les règles de gestion (zones sud et nord-centre), les segments de la flottille (acier versus bois) et la personnalité des capitaines (preneurs de risques versus suiveurs). La variance expliquée par les clusters désignant ces facteurs est égal à 61%. À l'échelle de la saison de pêche, la biomasse et la distribution d'anchois, et des facteurs environnementaux tels que la température de surface de la mer, la chlorophylle-a et la profondeur de l'oxycline ont conditionné significativement le comportement spatial des pêcheurs (0.63 et 0.55 d'association entre pêcheurs et poissons, et entre pêcheurs et environnement, respectivement). Evidemment ces résultats sont opportunistes, car ils dépendent des données disponibles à chaque échelle. Cependant, les niveaux de variance expliquée sont suffisamment élevés pour confirmer que les processus identifiés sont les plus importants.

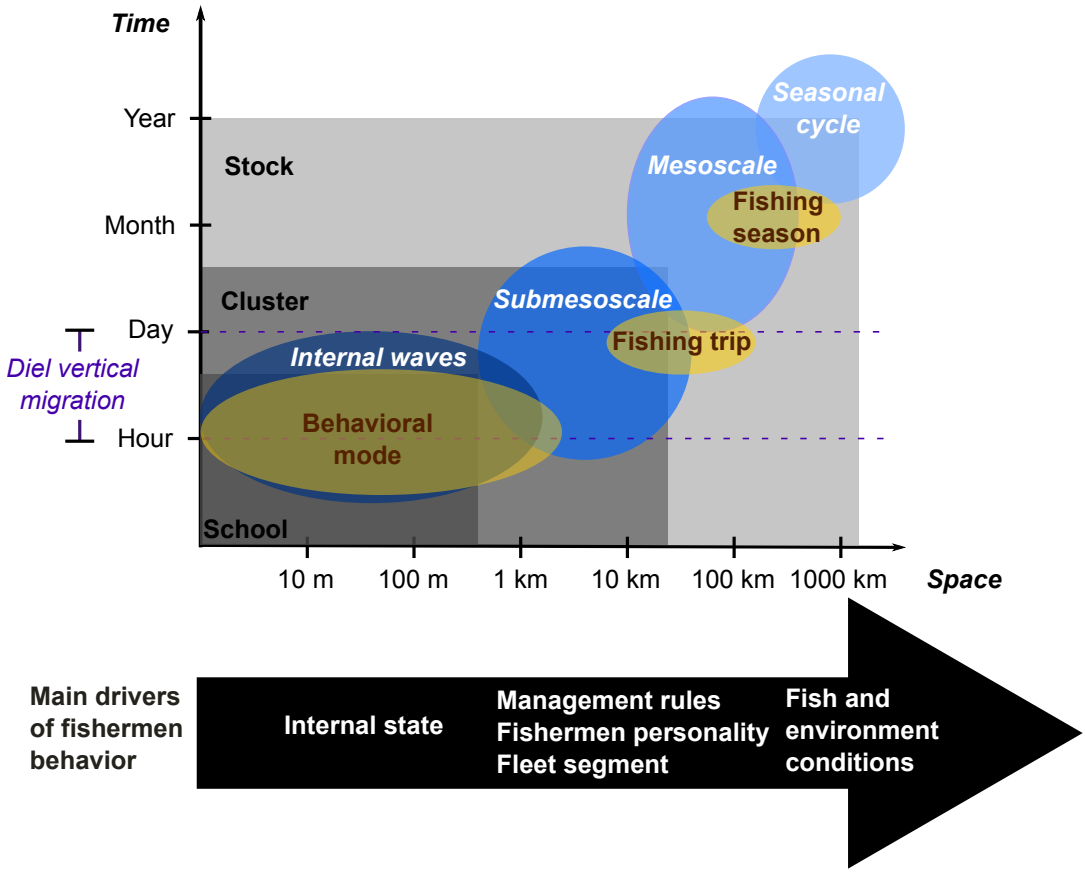


Figure 0.1: Schéma de Stommel pour les unités comportementales des pêcheurs et leurs moteurs. Les échelles spatio-temporelles pour les structures des poissons et de l'environnement sont basés sur Chelton (2001); Dickey and Lewis (2006) et Bertrand *et al.* (2008a).

Ce travail propose enfin une vision plus large du comportement de pêcheurs. Les pêcheurs ne sont pas seulement des agents économiques, ils sont aussi des fourrageurs, conditionnés par la variabilité de l'écosystème. Pour conclure, nous discutons dans ce travail de la façon dont les résultats obtenus peuvent servir de base à la gestion de la pêche, aux analyses de comportement collectif et à des modèles end-to-end.

Chapter 1

Introduction

Behavior is in the heart of many disciplines, such as ethology, comparative psychology, behavioral biology, behavioral ecology and evolutionary psychology. However, there is no unanimous definition of behavior. In an attempt to reach a consensus, Levitis *et al.* (2009) proposed behavior to be defined as the internally coordinated responses (actions or inactions) of living organisms (individuals or groups) to internal and/or external stimuli.

Behavioral ecology is the discipline that studies the variability of behavior in relation to ecological conditions (i.e., physical, biological and social conditions; Davies *et al.*, 2012). Ecology sets the stage on which individuals play their behavior, so the best way to behave depends on ecological selection pressures, such as the distribution in space and time of food, enemies and places to live. Behavioral ecology studies how variations in ecological conditions cause variations in the behavior that organisms display, whether at between-species, between-populations, between-individuals or even within-individual levels. Behavioral ecology also comprises the fitness consequences of behavioral strategies that organisms adopt; fitness is typically measured in terms of the survival, reproductive success or energetic return of an individual (Nettle *et al.*, 2013).

Human behavioral ecology focuses on how human behavior varies with the ecological context. Humans have been particularly successful for adjusting to diverse environments thanks to efficient adaptive learning and plasticity. Though recognizing humans' unique cognitive and behavioral capacities, the approach for studying human behavioral ecology is the same as for other animals: understanding the fitness costs and benefits given the ecological context, testing hypothesis of fitness maximization and making predictions that can be tested. At the beginning of human

behavioral ecology in the 1970s, research focused on studying foraging behavior and strategies in hunting, fishing and gathering populations (Cronk, 1991; Smith, 1983). Many aspects of foraging decision making were analyzed, such as: what they should eat, where they should forage, with whom they should forage, and how long they should forage (Cronk, 1991). In summary, how do their foraging strategies adapt to the surrounding context.

Nowadays, more than 40000 years after the first evidence on human fishing activities (O'Connor *et al.*, 2011), fishing may be the last human foraging activity in many places in the world (Cury and Miserey, 2008). Fishermen are peculiar foragers, as they typically rely on sophisticated technology and are strongly driven by economical factors. However, their foraging behavior still needs to constantly adapt to physical, biological, and social dynamics (Bertrand *et al.*, 2007). Fishing behavioral ecology has been mostly studied from anthropological and archaeological perspectives, where the object of study were ancient populations with primitive fishing techniques (Bird and O'Connell, 2006; Cronk, 1991). Due to the economical context, technological advances and globalization, analyzing behavior in modern fishing may seem more complicated. How could behavior be observed and analyzed? And how can its relationship with the ecological conditions be assessed?

Behavior can only be scientifically studied if it is performed – or analyzed – in the form of repeatable, publicly recognizable units (Ridley, 1995). Without recognized units of behavior, anecdotes might accumulate, but each of them would be closed to criticism, and rigorous testing of theories would be impossible. One way to conceive and interpret behavior is through movement (Ridley, 1995; Tinbergen, 1951). Indeed, in motile individuals like fishermen – or more specifically, fishing vessels – behavior can be expressed through movement. Movement paths can be decomposed into series of behavioral patterns, such as fishing, searching for prey, drifting, anchoring, for a fishing trip. Behavioral patterns can also be observed at larger scales through movement: when observing fishing trips, as individual units, some of them could be more sinuous than others, shorter than others, or go farther from the coast. Furthermore, fishermen behavior among several fishing seasons could also be compared and characterized. How ecological conditions shape fishermen behavior can be analyzed at any of these levels.

Those analyses would ideally require a natural laboratory, where fishermen trajectories and observations on their behavior are available, and the variability in key

components of the ecosystem can be thoroughly assessed. The Northern Humboldt Current System (NHCS) off Peru offers such conditions. First, because it sustains the world's largest monospecific fishery (Peruvian anchovy or anchoveta, *Engraulis ringens*). Second, because the NHCS is submitted to an intense regional climatic variability at a variety of spatio-temporal scales (Chavez *et al.*, 2008). And third, because the intense monitoring of the ecosystem and the fishery translates into a great amount of available data. The ecosystem monitoring comprises satellite information on environmental conditions (e.g., sea surface temperature, Chlorophyll-A, sea level anomaly) at daily and weekly resolutions; fish population distribution and biomass are monitored through scientific acoustic surveys (two to three times a year); and fishermen movement is monitored through a Vessel Monitoring System (VMS). Since 2000, the use of VMS tracking devices is mandatory for industrial purse-seiners. Vessels positions ($\pm$100m of accuracy; $\sim$ 1 record per hour) for hundreds of thousands of fishing trips are available for scientific purposes since. Fishermen behavior at sea is also documented through a program of on-board observers for a sample of $\sim$ 25 vessels per fishing season ($\sim$ 2% of the fleet).

In this work, we propose an approach to fishermen behavioral ecology at several spatio-temporal scales. We aim at characterizing Peruvian anchovy fishermen behavior by means of their trajectories and analyze how their movement behavior adapt to different ecosystem scenarios. The dynamics of fishermen behavior are explored at four scales: (1) the behavioral modes within fishing trips (which could be regarded as the elemental units of behavior observed in fishermen trajectories); (2) the behavioral patterns among fishing trips; (3) the behavioral patterns by fishing season conditioned by ecosystem scenarios; and (4) the spatial patterns of behavioral mode positions, that we use for building maps of anchovy presence proxy. The thesis is organized into six chapters as follows[1]:

Because behavior can be studied through movement, we will study the behavioral ecology of fishermen through movement ecology. **Chapter 2** provides a brief introduction to movement ecology, the different approaches (Lagrangian and Eulerian) to modeling movement, the trajectory data and the approaches taken in ecology for movement analysis. Within the movement ecology framework, we introduce the analysis of fishermen behavior, for which the VMSs have played a fundamental role. After a brief description of how a VMS works and the type of data that it provides, we review the modeling approaches to fishermen movement, the contributions of

[1]Most of the chapters (all of them excepting 2 and 3) are written in the form of scientific articles

movement data on fisheries science and management, and the remaining challenges.

Since the present work is applied to the fishermen in the Northern Humboldt Current System, in **Chapter 3**, we provide an outline of the NHCS that should help the reader to contextualize and interpret the following chapters, particularly the last two chapters. We describe the oceanography and biology, as well as their dynamics at multiple scales. We then describe several aspects of the anchovy fishery, the governmental management and the tools for monitoring the fishery and the ecosystem.

In **Chapter 4**, we study the behavioral modes within fishing trips. In order to do that, the modes should be a priori defined. For the Peruvian anchovy fishery, on-board observers in conjunction with skippers, defined a list of the possible activities made at sea during a fishing trip: cruising, fishing, searching, drifting, helping other vessels, and receiving/giving fish from/to other vessels. Since these activities represent distinct behaviors, we consider them behavioral modes. The sequences of behavioral modes associated with the fishing trips are only registered by on-board observers for a sample of $\sim$ 25 fishing vessels ($\sim$ 2% of the fishing fleet). We thus turn to VMS position records, which are available for 100% of the fishing fleet. Disposing of concomitant VMS and on-board observers data, we build a groundtruthed dataset, i.e., samples of tracks or positions for which behavioral modes are known. The goal is to use the groundtruthed dataset for training and validating several models, to choose the model with best performance and to use it for inferring the behavioral mode sequences in all the fishing trips.

In **Chapter 5**, we analyze the variability in fishermen behavior at the scale of fishing trips. We hypothesize that, at this scale, the skipper's personality, vessel characteristics and management restrictions generate different patterns in fishermen behavior, and thus that fishing trips could be grouped according to these patterns. We characterize each trip by a series of features such as its duration, maximum distance to the coast, maximum and minimum longitudinal and latitudinal locations, and time spent in each behavioral mode (estimated with the methods from the previous chapter) and reveal different patterns in fishermen trips using cluster analysis.

In **Chapter 6**, we explicitly analyze how environmental and anchovy conditions shape fishermen behavior at the scale of a fishing season. Features describing average spatial and temporal behavior of fishermen throughout a fishing season are

computed. In most cases, they are averages of the fishing trip features computed in the preceding chapter. We describe and quantify the associations between these features and descriptors of the environment (sea surface temperature, Chlorophyll-A and oxycline depth) and the anchovy biomass and distribution (assessed through acoustic surveys).

In **Chapter 7**, we examine a broader scale: we study the set of positions – and their corresponding behavioral modes – of the whole fishing fleet at a time range of $\sim$ 30 days. We hypothesize that fishermen spatial behavior reflects anchovy spatial distribution and thus, that maps of proxy of anchovy presence can be built from fishermen spatial behavior. Because VMS data is available in almost real time, the maps of presence proxy could be highly appealing for management purposes. We build the maps of presence proxy and compare them to maps of anchovy acoustic biomass. We then evaluate the co-variation of spatial descriptors between the presence proxy and the acoustic biomass.

General conclusions on the scientific contributions of this work and perspectives are presented in a final chapter.

Chapter 2

Movement ecology and fishermen

"All animals impart movement and are moved for the sake of something, so that this is the limit of their movement, the thing for-the-sake-of-which"

– Aristotle (*De Motu Animalium*)

2.1 Movement ecology

In motile individuals, behavior can be expressed through movement. Movement ecology aims at studying the process of movement as a result of a dynamic and continuous interaction between the environment and the organism's internal states at multiple scales.

Nathan *et al.* (2008) propose a movement ecology framework for exploring the causes, mechanisms and patterns of movement, that should also facilitate the understanding of the consequences of movement for the ecology, adaptability and variability of individuals, populations and communities (Fig. 2.1). Movement is composed of 4 components: an internal state, a motion capacity, a navigation capacity and external factors. The internal state accounts for the physiological and psychological states that drive the organism to fulfill certain goals; it addresses the question of why to move? The motion capacity accounts for the mechanisms that enables moving (how to move). The navigation capacity accounts for the ability to orient in space and time, deciding when (initiation and cessation) and where (direction and position) to move. And finally, the external factors comprise all possible spatio-temporal structures and dynamics that can make the individual modify its movement; it includes other individuals. Movement paths are produced as a result of the interplay between those 4 components.

Moreover, given that movement is performed by an individual whose behavior adapts and changes through certain time scales, movement should also be expected to vary and adapt. Hence, the causes and consequences of movement should be investigated as well. Therefore, describing movement patterns is essential and criteria should be established for relating those patterns to their causes and consequences (Levin, 1992). There is no single correct scale, approach or method for analyzing movement patterns. In the next section, we introduce the data needed for studying movement, the main existing approaches and methods.

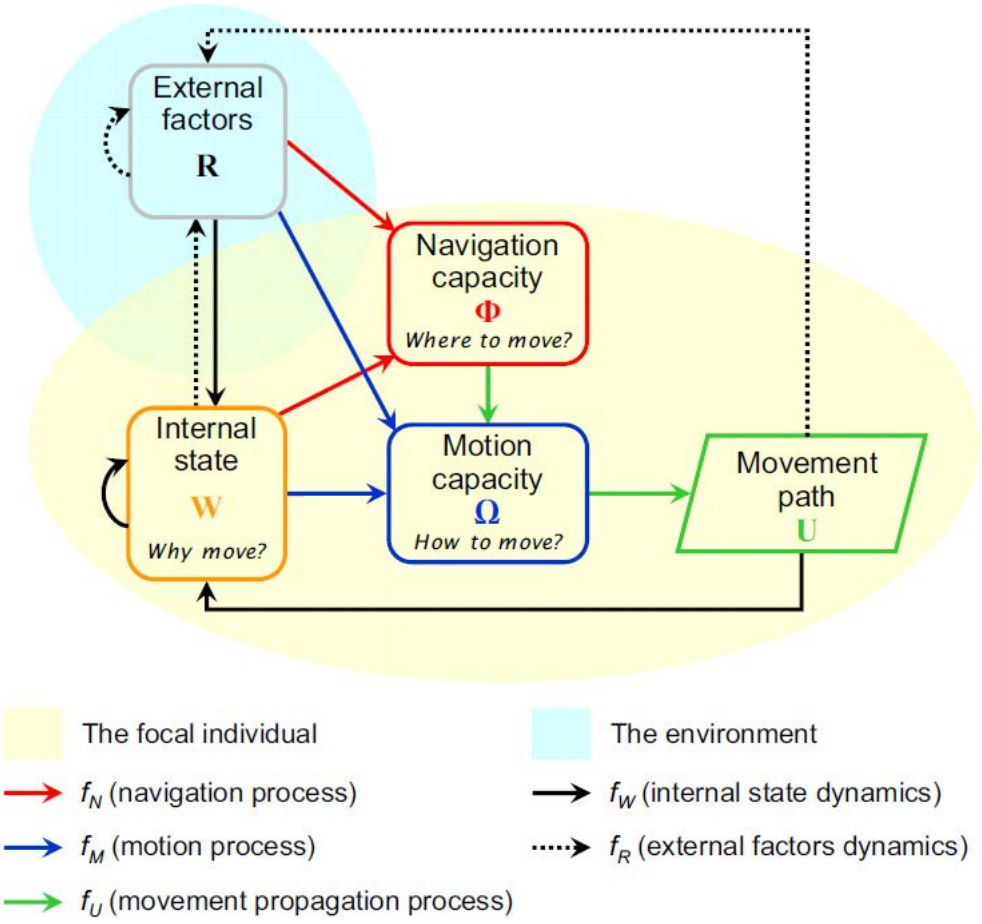


Figure 2.1: General conceptual framework for movement ecology proposed by Nathan *et al.* (2008)

2.2 Frameworks for studying movement

Movement analysis can be regarded under two main frameworks: the Lagrangian and the Eulerian approaches (Turchin, 1998). Lagrangian and Eulerian points of

view roughly refer to the individual and population levels of organization. The Lagrangian point of view is centered on the moving individual; i.e., the movement is studied by following the trajectory through space and time. Under this framework, detailed characteristics of the individual's behavior can be explicitly analyzed. For empirical applications, the major drawback of that framework is that it requires relatively large series of trajectory data; for simulation studies, the major drawbacks are the computational costs. By contrast, the Eulerian point of view is centered on a point in space. This point is characterized by densities and fluxes of moving organisms; i.e., it is like taking a photograph of the fluxes field at a particular fixed time. It may allow a more practical calculation of the spatial variability of the organisms distribution as a function of time (Hernández, 2012). Yet, individual properties are not explicitly taken into account.

In the past, the Eulerian framework was largely preferred in movement analysis; but with the increasing availability of trajectory data and the analytical and computational advances, the Lagrangian framework has become increasingly popular (Nathan *et al.*, 2008). Still, the Eulerian framework may remain the only practical way to study passively transported organisms such as microorganisms, airborne insects and seeds. Some works use mixed Eulerian and Lagrangian approaches for modeling populations considering the level of individual detail that the Lagrangian point of view provides (e.g., Black and McKane, 2012; Morales *et al.*, 2010)

In this work, we analyze fishermen movement within a Lagrangian framework. Following each individual through its trajectory will allow us understand the drivers of fishermen behavior individually, considering their fishing trips as composed of a sequence of behavioral mode units, and then considering the trips themselves as units of behavior. Beyond the scope of this work, combining Eulerian and Lagrangian frameworks would be desirable for understanding both the individual and collective movement. We now provide an introduction to tracking data and analyses in movement ecology.

2.3 Trajectory data

In general, the movement path of an individual is continuous and bounded in space and time. In order to document that path, observers must follow and record the individual without losing it, to get a spatio-temporal representation of its track

(Turchin, 1998). Nowadays, practically all studies use electronic devices for tracking tasks. Most of them are associated with global positioning systems (GPS), which are becoming more accurate and more available (i.e., cheaper), allowing the tracking of individuals practically anywhere in the planet (e.g., Block *et al.*, 2011). Thus, space does not represent a constraint to the observation, whereas time does: in almost all cases, individuals cannot be tracked throughout their whole life. Another time-related limitation resides in the discrete sampling inherent to observation. In order to record the observed path, it has to be represented in a discrete form suitable for computer storage and analysis (Turchin, 1998, Fig. 2.2). Time between steps (i.e., displacements between two consecutive positioning records) vary among studies (from infra-second frequencies in Wilson *et al.* (2008) up to 6 hours in Tew Kai *et al.* (2013)) and are not necessarily regular (Jonsen *et al.*, 2013).

Approximating a continuous movement by a discrete series of positions presents two types of flaws which depend on the sampling frequency and have an impact on the spatial information (e.g., Palmer, 2008; Ryan *et al.*, 2004). High-frequency data are sensitive to both measurement errors and low precision, leading to overestimation of distances traveled (i.e., length of the movement paths). Conversely, low-frequency data may translate into information loss, which leads to underestimation of the distances traveled. When too high-frequency data is available, error correction methods such as resampling or step aggregations are recommended (Turchin, 1998). State-space models are alternative methods for correcting error while modeling the movement process, and are claimed to be more rigorous and powerful (Jonsen *et al.*, 2013, Fig. 2.3). For low-frequency data, the ideal solution would be to increase the sampling frequency, which is often a financial problem rather than a technical one. If this is not possible, several interpolation techniques (besides simple linear interpolation) could be used for improving the representations of movement paths (e.g., Hintzen *et al.*, 2010; Russo *et al.*, 2011a; Tremblay *et al.*, 2006, Fig. 2.4).

2.4 Movement paths and random walks

Movement paths can be described by simple statistics (e.g., duration, distance traveled), measures of path sinuosity (Benhamou, 2004), and of autocorrelation (Legendre and Legendre, 1998; Perry *et al.*, 2002). Another type of measure is the fractal dimension (Bez and Bertrand, 2010; Mandelbrot, 1977), that can be used as an indicator of spatial coverage. As stated by Mandelbrot (1979), the fractal dimension

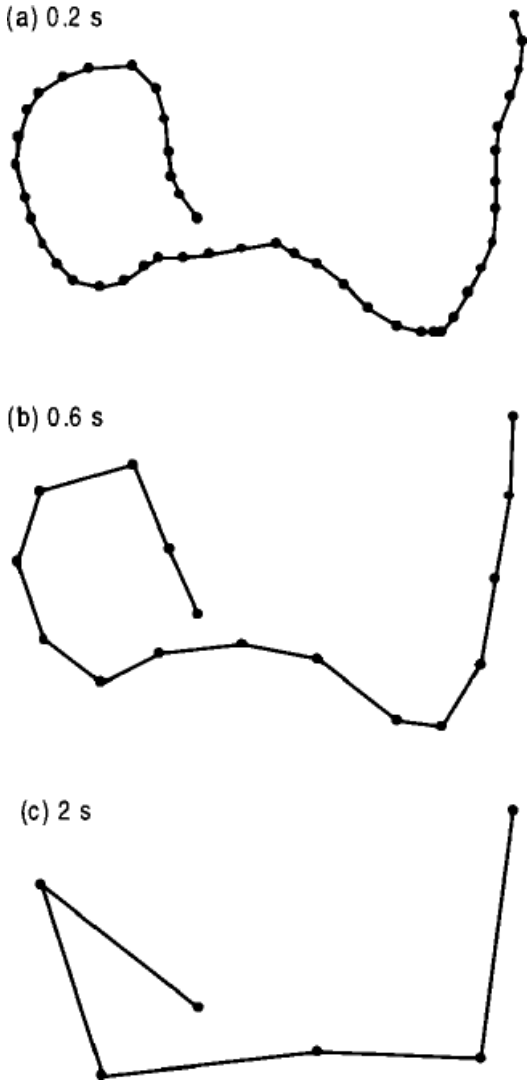


Figure 2.2: The path of a German cockroach (*Blatella germanica*) sampled at different time resolutions. Source: Turchin (1998)

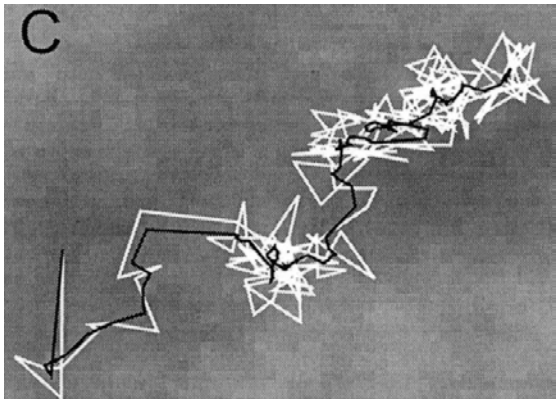


Figure 2.3: Graphical results of fitting a state-space model (SSM) to simulated movement data of a marine turtle. Plots of estimated (black) and observed (white) pathways, overlaid. Source: Jonsen *et al.* (2003)

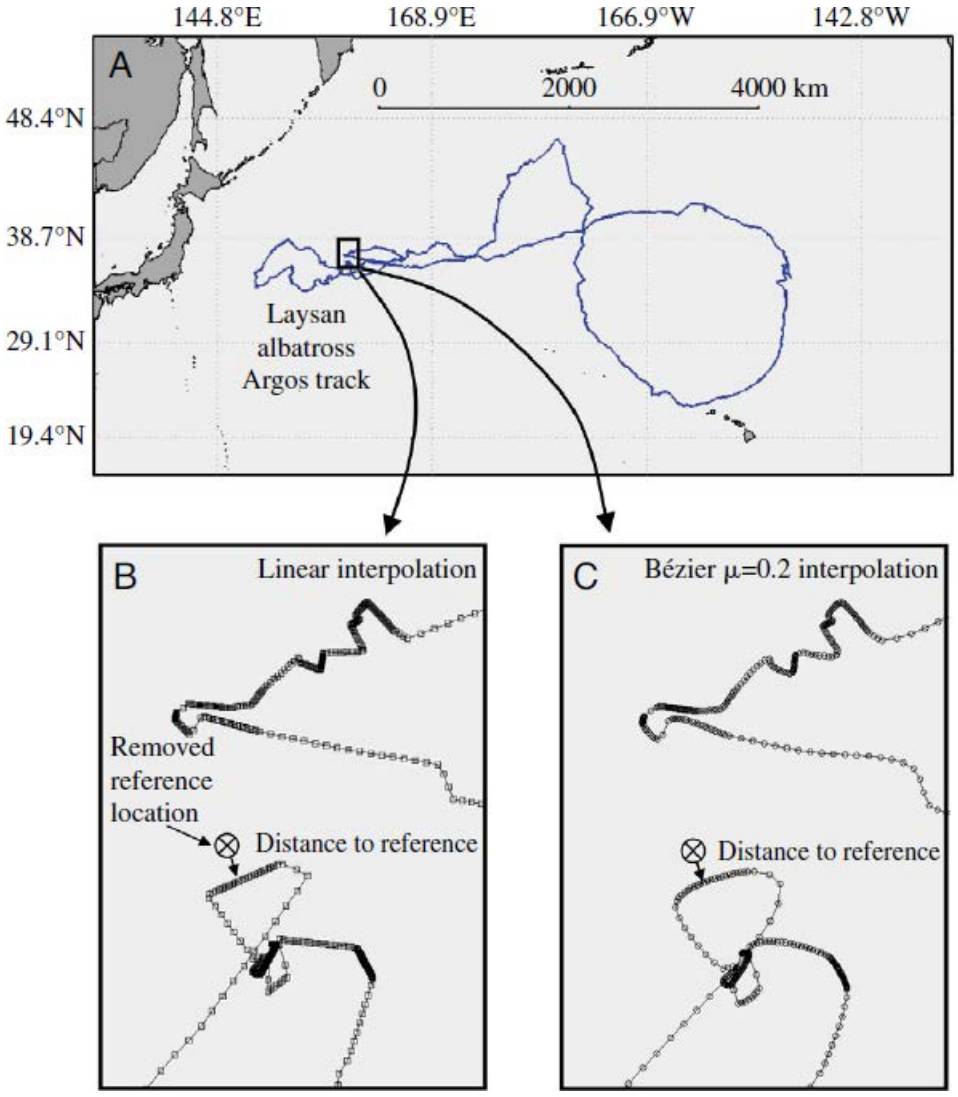


Figure 2.4: Laysan albatross Argos track (A), and selected examples of linear (B) and Bézier (C) interpolation of this track (every 10 minutes). Source: Tremblay *et al.* (2006)

of a set of two-dimensional points (e.g., a 2-D trajectory) can be seen as a measure of its propensity to cover the plane, with a value of 1 for the lowest plane coverage (a straight line or a circumference, for example) and 2 for the fullest coverage of some area in the plane (a Hilbert curve, for example).

Another approach for characterizing movement paths is random walk modeling. Random walks have been extensively used for modeling movement (see Codling *et al.*, 2008, for a review). It can be traced back to the analyses of Brown (1828) of pollen particles motion in what would later be called Brownian motion (Hänggi and Marchesoni, 2005). A random walk model is a formalization of the intuitive idea of taking successive steps, each one of a random length and in a random direction (Bartumeus, 2007). Thus, a trajectory can be regarded as a discretized sequence of steps of variable length, separated by variable turning angles. These discrete variable-length steps, reflecting different behavioral units, will be henceforth called moves, to distinguish them from the step defined previously (displacements between two consecutive recorded positions). A random walk is then a stochastic process defined by the statistical distributions of the move lengths and the turning angles.

The simplest models of movement using random walks are uncorrelated and unbiased; i.e., there is no autocorrelation in direction and there is no preferred direction (Codling *et al.*, 2008). Correlated random walks (CRWs), on the other hand, involve a correlation between the orientations of successive steps, what is termed persistence (Patlak, 1953). Random walk models with a consistent bias in a preferred direction are named biased random walks (BRWs), or biased and correlated random walks (BCRWs) if persistence is also observed.

The choice of the most appropriate random walk model for describing the movement of animals and humans (Brownian, Lévy, Poisson-like, as well as correlated, biased and bounded versions of all the others) has been keenly debated (Benhamou, 2007; Edwards *et al.*, 2007; Humphries *et al.*, 2010; Plank and Codling, 2009; Travis, 2007). Part of the debate focuses on the methodology for fitting and selecting the best model and the other part is centered on whether the individuals actually perform a given random walk. Fitting a random walk to an individual's movement basically consists in computing the length of the moves in the movement paths, and modeling the tail of the distribution of the move lengths (Lévy, Brownian and Poisson-like walks are characterized by power-law, Gaussian and exponential tails, respectively). The methodological discussions encompass data pre-treatment (Ed-

wards *et al.*, 2007), the choice of the start of the tail of the distribution (Bertrand *et al.*, in review), the method used for fitting a model to the data and estimating the parameters (Edwards *et al.*, 2007), the method for testing the goodness-of-fit (Bertrand *et al.*, in review), and the selection of the model. This last issue is not focused on the method for model comparison, but on the choice of the group of models to compare, since it differs among studies. An illustrative example is the work of Reynolds (2012), who compares Lévy walks with composite correlated random walks, claiming that Lévy walks remained contentious models partly because they were often compared with simplistic alternative models, rather than against strong alternative models.

Instead of determining 'the best' model for each type of organism, it is plausible that an individual, depending on the conditions and the spatial and temporal scales at which its movements are observed, can produce trajectories that are consistent with several distinct models (Benhamou, 2007; Humphries *et al.*, 2010). Bertrand *et al.* (in review) proposed an approach that lets the most likely random walk model emerge from the data instead of confronting a limited set of candidate models. They proposed the Generalized Pareto distribution (GPD; Pickands, 1975) for modeling the tail of the distribution of the move length distribution. The GPD distribution includes as specific cases the exponential, Gaussian and power-law distributions which underly Poisson-like, Brownian and Lévy walks, respectively (Fig. 2.5).

The question of whether individuals actually perform random walks or if what we observe are just random-walk-like patterns, has been eagerly addressed (see for example Bartumeus, 2009; Benhamou, 2007; Plank and James, 2008; Travis, 2007). Behind this question lie two perspectives: a phenomenological and pattern-oriented one, focusing on global and parsimonious models for each study case; and a biomechanical and process-oriented perspective, calling possibly for more complex models. In any case, the value of random walks should lie on their contributions for characterizing movement and behavior. In the words of Matheron (1978): 'Randomness is in no way a uniquely defined, or even definable property of the phenomenon itself. It is only a characteristic of the model or models we choose to describe it, interpret it, and solve this or that problem we have raised about it. (...) The only real problem is to know whether a given model, within a given context, does or does not possess an objective meaning to be able, if necessary, to perform a 'sorting out' operation'.

Indeed, the type of random walk model fitted to individuals movement has im-

portant implications on the interpretation of their foraging strategies. For instance, Lévy walks are characterized by heavy-tailed (power-law) move length distributions, due to the occurrence of numerous short moves, and rare and very large moves within the trajectory. This evidences super diffusive spatial behavior (Codling *et al.*, 2008) which may emerge from the patchy distribution of prey (Bertrand *et al.*, 2005); i.e., if prey are patchy, more time is spent within patches, and less patches are visited, what reflects a Lévy-like movement path. By contrast, Poisson-like models are characterized by exponential tails, and thus a smoother decay in the probability density function of the move-length distribution. Therefore, there are more long moves (compared with the number of short moves) than in the Lévy case. Translated into foraging movements, it implies that they visit more patches, probably because the latter are small and dispersed. The GPD approach proposed by Bertrand *et al.* (in review) provides a synoptic characterization of the observed movement through two parameters describing the spatial range and the diffusive property. Those two parameters constitute relevant metrics (1) as potential indicators of changes in the spatial distribution of their prey and (2) for studying the variability of the spatial behavior among species or among individuals with different personalities.

The GPD approach seems highly appealing because it contemplates a continuum of distributions and can thus encompass a continuum of diffusive behaviors that can be found in organisms' movement. It also enhances the interpretation of foraging strategies by providing a spatial range parameter in addition to the sinuosity parameter. These parameters will be later used in our work for characterizing fishermen behavior (Chapter 6).

2.5 Identifying behavioral modes in movement

In theory, a movement path results from the succession of distinct types of behavior or behavioral modes, e.g., standing, walking, running, searching for prey, eating a prey. In practice, a movement path is sampled through a movement track, i.e., a series of positions with either regular or irregular steps. If a movement track was almost continuous, we could consider that the points of transition between behavioral modes are always recorded. Under this assumption – implicit in most works concerning behavioral modes – we define a segment as a succession of steps associated with a distinct behavioral mode.

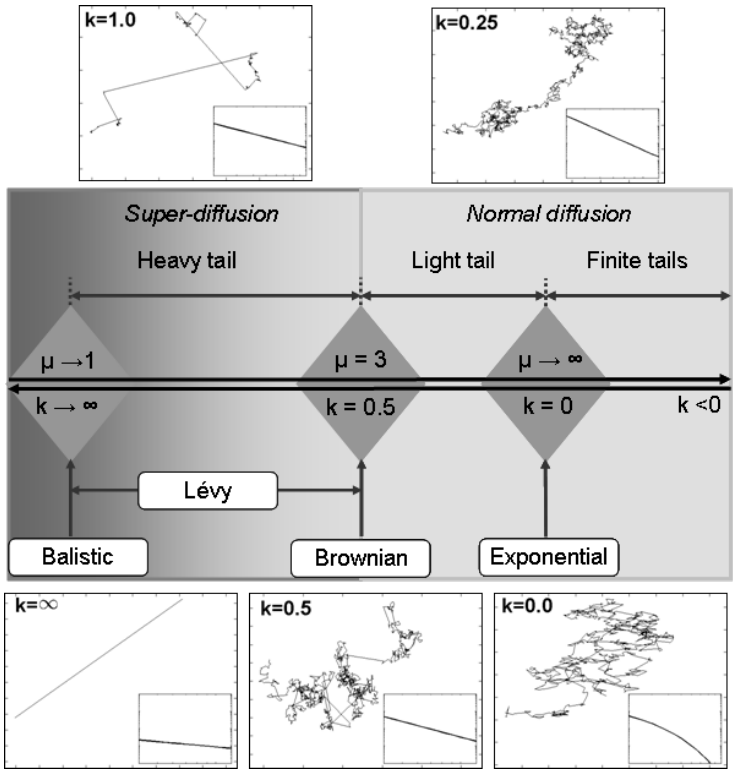


Figure 2.5: Generalized Pareto Distribution, a continuum from Exponential-Poisson to Power-Lévy walk patterns. Parameter k of the GPD defines a continuum of distributions from low-tailed ($k < 0$) to heavy-tailed ($k > 0.5$). We show typical trajectories emerging from random realizations of those different move distributions, including a Poisson-exponential motion ($k = 0$), a Brownian-Gaussian motion ($k = 0.5$) and a Lévy-power walk ($k > 0.5$) with $k = 1$. For each case, we also show in the lower right inset the log-log plot of the corresponding move length probability density function. Source: Bertrand *et al.* (in review)

In order to study the behavioral modes in a movement track, two approaches can be taken (Fig. 2.6). The first one consists in observing a spatial or a spatio-temporal representation of the movement path, and associating certain observed patterns (computed from the tracked observations) to distinct behavioral modes (Fig. 2.7). The second one consists in defining the behavioral modes a priori and then identifying which steps relate to each mode.

The recognition of behavioral modes without a priori defining them can be done visually, but it is more objective to use analytical methods (Turchin, 1998). One family of methods seeks to identify shifting points in the trajectories. For example, Thiebault and Tremblay (2013) proposed a method where movement tracks are divided into segments based on consistency in speeds and headings. They applied it to seabird foraging tracks and compared the resulting segmentation with bird-borne video cameras observations to evaluate if the segment limits matched changes in behavior. Gurarie *et al.* (2009), alternatively, introduced a behavioral change point analysis based on continuous stochastic processes for identifying structural changes. They applied this method to northern fur seal trajectories and then gave ecological interpretations for the identified behavioral units.

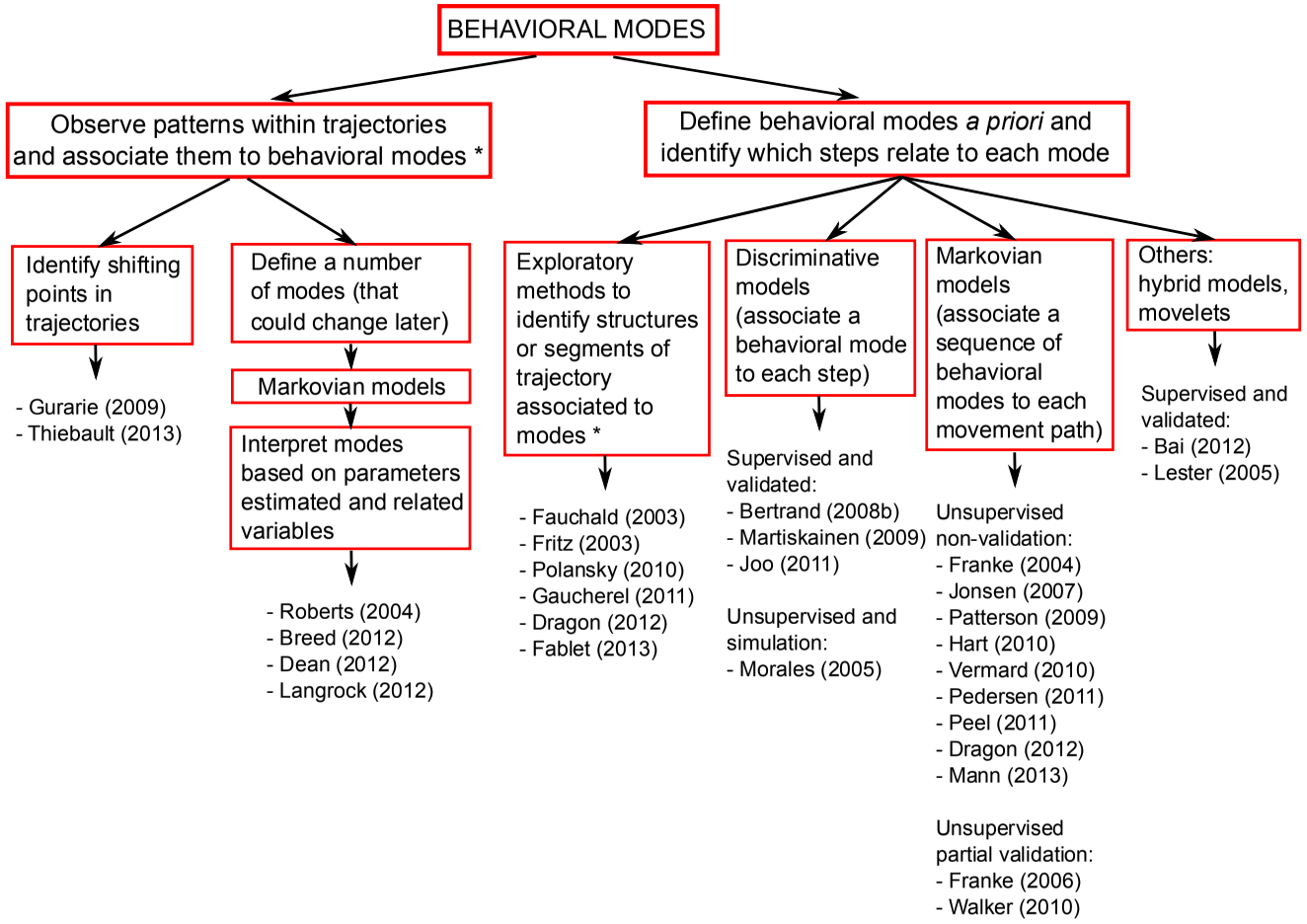


Figure 2.6: Approaches for analyzing behavioral modes in movement paths and examples of studies under each approach. Only first authors are mentioned. * Under these approaches, only supervised learning is possible.

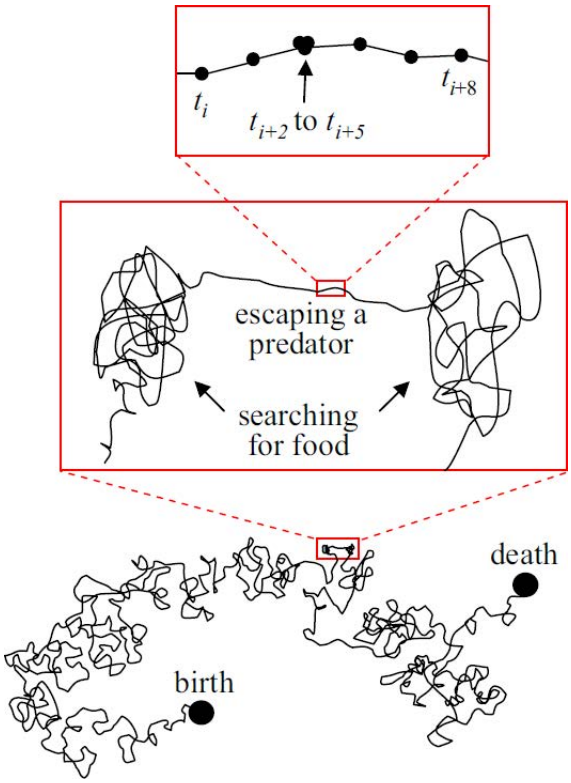


Figure 2.7: Behavioral modes observed in a movement path. Source: Nathan *et al.* (2008)

Another family of methods consists in: (1) defining a number of behavioral modes (that could or could not change later), (2) fitting a model that infers those modes based on the observed trajectory, and (3) interpreting the behavioral modes relying on the parameters estimated and the related observed variables. Among the most common models within this family are Hidden Markov models (HMMs), which account for the temporal dynamics of the behavioral modes via transition probabilities. In some studies, the number of states (behavioral modes) in the HMMs was clear from the beginning (e.g., Breed *et al.*, 2012; Langrock *et al.*, 2012, for sea lions and American bison, respectively) and the modeling results helped giving ecological interpretations to those states. In other studies, the choice of the number of states was not trivial. Roberts *et al.* (2004), for instance, analyzed bird navigation behavior using a variational learning HMM, where the number of states was fixed to 10, although they expected it to be lower. Since the modeling results showed that only 3 of the 10 states were significantly visited, they concluded that the birds' behavior was naturally organized into 3 states. Dean *et al.* (2012), on the other hand, analyzed at-sea behavior of seabirds and fitted a series of candidate HMMs, where each one had a different number of states (2-10). Then, the log-likelihood of each model was computed, and a 3-state model was chosen based on a knee-point

criterion.

Nonetheless, in a majority of studies, behavioral modes are defined a priori. In almost all cases, a foraging or searching mode is included among them. Exploratory techniques can be used for identifying structures or segments of the trajectory associated to one or several modes. Fauchald and Tveraa (2003) and Dragon *et al.* (2012), for example, used first-passage time (FPT) for identifying area-restricted search (and their associated steps) in antarctic seabird and elephant seal paths, respectively. Fritz *et al.* (2003) used fractal analysis for identifying searching behavior patterns in GPS tracks of albatrosses at several scales throughout their foraging trips. Wavelet analyses have been used for identifying the multiple scales at which behavioral modes occur. Fablet *et al.* (2013) used wavelet techniques for identifying reorientation patterns associated with foraging bouts in seabird tracks. Polansky *et al.* (2010) used them for studying changes in resting, feeding and moving behavioral modes in lion and buffalo tracks; and Gaucherel (2011), for analyzing resting, ballistic and searching behavioral modes of albatrosses.

Discriminative models, that state the classification of a step into a behavioral mode independently of the other steps in the trajectory, have also been used. Groundtruthed data, i.e. samples of tracks or positions with known behavioral modes, available in some studies allowed for supervised learning. For example, in Bertrand *et al.* (2008c) and Joo *et al.* (2011), steps associated with fishing behavior were identified using artificial neural networks (ANNs) and validated with on-board observers data. Martiskainen *et al.* (2009) obtained great performance from support vector machines for recognizing various behavioral modes from accelerometry data on cows, where the model outputs were contrasted with direct observations and video recordings. In an unsupervised learning framework, Morales *et al.* (2005) used ANNs and genetic algorithms for identifying eating, foraging and exploring behavioral modes in simulated elk populations. Here, movement efficiency was assessed by the effectiveness in avoiding predators.

Unsupervised two-state Hidden Markov models were commonly used for inferring the behavioral modes, e.g., for classifying penguin dives into searching or foraging (Hart *et al.*, 2010), elephant seal steps into foraging or not foraging (Dragon *et al.*, 2012), turtleback steps into transiting or foraging (Jonsen *et al.*, 2007), tuna steps in searching or foraging (Patterson *et al.*, 2009), or in foraging or migrating (Pedersen *et al.*, 2011). It should be noted that not all of these works used the same concept

of 'foraging' (for some of them, it is just feeding, while for others, it involves searching for prey). Two-state HMMs were also used for identifying types of movement orientation in glass prawn movement (Mann *et al.*, 2013b). Unsupervised HMMs with more states were used for identifying several behavioral modes: e.g., bedding, feeding and relocating within caribou and wolf tracks (Franke *et al.*, 2004, 2006); fishing, cruising and stopping within fishermen tracks (Peel and Good, 2011; Vermard *et al.*, 2010); and fishing, cruising and searching also within fishermen tracks (Walker and Bez, 2010).

Groundtruthed data were used for validating the models when they were available. Franke *et al.* (2006) used data on wolf killing sites for validating two HMMs inferring kill/not-kill modes and locally-active/bedding/relocating modes (two- and three-state HMMs, respectively), by assessing if the killing sites were within a given radius from positions identified as kill and locally-active. Walker and Bez (2010) calibrated and validated HMMs for inferring fishing, cruising and searching modes by comparing the inferred fishing modes with collected fishing observations. Lester *et al.* (2005) used supervised hybrid HMM/discriminative models for identifying a series of everyday activities performed by humans wearing multi-sensor boards. The hybrid models were trained on an independent sample of video recorded data on the volunteers; during the recorded sessions they were asked to do specific activities.

A less conventional but supervised approach proposed by Bai *et al.* (2012) consisted in building a reference database of human behavioral modes and their observed patterns (in this case, from an accelerometer), then comparing movelets (i.e., moving windows in the accelerometer time series) to the references using a distance measure, and associating behavioral modes to the movelets so that the distances are minimized.

All of these methods are valuable tools for analyzing movement, depending on the goals motivating the behavioral mode identification (Fig. 2.6). It would be appealing to compare the efficiency of several methods for inferring behavioral modes, using groundtruthed data for validation and comparison of the models. This issue will be addressed in chapter 4, when comparing discriminative and Markovian models for inferring behavioral modes in fishermen movement.

2.6 Movement in an ecological context

As stated in section 2.1, movement is not an isolated action from an individual. External factors are important for understanding movement. In this section, we review the works studying how environmental factors shape individuals' movement through modeling and descriptive frameworks. Then, we review the approaches adopted for analyzing social behavior and collective movement.

Various studies have incorporated environmental covariates into their animal movement models and evaluated the effect of those covariates in the movement patterns. For example, Morales *et al.* (2004) used the type of habitat as a covariate for inferring encamped and exploratory behavioral modes in elk movements. While the animals encamped in open habitats, their exploration was not strongly associated with any particular habitat type. Bestley *et al.* (2010) used several environmental covariates for inferring feeding behavior in tuna movement. Although feeding mostly occurred in the coastal waters, tuna also showed a pattern of high foraging success throughout their migratory range, providing evidence of opportunistic feeding. Using a state-space model (SSM) with environmental covariates, Bestley *et al.* (2012) showed that switches from directed to resident behavioral modes in elephant seal movement were associated with cold water temperatures. In turn, Hanks *et al.* (2011) used a velocity-based approach for modeling animal movement in space and time that incorporates environmental variables. Applying it to northern fur seal movement, they showed sex differentiation, with females exhibiting stronger response to the environmental variables (sea surface temperature, Chlorophyll-A, and net primary production).

Other works examined the association between movement patterns and environmental conditions in a descriptive framework. Fritz *et al.* (2003), for example, analyzed the relationships between the fractal dimension of albatrosses tracks at several spatial scales, wind conditions, and distance to the continental shelf. A significant association between the latter and the fractal dimension at one of the scales was found. Pichegru *et al.* (2007) analyzed the foraging zones of Cape gannets from 2 different islands (obtained from GPS and time-depth recorder data) and the pelagic fish abundance (obtained from acoustic surveys). They found that gannets from one of the islands foraged in an area that was closer to their island and had higher sardine densities, which could have contributed to their growing population trends; whereas other gannets foraged farther away and in low-abundance zones, and

had a declining population. Jonsen *et al.* (2007) compared temperature and depth conditions between transiting and foraging behavioral modes of leatherback turtles; turtles transited in deeper and warmer waters than when foraging. Benhamou (2007) showed via simulation that environments with patchy prey could generate Lévy-like movements. Humphries *et al.* (2010), on the other hand, investigated the horizontal tracks of various open-ocean predatory fish and found Lévy behavior to be associated with less productive waters (sparse prey) and Brownian movements to be associated with productive shelf or convergence-front habitats (abundant prey).

Wittemyer *et al.* (2008) used a spatio-temporal approach for analyzing the interplays between individual and ecological (social rank, predation risk and seasonal variation in resource abundance) characteristics. They tested hypotheses on whether ecological conditions shaped autocorrelation patterns of step-length in elephant movement, obtained from Fourier and wavelet analyses. They showed that autocorrelation was weaker during the wet season, indicating that random movements were more common in good ecological conditions. They also found that diurnal movement correlation was more common with protected wildlife areas, and multiday movement correlations found among lower rank individuals were outside of protected areas where predation risk was higher. Finally, they found that the observed patterns were all related to the distribution of critical resources (i.e., forage and water).

For studying collective movement, Polansky and Wittemyer (2011) presented a Fourier and wavelet approach for analyzing pairwise movement patterns. They applied it to step lengths in the movements of African elephants; the degrees of movement synchrony found were consistent with the observed strength of the social bonds in each pair. Freeman *et al.* (2011) studied the flights of pigeons released alone and then in pairs. They found that, in the pairs, the one that showed the lowest variability in their solo path (and hence memorized it best) became the leader, and the other one, the follower.

Within the Lagrangian framework, several studies have addressed collective behavior modeling of animal movement, mostly inspired from self-propelled particles (SSP; Vicsek *et al.*, 1995) or in individual or agent based models (IBM or ABM; Reynolds, 1987). In these models, a group (e.g. a swarm, school, flock or crowd) is modeled by a collection of individuals (or particles) that move in a one, two or three dimensional space. Each individual has a local interaction zone within which

it responds to other individuals. The exact form of this interaction varies between models, but typically, individuals are attracted to, aligned with and/or attracted to other individuals within one or more different zones. This modeling approach has been used in many applications, such as pedestrian motion trails in a park (Helbing *et al.*, 1997), human crowd stampedes (Helbing *et al.*, 2000), locusts marching movement (Buhl *et al.*, 2006), bird flocks synchronized landing (Bhattacharya and Vicsek, 2010); and mostly shoaling fish dynamics (see Lett and Mirabet, 2008, for a review). In some works, video recording were used for confronting the simulation results with real data (Lett and Mirabet, 2008).

Several other modeling approaches have been undertaken. For example, Herbert-Read *et al.* (2011) used a neural network based learning algorithm for identifying the factors responsible for fish speed and direction in shoals of mosquitofish. They found that fish adapted their speed and heading primarily in response to their nearest neighbor. Mann *et al.* (2013b) modeled the collective motion of glass prawns, and used Bayesian model selection to show that the individual's memory of its recent interactions is key for determining its future direction choices; based on this methodology, human clapping dynamics were also modeled (Mann *et al.*, 2013a). A framework for modeling collective animal behavior is presented by Sumpter *et al.* (2012), focusing on the interplay between what happens on the individual level and the behavior of the group as a whole. Examples of this approach are given, based on past works of the coauthors. Quantitative model validation presents serious difficulties in collective behavioral modeling, mostly because of data unavailability: tracking data on collective movement seems very hard to obtain for animals that are not observed in laboratories or domesticated.

The works reviewed in this section highlight the role of environmental factors and social interactions in movement. Still, few studies have addressed these issues. The increasing amount of data in ecosystem monitoring and the methodological advances will allow for a better understanding of movement in an environmental and social environment.

2.7 Fishermen movement

We now introduce the analysis of fishermen behavior based on their movement. The implementation of Vessel Monitoring Systems in many industrial fisheries and their

availability for research purposes has played a fundamental role for fishermen movement analysis. We will provide a short description of VMS data in section 2.7.1. Relying on the movement ecology framework introduced before, we will review the existent studies characterizing fishermen behavior alone (section 2.7.2), then its response to environmental conditions and social interactions (section 2.7.3).

It has been widely recognized that understanding the spatial behavior and strategies of fishermen is key for fisheries science and management (Bertrand, 2005; Salas and Gaertner, 2004; Smith and Wilen, 2003; Wilen, 2004). Indeed, fisheries science is mostly motivated by management purposes. Management policies can act upon the catches or the effort. Traditionally, effort has been more complicated to monitor and so to control. For that reason, management has been mostly based on catch limits (e.g. total allowable catches or individual vessel quotas). Thanks to VMSs, fishermen can be monitored in almost real-time, with good spatial accuracy and at relatively fine temporal resolutions. For fisheries management, VMS data has been used for assessing the spatial distribution and dynamics of effort and thus introducing management policies related to effort reduction. We will review the management-centered works in section 2.7.4.

2.7.1 VMS data

A VMS is a real-time continuous tracking system of fishing vessel movements. VMSs have been extensively implemented in dozens of industrial fisheries throughout the world (Fig. 2.8) since the end of the nineties. Vessels participating in the VMS programs carry a shipboard electronic equipment with a unique identifier. Most shipboard units use satellite communications systems with an integrated Global Positioning System that calculates the unit's position and sends a data report to shoreside users. The standard data report includes the VMS unit's unique identifier, date, time and position in latitude and longitude (FAO, 2013a). Initially implemented for law enforcement and security purposes, VMSs provide high quality and low cost information on fishing vessels' trajectories. They have enabled great progress in the understanding of the spatial mechanisms involved in the fisheries dynamics mainly because: (1) these observations are collected continuously and at good temporal resolution (1 record per hour on average); (2) the information is independent of fishermen's declarations; and (3) the nature of the data allows applying a movement ecology framework to the analysis of fishermen's spatial behavior

(Bertrand *et al.*, 2007).

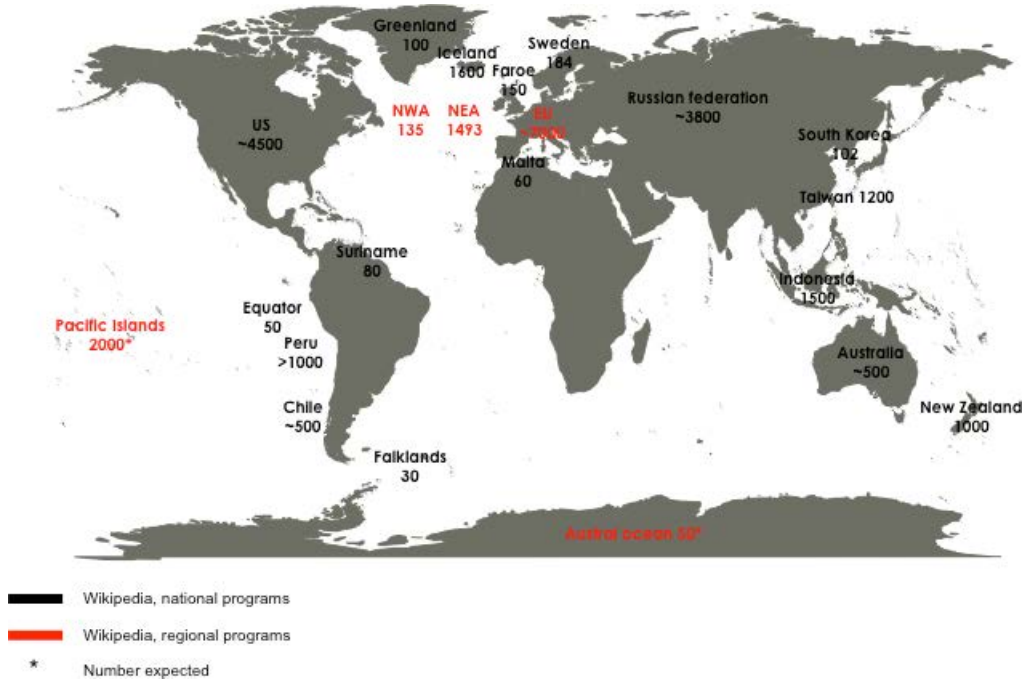


Figure 2.8: Non-exhaustive list of the main VMS in the world, synthesized from Smith (2001), FAO (www.fao.org) and wikipedia (www.wikipedia.com/xx). Only the VMS programs with publicly available number of vessels are shown (the figures are shown in the map). Source: Bertrand *et al.* (in review)

On one hand, never before has spatialized information on fishermen, at that time resolution been available for entire fleets. On the other hand, concerning the spatio-temporal representation of the movement path, the average sampling rate of 1 hour can only provide a coarse approximation to the real movement path. Linear interpolations of the fishing trajectories tend to underestimate the real traveled distances and thus their associated speeds (Palmer, 2008). For the same reason, the estimation of trip sinuosity indicators is not accurate neither. Sinuosity is rarely computed as a fishing trip descriptor; we will see in the next section how the speed underestimations can affect fishermen movement analysis.

2.7.2 Fishermen movement analysis

Classifications of movement paths into groups has been an issue addressed by Russo *et al.* (2011b), for identifying métiers based on vessel gear and fishing trip characteristics computed from VMS data (e.g., speed, turning angle, distance to the coast,

duration of the trip) using artificial neural networks. Several clustering methods for métier classification were evaluated by Deporte *et al.* (2012), but they did not rely on movement features but rather on landing profiles for the classification.

A few studies have used random walk modeling for characterizing fishermen movement. As explained in section 2.4, random walk modeling basically consists in computing the lengths of the moves in the movement paths and modeling the tail of the distribution of the move lengths. Moves are mostly computed by resampling the positioning records, using turning angles for detecting points of changes in the trajectories (Turchin, 1998). The existing applications of random walks correspond to the Peruvian anchovy fishery. In this fishery, fishing trips usually last $\sim$ 24 hours. Tracks with one positioning record per hour in average (i.e., the VMS data resolution) result in a small amount of moves. There are not enough move lengths for modeling the tail of their distribution. Then, the approach taken by Bertrand *et al.* (2007, 2005, in review) consists in characterizing not a movement path, but the movement of a fisherman (or fishing vessel), modeling the move lengths corresponding to a set of his movement paths by random walks. Under that approach, Bertrand *et al.* (2007, 2005) fitted Lévy walks to fishermen move lengths and used the parameter of the tail of the distribution of the move lengths for characterizing the sinuosity of the movement of each fishing vessel. Afterwards, Bertrand *et al.* (in review) fitted a Generalized Pareto distribution to the vessels move lengths and found evidence of more diffusive movement for the steel-hulled fleet than for the wooden-hulled fleet.

Behavioral modes in fishermen trajectories are defined as activities within the fishing trips, like fishing, cruising, searching or drifting. The vast majority of VMS-related works focused on determining which positioning records were associated with fishing operations (Bertrand, 2013). Most of those works used a simple criterion on speed (see Lee *et al.*, 2010, for an overall review on speed criteria for identifying fishing behavioral modes). However, speed is computed on distances and time, and as stated before, distances may be underestimated because of the relatively low resolution of the data, especially during sinuous segments (trajectories tend to be more sinuous in fishing zones); hence, speed may also be underestimated. Thus, determining fishing activity solely on a speed criterion increases the risk of false positives and thus of overestimation of the number of fishing sets. Only two works have quantified these estimation errors. Palmer and Wigley (2009) found false positive rates of 32% and 69% for bottom otter trawl and scallop dredge fishing activity, respectively, in the northeastern USA. Bertrand *et al.* (2008c) found an overestimation of 182% of

the number of fishing sets for the purse-seine anchovy fishery in a northern region in Peru. Those rates may explain why high percentages of the total of VMS records are identified as fishing sets: 47%, 79% and 80% for UK VMS corresponding to 2005 (Witt and Godley, 2007), 2006 and 2007 (Lee *et al.*, 2010).

More accurate estimations were obtained using artificial neural networks with time, speed, acceleration and turning angles as input variables for detecting fishing set positions in the Peruvian purse-seine anchovy fishery (Bertrand *et al.*, 2008c; Joo *et al.*, 2011). Aiming at minimizing the over/under estimation error, the calibrated neural networks gave 1% of average overestimation (compared to the 182% obtained with the speed threshold) and 24% of average false positives; the very small overestimation error implies that the false positives could be regarded as positioning errors of the fishing sets.

Other studies aimed at inferring the sequence of behavioral modes of each fishing trip, i.e., from a set of a priori defined behavioral modes, identifying which one most likely corresponds to each movement step. Inferred behavioral modes were mainly fishing, searching, cruising and staying still, for French pelagic trawlers (Vermard *et al.*, 2010), French tropical tuna purse-seiners (Walker and Bez, 2010), Australian trawlers (Peel and Good, 2011). Hidden Markov models were used for inferring the sequences of behavioral modes in all three scientific papers. Among them, Walker and Bez (2010) were the only ones to partially validate their results with groundtruthed data. Partially, because (1) the groundtruth was only available for the fishing mode, which impeached validation of the inference of the other activities; and (2) the groundtruthed tracks were simultaneously used for calibration and validation, so validation was not independent.

Concerning random walk modeling, this work is also situated in the Peruvian anchovy fishery context, and thus constrained by the number of moves within each fishing track. Notwithstanding, at a fishing season scale, the high number of moves per vessel does allow a random walk characterization; using the GPD approach in Bertrand *et al.* (in review) we extract the two parameters described above as additional indicators of fishermen behavior for chapter 6.

Regarding behavioral mode inference, it has been mostly done using speed-based criteria (for fishing behavior), but artificial neural networks and Hidden Markov models have shown to be appealing alternatives. Validation of the methods has

been regarded in very few works. It should be given more importance to model validation, especially since the results are later used for management purposes. We will address these issues in chapter 4, when comparing discriminative and Markovian models for inferring behavioral modes within fishing trips.

2.7.3 Fishermen movement in an ecological context

As for every other predator, the movement of fishermen responds to ecological factors. Movement tracks collected on fishermen are conditioned by prey and environmental conditions, and that has been the focus of a few studies. Bertrand *et al.* (2005) fitted Lévy random walks to move-length distributions of tracks corresponding to 3 anchovy purse-seiners. The super-diffusive behavior of fishermen – represented by a parameter from the power-law tail distribution – was significantly correlated with the fractal dimension of fish distribution, showing that fishermen spatial behavior reflected to some extent the spatial distribution of fish. Bertrand *et al.* (2008c) compared the spatial distributions of fishing sets from anchovy purse-seiners and fish acoustic biomass; they found significant correlations in distance to the coast and a clustering index. Bertrand *et al.* (2008b) described the effects of large scale oceanic forcing, via Kelvin waves, in all components of the coastal Humboldt Current system, from oceanography to fishermen. They showed that the arrival of a downwelling Kelvin wave generates a contraction in the extent of cold coastal waters, a contraction in the spatial distribution of anchovy and a deepening of their aggregations, triggering less diffusive fishermen trips (i.e., a Brownian-like motion indicating the low effort for finding fish and catching it, since anchovy is closer to the coast and highly aggregated).

Sharples *et al.* (2013a) analyzed the spatial patterns of fishing activity in the Celtic Sea and described how oceanographic characteristics (i.e., tidal currents, batimetry, sea surface temperature and chlorophyll) led to consistent patchiness in the fishing effort. The pattern of fishing showed the most heavily fished area was along the edge of the continental shelf, where a strong internal tide mixes nutrients towards the sea surface, increasing primary productivity and altering the species structure of the phytoplankton community. Sharples *et al.* (2013b) produced a synoptic description of the ecological processes in a Celtic Sea area, where the dynamics of productivity are highly influenced by internal Lee waves over a shelf sea bank. They showed that the dynamics of these internal waves strongly influence the pri-

mary production in the area, fish distribution and the distribution of a bird species, storm petrel. VMS data were analyzed for assessing whether the fishing vessels showed preferences in fishing grounds in that area. The results showed that while some marine organisms are able to quickly adapt their distribution to the spatial structures generated in the water masses by internal Lee waves, the commercial fishing vessels respond to a larger spatial scale, since they were concentrated generally around large fish aggregations, but with no particular preference for areas where intense Lee waves locally increase primary production.

The interactions between fishermen and their repercussions on their own spatial and temporal behavior has been analyzed to some extent. Horta and Defeo (2012) analyzed the spatial overlap between industrial and artisanal fleets targeting a same species (whitemouth croaker) in coastal Uruguay. The overlap negatively affected the artisanal fleet which was less competitive, so management recommendations for limiting overlapping were given. Poos and Rijnsdorp (2007) studied the dynamics of the Dutch beam trawl fleet spatial distribution when one fishing area was temporally closed, using a logbook dataset. They focused on the effect of interference competition and proposed a mean crowding index for measuring the strength of fishermen interference competition. In the first weeks after the area closure, catch rates significantly lowered, which corresponded to a strong interference competition between trawlers. They also showed that the fishermen with specialization in the closed area experienced lower catch rates than fishers with previous experience in the open areas, and were more likely to stop fishing during the closed period.

Rijnsdorp *et al.* (2011) studied the dynamics of exploitation of fishing grounds based on data of individual tows for a sample of the Dutch beam trawl fleet. In conformity with the marginal value theory, the exploitation of a fishing ground continued until the capture rate became lower than a certain threshold, called the giving up catch rate. However, the observed value of the giving up catch rate was inferior to the predicted value. This difference could have been explained by the bias introduced in fishermen behavior by an individual quota system. Poos *et al.* (2010b) used optimal foraging theory to examine the spatial behavior of beam trawlers observed by VMS. They studied the causes of the divergence between the observed spatial distribution of the fleet and the ones estimated by the ideal free distribution theory (Fretwell and Lucas, 1969). The results revealed potential segregation between two fleet segments with unequal competitors through spatial segregation in prey species, price differences, and interference competition through prey depression. In the ab-

sence of interference competition, the segregation between fleets showed dependence on differences in the distribution and prices of the prey species, combined with differences in catch efficiency.

Several other works, that are presented in the following section, analyze the relationships between fishermen, fish and the environment. However, most of them focused on the impacts of fishermen on the ecosystem; only the studies mentioned in this section focused on how fishermen respond to variation in ecological conditions. In this work (Chapter 6), we use a descriptive framework to assess how coastal environmental processes and prey conditions shape fishermen behavior. Although collective behavior is not addressed here, it should be considered in future works.

2.7.4 Contributions to fisheries management

Control and surveillance

Introduction of VMS in fisheries was mainly motivated by surveillance and control purposes: for monitoring foreign fleets in national exclusive economic zones (e.g., Japanese fisheries in Australian waters, international fleets off South Pacific countries), monitoring fishing activities in environmentally sensitive areas (e.g., Great Barrier Reef, Australia), monitoring the exclusion of industrial fishing fleet activities within the first 5 mn from the coast (e.g., Peruvian anchovy fishery), etc. This monitoring is mostly focused on the identification of illegal captures.

Willems *et al.* (2009) proposed a visualization method for other surveillance purposes: to know where significant maritime areas, like highways and anchoring zones are located. This visualization is based on density fields which are derived from convolution of the dynamic vessel positions with a kernel. A combination of two fields, with a large and small kernel, provides overview and detail. A large kernel provides an overview of area usage revealing vessel highways. In turn, a small kernel shows details of speed variations of individual vessels, highlighting anchoring zones where multiple vessels stop. Using GPS data on vessels off the Dutch coast, the authors investigated traffic patterns in different weather conditions (smooth and stormy weather), characterizing movement behavior in anchor zones and highways (highways with slow movers and highways with normal functioning were identified as well). Vessel surveillance would be greatly benefited from this tool.

Besides, control and surveillance, fishermen movement data has proven to be very useful for many management purposes.

Impact of trawling activity in the marine environment

Trawling activity has been seriously questioned because of its impact on the marine environment (Jones, 1992), directly (scraping the substrate, sediment resuspension, destruction of benthos) and indirectly (post-fishing mortality and long-term impacts on benthic communities). For that reason, a large number of studies used spatio-temporal behavior of fishermen for assessing impacts of the fisheries on the sea bottom. Various works elaborated indicators of fishing effort (e.g., Gerritsen *et al.*, 2013; Hiddink *et al.*, 2006a,b; Jennings and Lee, 2012; Jennings *et al.*, 2012; Marrs *et al.*, 2002; Piet *et al.*, 2000; Piet and Hintzen, 2012; Stelzenmuller *et al.*, 2008; Vinther and Eero, 2013), involving the number of times or hours a trawl passed through defined cell grids or areas in the map, the areas not fished, or the areas fished at specific intensities. In this respect, Piet and Quirijns (2009) showed that the absolute amount of effort and the choice of spatial and temporal scales determine the perception of fishing impact both in terms of the spatial distribution of fishing effort as well as estimated fishing-induced mortality. The accuracy of VMS data for estimating the distribution of fishing effort and impact was studied by Deng *et al.* (2005) and Skaar *et al.* (2011); they showed that for accurate estimation of fishing indicators at small spatial scales, fine resolution of VMS data are required. Jennings and Lee (2012) assessed how the choice of criteria for defining fishing grounds can influence their size, shape and location.

Information on benthic communities and environment has been used in several works to account for the fisheries impact in an ecological perspective. Hiddink *et al.* (2006a) proposed a size-based model to predict the effects of trawling on benthic biomass, production and species richness in grid maps throughout the main fishing grounds exploited by the North Sea beam trawl fleet. The model incorporates trawling effort computed on VMS data and environmental parameters of habitat type (sediment, depth, shear stress, chlorophyll-a). Hiddink *et al.* (2006b) developed indicators accounting for the interaction between trawling pressure and capacity for recovery of the benthic biomass. The size-based model from Hiddink *et al.* (2006a) was used for predicting the recovery time of these communities after trawling. They proposed a state indicator measuring the proportion of an area where benthic invertebrate biomass is greater than 90% of pristine benthic biomass; and a pressure

indicator measuring the proportion of the area where trawling frequency is sufficiently high to prevent biomass from reaching the 90% of pristine benthic biomass. Hiddink *et al.* (2007) developed a method for assessing the sensitivity of seabed habitats to bottom-trawling disturbance; the method also takes account of the effects of natural disturbance on habitat characteristics. Vinther and Eero (2013) developed an approach for obtaining proxies for changes in fishing mortality (based on effort information), spatial modeling of species distributions (based on trawl surveys) and a retention probability available in the literature.

For comparing the impact of the main human activities operating in England and Wales waters, Eastwood *et al.* (2007) provided an assessment of their direct physical pressure on the seabed. Pressure was estimated as the spatial extent of each activity. Among all activities, demersal trawling had the greatest associated pressure, ranging from 5% to 21%. Regarding large scale and long term impacts, Tillin *et al.* (2006) examined the impact of chronic trawling on the functional composition of benthic invertebrate communities, using information of fishing effort from VMS data, and information on the life history and ecological function traits of sampled taxa. Univariate and multivariate analyses were used for examining changes in the distribution of traits over gradients of trawling intensity. They demonstrated that chronic bottom trawling can lead to large scale shifts in the functional composition of benthic communities, with likely effects on the functioning of coastal ecosystems.

Complementary effort indicators

VMS-based effort indicators have been proposed in different fisheries. For example, Mullowney and Dawe (2009) developed a VMS-based fishing catch per unit of effort index (kg/fishing hour) and an index of efficiency (number of pot hauls per hour) in a snow crab fishery. They showed that VMS-based fishing effort and CPUE (catch per unit of effort) indices can be interpreted to provide reliable complementary or alternative indices to logbooks for assessment of fishery performance in the fishery. Furthermore, Bez *et al.* (2011) introduced several effort indicators for studying and quantifying the spatial dynamic of the tropical tuna purse seine fishing activity. For each activity at sea (searching, fishing and cruising), they computed: time spent (in each activity), spatial extension, spatial range (computed from variogram), Gini index (for homogeneity in time spent by activity by pixel) and a collocation index (as a measure of spatial correlation).

Evaluation of management measures and spatial marine planning

Fishermen movement analysis has also been used for evaluating the effects of different management measures. Hiddink *et al.* (2006a) developed a model for predicting the effects of area closures and effort reduction on the biomass, production and species richness on benthic comunities. Dinmore *et al.* (2003) studied the changes in the spatial distribution of effort of the North Sea trawl fleet when seasonal area closures occur. They showed that repeated seasonal closures could produce a more homogeneous or less 'patchy' distribution of fishing effort, when the effort moves into previously unfished areas. The authors therefore concluded that the repeated use of these seasonal closures could lead to a negative effect of increased trawling activity on the total production of benthic invertebrates in the North Sea. Gerritsen *et al.* (2012) investigated the effect of a seasonal closures in several regions with high cod CPUE on the number of catches and landings of cod by the Irish demersal otter trawl fish. Due to the joint analysis of VMS and logbooks, two areas are identified where effort is relatively low, but cod bycatch is high. Fishing closure in those areas was recommended, since it would demand relocating between 3% and 9% of the effort, while cod bycatch would be reduced in 8%-22%.

The effects of real-time closures (RTCs) has also been evaluated using data on fishermen movement. Holmes *et al.* (2011) presented the RTC system applied in Scotland for reducing cod capture. VMS data served as an input for deciding the area closures. They also used VMS data analysis to show that compliance with these closed areas was good, although the decline in catches of cod was not as large as expected. Needle and Catarino (2011) used the VMS data from Scottish vessels for constructing space-dependent relative cod-importance indexes (RCII), which were then used to determine whether the areas to which vessels move have a higher or a lower RCII, and how far away they move when an RTC is activated. They showed that the RCII of the areas vessels moved to tended to be lower than that of the closed ones, and that vessels traveled farther when moving away from a closure than when moving back after reopening. It evidenced that RTCs may impact beneficially on cod mortality.

Marine protected areas (MPAs) are a common type of spatial closure given by management programs for protecting sensitive marine habitats and associated fauna. Spatial information on the fisheries is key for MPA design and efficiency (Costello *et al.*, 2010). For improving spatial planning, various works analyzed the changes in fishing effort that would generate potential MPAs. For a local fishing in

Solomon Islands, Aswani and Lauer (2006) use a geographical information system (GIS) database to incorporate socio-spatial information, such as indigenous knowledge and artisanal fishing data, along with biophysical and other information to assist in MPA design. Converting peoples' knowledge and socio-ecological behavior into geo-spatial data allowed researchers to formulate hypotheses regarding human responses to inter- and intra-habitat variability, along with other marine ecological processes, and helped in the designing and implementation of resource management strategies in a cost-effective and participatory way.

Shephard *et al.* (2012) combined fisheries-independent survey and fine-scale VMS data, for modeling elasmobranch biomass as a function of environmental variables and local fishing effort. They evidenced that spatial heterogeneity in fishing effort can lead to a temporally stable mosaic of fish and unfished areas that would generate *de facto* refugees for elasmobranches in the Celtic Sea. Such refugee may represent sites where establishment of formal MPAs would result in minimal fishing effort displacement. Nonetheless, potential MPAs in other zones of the Celtic Sea could cause effort to shift to the *de facto* refugee zones which would make them stop being refugee zones. Therefore, they advised to consider setting an MPA in the zone.

Jennings *et al.* (2012) used VMS data for measuring fishery footprints and assessing habitat sensitivity and trawling impacts in an area of the North Sea where marine spatial planning was underway and a network of MPAs had been proposed. Interannual and fleet-related differences in the distribution and intensity of trawling activity, driven by location choice and fisheries regulations, had more influence on overall trawling impacts than the exclusion of beam trawlers from a proposed network of MPAss. They concluded that direct management of trawling footprints has potential to support the achievement of environmental objectives efficiently and at low costs. Maiorano *et al.* (2009) presented a systematic conservation planning in the Mediterranean context for the identification of no-take MPAs. They used trawl survey data for obtaining information on juveniles and spawners of commercial species, and VMS data for information on the spatial distribution of the fleet. Their framework aimed at optimizing both the conservation of species and the economic activity.

For assessing the consequences of a MPA imposed for over a decade, Murawski *et al.* (2005) evaluated the spatial distribution of otter trawl fishing effort and catches off North East USA. The positions of the vessels were documented by logbooks and VMS. The high resolution of the VMS data allowed observing CPUE gradients

around the MPA, suggesting a positive impact of the MPAs. The analyses confirmed that large-scale year-round closed areas affect the abundance and spatial distribution of some target species, and the allocation of trawling effort. For evaluating potential impacts of spatial planning measures, Fock (2008) and Pedersen *et al.* (2009) presented methods for mapping fishing effort and catches combining VMS and logbook data. They allowed exploring the spatial and temporal variability of fishing activities in the German Exclusive Economic Zone.

Ecosystem approach to fisheries

The availability of VMS data has opened great perspectives for an ecosystem approach to fisheries (EAF; Browman and Stergiou, 2004; Garcia and Cochrane, 2005). The impact of the fisheries in the marine environment and most of all, on the marine organisms, has been extensively studied, as we have shown in the examples above. An important issue that is not always considered is the interaction of fishermen with other marine top predators. Catry *et al.* (2013); Granadeiro *et al.* (2013, 2011) studied the spatial and temporal extent of the interactions between individual albatross and fishing vessels, using VMS of the Falklands fishing fleet, seabird GPS tracking, blood and feather samples on the birds. In general, they found little time-space overlap between the vessels and the birds; although a few individuals repeatedly visited the vessels. The isotopic analyses from blood and feather samples suggested no specialization of individual albatrosses with regard to fisheries. However, because of the presence of a few individuals that follow the vessels, the authors recommended management actions leading to a reduction of discards for reducing the risk of incidental mortality.

Sonntag *et al.* (2012) addressed the issue of bird bycatch in set-nets in the southern Baltic Sea. They used a bird-counting database and VMS records from the same period of time (2000-2008). A spatial overlap between set-net fishing activities and diving birds was used to indicate potential conflicts between the two. Conflict showed to be higher during winter and spring in coastal waters and around shallow offshore grounds. Local bycatch studies validated the usefulness of their approach, which can provide a valuable tool for conservation purposes. Torres *et al.* (2011) used GPS tracking data from foraging trips of albatross within sub-Antarctic New Zealand and VMS-based fishing effort distribution data for quantifying fine-scale overlap between individual albatrosses and individual vessels and characterizing behavioral changes in albatrosses when they are associated with the vessels. High

variability in the foraging destinations of albatrosses and in association rates with the fishing activity was found. Foraging behavior patterns of albatrosses relative to commercial fishing activity revealed that birds moved in straighter paths and at slower speeds when overlapping a squid trawler than when foraging naturally, probably for maintaining the vessel's heading and pace. The fact that albatrosses adjusted their foraging behavior when associated with fishing vessels, suggested that the ecological consequences of fishing activities on seabirds that interact with fishing gear, may extend beyond the direct effects of mortality and injury.

Votier *et al.* (2010) investigated the behavioral responses of gannets to trawlers using GPS tracking on gannets and VMS data on the trawlers near a UK island where the colonies of gannets were located. Analysis of conventional diet samples, as well as stable isotope ratios of carbon and nitrogen in blood evidenced marked individual differences in the proportion of fishery discards in the diet; although fisheries waste did not form the majority of prey for breeding gannets, it formed a significant component of the diet for certain individuals. They also showed that gannets adjust at-sea path tortuosity and flight speed in relation to fishing vessel positions, with evidence indicating that the gannets flied towards the vessels. Overall, linking GPS tracking data with VMS data indicated that fishing vessels shaped the at-sea foraging behavior of seabirds. The authors highlighted the potential of VMS data for improving ecosystem monitoring and management. Some of the monitoring and management measures they state that should be prioritized are: identifying vessels with particularly high rates of scavenger co-occurrence and therefore entanglement risk, giving priority to conservation policies of bycatch reduction, and ensuring that scavenging species have sufficient food to meet their energetic requirements in the absence of fishery waste, to reduce possible indirect ecosystem effects.

Tew Kai *et al.* (2013) studied the foraging behavior of Cape gannets breeding off the coast of South Africa using high-resolution GPS tracking, in relation to (1) pelagic fish availability assessed through acoustic surveys, and (2) VMS-based fishing effort by pelagic purse-seiners that compete with seabirds for fish and demersal trawlers that discharge fish waste. They found substantial inter-annual variability in spatial use by breeding gannets, which was driven primarily by pelagic fish availability. At the mesoscale, birds and purse seiners exploited similar marine areas, but no fine-scale dependence of birds on purse seiners was detected. It was also found that birds only sought trawlers when pelagic fish availability was low, suggesting reversible dependency upon fishery waste. They thus highlighted the necessity to

promote sustainable fishing allowing the restoration of pelagic fish stocks.

Focused on competition for prey, Bertrand *et al.* (2012) studied Peruvian booby and Peruvian anchovy fishery interactions using tracking data on seabirds and VMS data; both the fishing season and the tracking experiment with the birds started on the same day. The authors show that if in the early days of the fishing season the birds and vessels share the same areas, birds then dissociate from the vessels and undertake trips increasingly distant from the colony. This increase in foraging effort was significantly related to the growing quantity of anchovy removals by the fishery (from landings data). Daily removals by the fishery were at least 100 times greater than the daily anchovy requirement of the seabird colonies. The authors concluded that the overcapacity of the fleet during the year of study could have led to local depletion of anchovies, forcing the birds to increase their foraging effort in more remote areas for compensating. They recommended the implementation of areas of temporary fishing closures during the core of breeding seasons and around the main colonies.

Other studies (Bertrand *et al.*, 2008b; Sharples *et al.*, 2013a,b), using VMS data for elaborating integrated analyses of marine ecosystems and therefore contributing to the EAFs, were already described in section 2.7.3.

2.7.5 Remaining challenges

Vessel monitoring system data has been available for management and scientific purposes in many fisheries from the beginning of the 2000s. In more than one decade, the use of this relatively high resolution data ($\sim$ 1 position per hour) on fishermen movement for fisheries science and management has progressively increased. Sections 2.7.2, 2.7.3, 2.7.4 intended to account for those advances. However, several challenges remain; they will be addressed here in the form of questions.

- **Can we trust fishing set identification? And what is the best method of inferring behavioral modes in fishermen trips?** We have shown that the highest interest in VMS data is the identification of fishing sets, for assessing the distribution of effective fishing effort in a second stage. Speed thresholds are still the most common criteria for classifying VMS records into fishing or not fishing positions (Lee *et al.*, 2010). The few works that quantified the

false positive and overestimation errors (Bertrand *et al.*, 2008c; Palmer, 2008) demonstrated that the speed criteria seriously overestimates the number of fishing sets. Many of the multiple indicators used in fisheries management are based on those classifications (e.g., for estimating trawling impact in seabeds) and may thus be biased. Hence, it is important to: (1) account for inference errors, and (2) choose a method that gives the best inference performance. Discriminative (e.g. artificial neural networks) and Markovian models (e.g. HMM), can be used for inferring not only fishing behavior, but several other types of behavioral modes occurring throughout the fishing trips. Indeed, other behavioral modes, such as searching, provide information of a certain degree of foraging effort. In this work (Chapter 4), we address this challenge by comparing several discriminative and Markovian models (both using supervised learning) and using independent groundtruthed data for their validation. We evaluate how well do the models infer three behavioral modes: fishing, searching and cruising.

- **Can fishermen trips be classified in strategic groups?** The analyses of fishermen movement have rarely addressed the heterogeneity in movement and how this heterogeneity could obey to different gear, strategies or external factors. Russo *et al.* (2011b) showed that different métiers could be reflected in different distributions of speeds and turning angles, among others. Bertrand *et al.* (in review) showed that different segments of the Peruvian fishing fleet differ in their diffusive behavior. In this work (Chapter 5), we examine the heterogeneity in fishing trips based on a series of descriptors (e.g. duration, distance traveled, time spent in each behavioral mode) and their association in different strategic groups.

- **How do ecosystem scenarios condition fishermen movement and behavior?** Fishermen do not only impact the ecosystem; they are also conditioned by it. Evidence of the similarities in fishermen foraging behavior to those of other animals has been already shown (Bertrand *et al.*, 2007). However, there are very few works addressing the influence of the ecological conditions on fishermen; instead, most works account only for the changes in fishermen behavior due to changes in management policies, as if it was the only factor affecting fishermen behavior. Instead, within an ecosystem approach to fisheries, it would seem natural to account for the effect of changes in ecosystem in fishermen behavior.

In this work (Chapter 6), we analyze and quantify the associations between fishermen, their prey and the environment; and describe how these two fish and coastal environmental conditions shape the movement of Peruvian anchovy purse-seiners.

- **Does the spatial distribution of fishermen effort reflect the spatial distribution of fish?** Thanks to VMS data, spatial distribution of effort is being intensively studied. However, and even though effort in commonly used as a proxy of abundance, the spatial associations between fishermen effort and fish distribution and abundance has been assessed in very few works. In Chapter 7, we analyze the co-variation and co-occurrence between fishermen spatial behavior and the spatial distribution of acoustic biomass of Peruvian anchovy.

- **Should descriptors of fishermen behavior be used as effort indicators?** As we described in section 2.7.4, many indicators of effort have been elaborated using VMS data. However, they are only based on fishing sets. Results from movement analysis have only been rarely incorporated into fisheries management. The work of Bez *et al.* (2011) may constitute one of the only examples using spatial and temporal information on behavioral modes for effort indicators. Random walk related works allow accounting for diffusive behavior and explored spatial range. But, among all possibilities, which indicators should be preferred? What insights could they offer for fisheries management? Although these questions have not been extensively studied in this work, we do offer important clues when showing in Chapter 6 the potential of some fishermen descriptors as ecosystem indicators, when examining their response to ecosystem scenarios.

- **How can fishermen collective behavior be studied? How to analyze competition and association?** VMSs have enabled access to rich data on fishermen movement (i.e., numerous trajectories from numerous individuals foraging at the same time), allowing to account for collective foraging behavior in a free environment. Despite first approaches that evaluate interference competition (Poos *et al.*, 2010a; Poos and Rijnsdorp, 2007), many issues about collective behavior remain unresolved. For example, how to distinguish between competition and association? At which scales are association and competition patterns observed? Observing the collective dynamics at a spatial scale (i.e., defining grid cells) would depend on the spatial scale (the size of the cells) and temporal scales fixed. It could be considered to use the

approaches investigated in animal movement. While wavelet analyses would be difficult to use since large time series are needed, the exploration of approaches based on self-propelled particles would be appealing. This challenge is not addressed in this work, however it should be examined in future works. Collective behavior analysis is not only attractive for behavioral science; it is also important for fisheries (for example, for evaluating collective responses to area closures or effort reduction policies).

- **How do fishermen interact with other marine top predators?** The EAF involves the interaction of fishermen with other marine top predators. The studies described in section 2.7.4 represent the first attempts for assessing fishermen impact in seabird diet, foraging behavior and by-catch mortality, using spatialized data on fishermen. However, an intensive monitoring of seabirds and marine mammals, and a deeper knowledge on fishermen impact in their behavior and mortality should be achieved. This challenge is beyond the scope of this work. Nonetheless, it is part of a collaborative work within the TOPINEME (Top predators as indicators of exploited marine ecosystem dynamics) project.

Chapter 3

The northern Humboldt Current system: components and dynamics

"No es la última ola con su salado peso / la que tritura costas y produce / la paz de arena que rodea el mundo: / es el central volumen de la fuerza, / la potencia extendida de las aguas, / la inmóvil soledad llena de vidas. / Tiempo, tal vez, o copa acumulada / de todo movimiento, unidad pura / que no selló la muerte, verde víscera / de la totalidad abrasadora."

– Pablo Neruda (*El Gran Océano*)

3.1 Introduction

Four major upwelling systems border the west sides of the continents, comprising the eastern parts of the oceans: the Humboldt Current system in the South Pacific, the Canary Current system in the North Atlantic, the Benguela Current system in the South Atlantic and the California Current system in the North Pacific (Fig. 3.1). Those four Eastern Boundary Upwelling systems (EBUS) represent $\sim$ 0.3% of the world surface oceans (Carr and Kearns, 2003) but produce about 20% of the world's fish catches, contributing significantly to securing food and livelihood strategies in many developing countries (Fréon *et al.*, 2009). In these regions, regular trade winds combined with the earth's rotation generate coastal upwelling, bringing cold, nutrient-rich water from the deep ocean ($\sim$ 200-300 m) to the surface. The arrival of this water to the sunlight-exposed surface layer fuels primary production, which supports a highly productive food web (Fréon *et al.*, 2009). It also contributes very significantly to gas exchanges between the ocean and the atmosphere, particularly CO_2, which makes EBUS particularly sensitive to climate change (Chavez and

Messié, 2009).

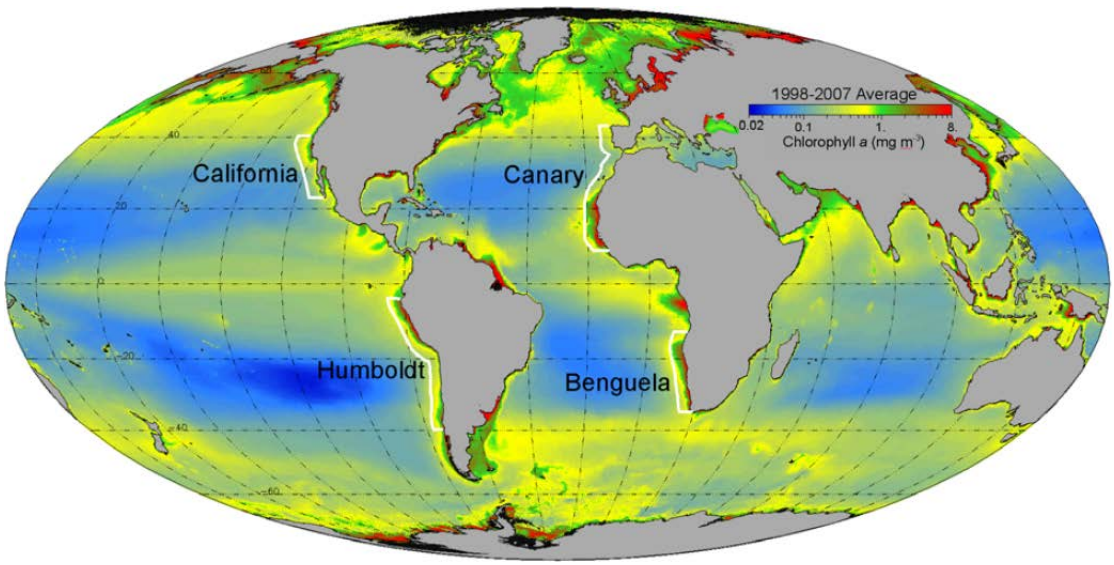


Figure 3.1: Chlorophyll-a concentration from SeaWiFS for the 1998-2007 period and EBUSs location arbitrarily delimited by the 200 nm offshore limit and latitudinal extensions including seasonal upwelling zones (courtesy of H. Demarcq, IRD, France). Source: Fréon *et al.* (2009)

The Humboldt ecosystem is the EBUS with the strongest interannual and interdecadal variability that affects its biological and abiotic components, ecosystem processes, and fisheries yield, in addition to the variability occurring at seasonal scales (Montecino and Lange, 2009). The Humboldt Current system has the weakest upwelling winds of all EBUS, and the greatest average upwelled volume. It also has the shallowest, more oxygen-depleted and more extended Oxygen Minimum Zone (Chavez and Messié, 2009, Fig. 3.2). As stated by Fréon *et al.* (2009), despite these apparently detrimental effects, and the fact that this ecosystem does not have the highest levels of primary production, it provides the highest contribution to fish production, mainly due to a single species: the Peruvian anchovy or anchoveta (*Engraulis ringens*), making the Peruvian anchovy fishery the world's largest monospecific fishery (average annual catch of 6 809 492 tonnes between 2000 and 2011; FAO, 2013b).

In this chapter we will briefly describe the main features of the Humboldt Current system, with a stronger focus in the Northern Humboldt Current system (NHCS), its oceanography, biology and fishery.

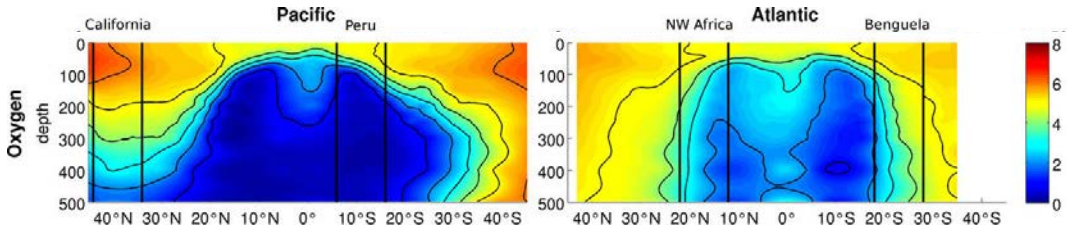


Figure 3.2: Northsouth section of oxygen in the Pacific and Atlantic Oceans following the coast 1000 km offshore of the eastern boundary. Location of the 10°-latitudinal bands chosen for comparison within each EBUS are shown in black. Data are from the World Ocean Atlas 2005 (Garcia *et al.*, 2006a,b). Source: Chavez and Messié (2009)

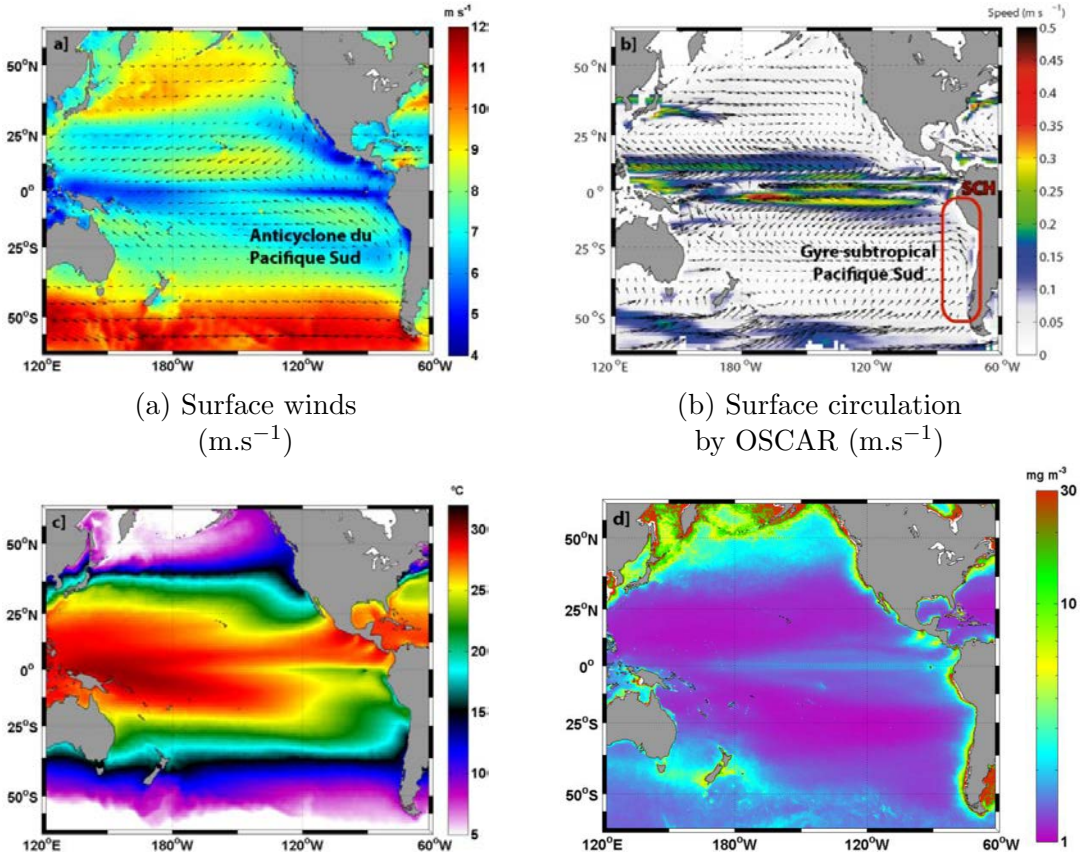
3.2 Oceanographic context

3.2.1 South Pacific gyre and oceanic circulation

Like other subtropical regions of the world, the South Pacific Ocean is surmounted by an area of high atmospheric pressures: the South Pacific subtropical anticyclone, which generates surface winds rotating counterclockwise around the South Pacific (Fig. 3.3a). This wind system carries the surface waters by friction, giving rise to an anticyclonic ocean circulation, known as the South Pacific subtropical gyre (Fig. 3.3b).

Its eastern flank comprises the equatorward Peru Oceanic Current (POC). In the offshore ocean off the Ecuadorian and Peruvian coasts, this current feeds the South Equatorial Current (SEC) that flows westward in the nearsurface layers (Fig. 3.4a). The SEC is forced by the permanent easterlies that pile up surface water toward the western equatorial Pacific creating a zonal eastward pressure gradient force that drives, in subsurface layers, the eastward flowing Equatorial Undercurrent (EUC) centered along the equator (Fig. 3.4b) and reaching the Galapagos Islands near 90°W. East of the archipelago, the EUC separates into two branches, one branch flows southeastward to reach the Peruvian coast at $\sim$ 5°S while the other branch remains trapped along the equator.

Below the thermocline and further South, are found the primary and secondary Southern Subsurface Countercurrents (SSCCs), also referred to as Tsuchiya jets, that flow eastward and enter the NHCS along nominal latitudes of $\sim$ 5°S and $\sim$ 7°S, respectively. Near the Peruvian coast, the dominant alongshore equatorward winds and the cyclonic wind-stress curl lead to an intense upwelling characterized by an



(a) Surface winds ($m.s^{-1}$)

(b) Surface circulation by OSCAR ($m.s^{-1}$)

(c) Surface temperature by MODIS ($m.s^{-1}$)

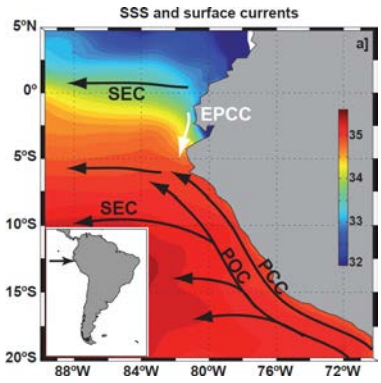
(d) Chlorophyll concentration by MODIS ($mg.m^{-3}$)

Figure 3.3: Main average characteristics of the Pacific Ocean. Source: Chaigneau (2013).

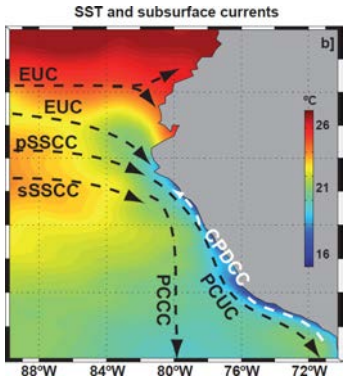
alongshore narrow strip of cold (Fig. 3.4b, color shading) and highly productive water and a current system composed of equatorward surface and mainly poleward subsurface flows. The equatorward surface circulation is composed of the Peru Coastal Current (PCC) that is mainly wind-driven, but also reinforced, through geostrophic adjustment, by the cross-frontal temperature (and density) gradient due to the upwelling (Fig. 3.4b, color shading). The subsurface poleward circulation is mainly composed of the Peru-Chile Undercurrent (PCUC) along the Peruvian continental shelf and slope (Fig. 3.4b) and a weaker secondary poleward flow, the Peru-Chile Countercurrent (PCCC), that flows south of 7°S and at 80°W - 85°W (Fig. 3.4b). North of 5°S, a near-surface coastal current flowing from Ecuador to Peru and associated with the surfacing of the EUC, has been suggested (Collins *et al.*, 2013; Lukas, 1986). This poleward surface-trapped current is called the Ecuador-Peru Coastal Current (EPCC; Chaigneau *et al.*, 2013, Fig. 3.4a). However close to the Ecuadorian coast, a northwestward oriented surface current (the Coastal Ecuadorian Current, not shown in Fig. 3.4a) can also take place (Allauca, 1990; Collins *et al.*, 2013). In the deep layer at 500 m depth, the average circulation is mainly northward, due to a relatively deep vertical extent of the POC in the offshore ocean and a deep equatorward near-coastal current, the Chile-Peru Deep Coastal Current (CPDCC; Chaigneau *et al.*, 2013, Fig. 3.4b), which flows below the PCUC and transports relatively fresh and cold Antarctic Intermediate Water northward.

Among the different currents that compose the NHCS, the PCUC is a key element because it advects seawater property anomalies from equatorial to extra-tropical regions and it plays a major role in the functioning of the NHCS ecosystem. The PCUC, that has been tracked along the continental shelf and upper slope from $\sim$ 5°S off Peru to $\sim$ 45°S off Chile, carries a relatively warm, salty, nutrient-rich, oxygen-poor and weakly stratified water-mass of near-equatorial origin (Silva and Neshyba, 1979; Tsuchiya and Talley, 1998). This water-mass, the Equatorial Subsurface Water (ESSW), flows southward into the PCUC and is the main source of the coastal upwelled waters in NHCS, promoting an intense primary productivity (Chavez *et al.*, 2008). Although the origin of the PCUC is still subject of scientific debate, it has been recognized that its main sources are the EUC and the two branches of the SSCC (Czeschel *et al.*, 2011; Montes *et al.*, 2010).

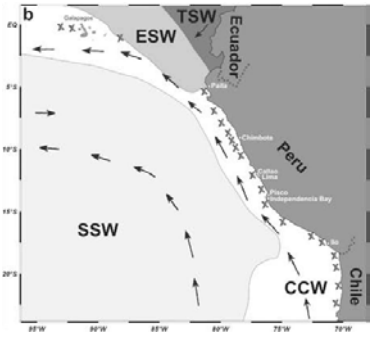
The presence of cold coastal waters at the surface has several consequences (Chaigneau, 2013). From a climate point of view, it stabilizes the overlying atmospheric layers and prevents the formation of convective systems and thus severely



(a) Sea surface salinity and surface currents.



(b) Sea surface temperature and subsurface currents.



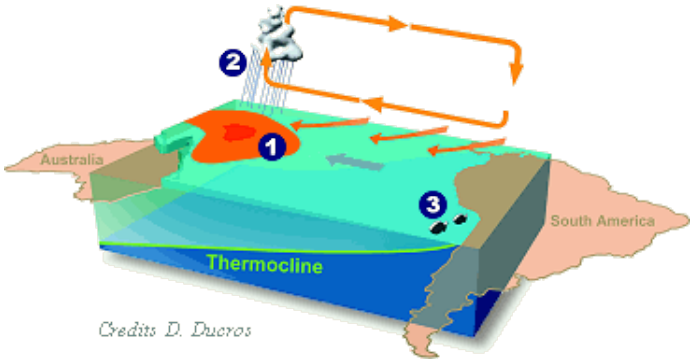
(c) Surface water masses

Figure 3.4: Schematic distribution of: the main currents, sea sea surface salinity and temperature (a and b); and the main water masses in the surface layers off Peru (c). The approximate locations for main upwelling areas for nutrient-rich waters are indicated with an 'X'. Currents: POC, Peru Oceanic Current; SEC, South Equatorial Current; EUC, Equatorial Undercurrent; pSSCCs and sSSCCs, primary and secondary Southern Subsurface Countercurrents, respectively; PCC, Peru Coastal Current; PCUC, Peru-Chile Undercurrent; PCCC, Peru-Chile Countercurrent; EPCC, Ecuador-Peru Coastal Current. Source: Chaigneau *et al.* (2013) (figures a and b) and Ayón *et al.* (2008a) (figure c).

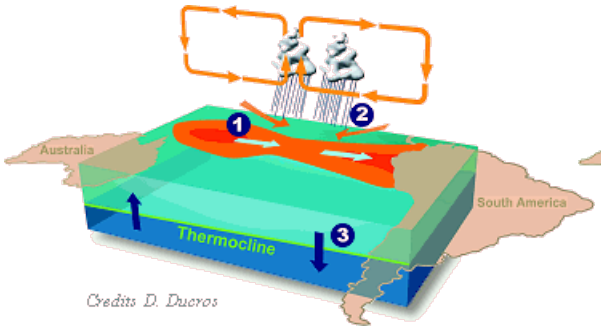
limits the precipitations. This coastal phenomenon reinforces the synoptic effect related to the subsidence of the subtropical anticyclone and thus the coasts of the north of Chile and south of Peru are among the driest in the world. From an ocean view, the surface temperatures observed in the coastal areas of the Humboldt Current System are the coldest in the world for similar latitudes; along the coast off Peru, they are more than 10°C lower than in west side of the Pacific Ocean (Fig. 3.3c). The strong temperature gradient between the coastal and the offshore areas increases currents forced by winds, giving rise to a relatively intense coastal surface current flowing to the north. From an ecological point of view, cold coastal waters rich in nutrients, by contact with the solar radiation in the euphotic layer, allow the development of an intense primary production (Fig. 3.3d). This primary production is the first link in the marine food chain that makes the Humboldt Current system the most productive region in the world. Thus, the marine resources play an ecologically, economically and socially vital role in the countries bordering this area. Finally, from a biogeochemical point of view, microbial degradation of organic matter associated with the strong biotic production consumes a large amount of dissolved oxygen. This reduces the oxygen concentration in subsurface waters which are already poorly ventilated, and gives rise to the presence of a thick anoxic layer that is close to the ocean surface (Fuenzalida *et al.*, 2009; Stramma *et al.*, 2010). This area, almost devoid of oxygen, is known as the Oxygen Minimum Zone (OMZ) and is a major source of greenhouse gas emissions (CO_2 and N_2O; Paulmier *et al.*, 2008) . The limit between the oxygenated surface layer and the OMZ, also forms a strong vertical barrier – the oxycline – that most living organisms cannot cross and thus restricts the vertical extension of the ecosystem in this region (Bertrand *et al.*, 2010, 2011).

3.2.2 Water masses in the surface layers

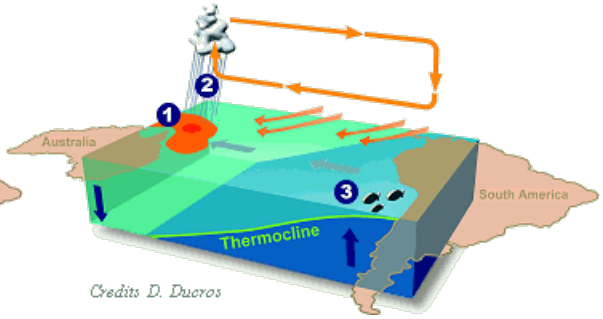
Regarding water masses (Table 3.1; Fig. 3.4c), a wide region off the Peruvian coast is dominated by Cold Coastal Waters (CCW), which are strongly influenced by coastal upwelling (upwelled waters originate from the PCUC; Echevin *et al.*, 2004a,b). The oligotrophic Subtropical Surface Waters (SSW) are located offshore the CCW. The Tropical Surface Waters (TSW) are found North of the Equator and characterized by higher temperatures and lower salinities than the other water masses present off Peru. The Equatorial Surface Waters (ESW) are found between the CCW and the STW.



(a) Normal conditions



(b) El Niño conditions



(c) La Niña conditions

Figure 3.5: Pacific Ocean - atmosphere interactions in the equatorial region in normal (a), El Niño (b) and La Niña (c) conditions. The numbers in each subplot correspond to different components in the interactions: (1) winds, (2) atmospheric convection and rain, (3) thermocline. The horizontal arrows are related to sea level variations while the vertical arrows are related to thermocline depth variations. Source: AVISO (2013)

Table 3.1: Main characteristics of the water masses present in the Humboldt Current system

Water mass	Salinity	Temperature	Ecological Type
Cold Coastal Water (CCW)	$[34.80; 35.05]$	$[15°; 17°]$ in winter $[15°; 19°]$ in summer	coastal productive
Subtropical Surface Water (SSW)	> 35.10	$[17°; 25°]$ in winter $[20°; 25°]$ in summer	oceanic oligotrophic
Tropical Surface Water (TSW)	$[34.00; 34.80]$	$[20°; 26°]$ in winter $[21°; 26°]$ in summer	mesotrophic
Equatorial Surface Water (ESW)	< 34.00	$> 23°$ in winter $> 26°$ in summer	oligotrophic

Source: Bertrand *et al.* (2004a)

3.2.3 Oxygen minimum zone

Oceans include vast areas called oxygen minimum zones where subsurface layers are depleted in dissolved oxygen (DO; Helly and Levin, 2004), resulting from the sinking and decay of surface-derived high primary production and poor ventilation. The OMZ forms a barrier to some animals, concentrating living resources near the surface. At the other end of the spectrum different forms of marine life have adapted to this harsh environment, some utilizing it as a refuge from predation. The OMZ also affects global nutrient budgets, as nitrate instead of oxygen is used by bacteria as a terminal electron acceptor (Chavez *et al.*, 2008).

Among the EBUS, waters off Peru are the oldest and least ventilated (Chavez and Messié, 2009). The OMZ reaches a maximum thickness of > 600 m, which extends about 1000 km into the open ocean, between 5°S and 13°S. The OMZ is the shallowest there, reaching depths of 150 m (Fig. 3.6 Fuenzalida *et al.*, 2009).

3.3 Biological context

3.3.1 Primary production

The distribution of phytoplankton biomass roughly follows a decreasing gradient from the coast towards offshore (biovolumes larger than 3 $ml.m^{-3}$ inshore) . Pri-

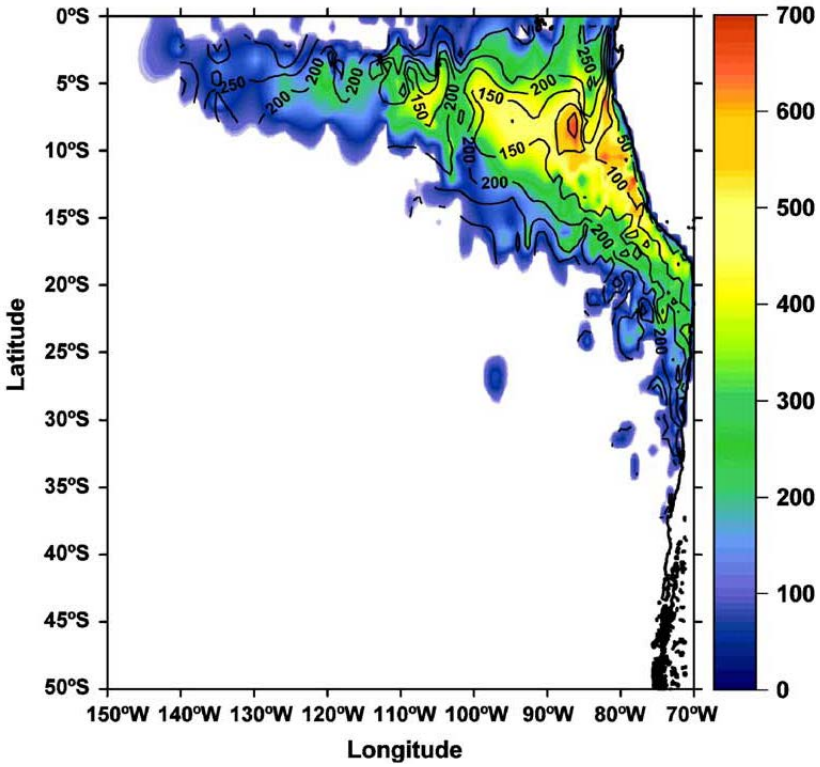


Figure 3.6: Oxygen minimum zone thickness and upper boundary in the Humboldt Current System. Thickness is color-coded according to the color bar on the right-hand side of the figure; units are in m. The upper boundary is shown in black contour lines with 50 m intervals. Source: Fuenzalida *et al.* (2009)

mary production is higher in austral summer and lower in austral winter (Echevin *et al.*, 2008).

Concerning phytoplankton communities, diatoms (microplankton) dominate the coastal upwelling zones, while dinoflagellates and coccolithophores (nanoplankton) dominate offshore in less turbulent zones (Sánchez Ramirez, 2000). Because of their marked preferences in terms of temperature and salinity, some species of dinoflagellates are commonly used as biological indicators of water masses (Arntz *et al.*, 1996; Delgado *et al.*, 2001; Sánchez Ramirez, 2000): *Protoperidinium obtusum* for cold coastal waters; *Ceratium breve*, *Ornithocercus steinii*, *Ornithocercus thumii* and *Amphisolenia Thrinax* for equatorial surface waters; and *Ceratium praelongum* and *Ceratium incisum* for sub-tropical surface waters.

3.3.2 Secondary production

The distribution of zooplankton biomass may be classified in three regional groups (Santander, 1981): (1) a continental shelf group dominated by *Acartia tonsa* and *Centropages brachiatus*; (2) a continental slope group characterized by siphonophores, bivalves, foraminifera and radiolaria; (3) and a species-rich oceanic group. The highest zooplankton abundances and biomasses have been often found between 4°S - 6°S and 14°S - 16°S, where continental shelves are narrow. Species composition changes with distance from the shore. Both species composition and biomass are highly variable at several time scales (seasonal, inter-annual and multi-decadal scales; Ayón *et al.*, 2008a).

Ballón *et al.* (2011) developed an acoustic method for estimating macrozooplankton biomass in the NHCS and estimated a biomass two to five times higher than previous estimates. The high biomass estimated is consistent with the trophic ecology findings indicating that forage fish mainly consume macrozooplankton (Espinoza and Bertrand, 2008; Espinoza *et al.*, 2009); it is also consistent with the current hypotheses (intermediate wind intensity, near-equator location and El Niño) explaining the NHCS high fish production (Bakun and Weeks, 2008; Chavez *et al.*, 2008).

3.3.3 Higher trophic levels

Anchovy

Anchovies are from the order of Clupeiformes and from the family *Engraulidae*. In Peru, there are two species, *Anchoa nasus* or anchoveta blanca, and *Engraulis ringens* or anchoveta (Fig. 3.7). *Engraulis ringens*, which we will henceforth refer to as anchovy, is the main fishing resource (average annual catch of 6 809 492 tonnes between 2000 and 2011, compared to the 15 632 tonnes of *Anchoa nasus* and the 7 381 681 of marine fishes in general for the same period of time; FAO, 2013b). Anchovy has a maximum life span of 4 years and may reach a maximum size of 20 cm. It experiences a fast growth and an early maturity (at 1 year and measuring approximately 12 cm). Therefore, individuals of less than 12 cm are considered as juveniles by the fisheries. Anchovy mainly feeds on zooplankton, although it may also prey on phytoplankton (they represent $\sim$ 98% and 2% of the food energy, respectively; Espinoza and Bertrand, 2008). According to Espinoza and Bertrand (2008), the fact that anchovy is able to feed from several trophic levels, choosing energetically advantageous food types, and to adjust its foraging period and duration to prey availability, represents an evolved adaptive strategy.

Figure 3.7: Peruvian anchovy or *Engraulis ringens*.

Anchovy is also flexible for reproduction, since they are able to spawn all year-round, with peaks between July and October, and at the whole latitudinal range off Peru (Bouchon *et al.*, 2010a). Other adaptive strategies include the ability to track and concentrate in refuge areas when conditions are adverse (Bertrand *et al.*, 2004a) and to spatially distribute its population over a large temperature range (Bertrand *et al.*, 2004a; Gutiérrez *et al.*, 2008).

As stated by Espinoza and Bertrand (2008), these combined characteristics may explain the 'anchovy paradox: how a fish which (i) performs very small migrations and cannot escape adverse conditions, (ii) is mainly distributed in dense surface aggregations and is thus highly accessible to predators (fish, cephalopods, birds, mammals and fishers), and (iii) is very slow in its avoidance reactions to predators (Gerlotto *et al.*, 2006), can achieve such enormous biomass in a relatively short time.

In addition, Bertrand *et al.* (2011) hypothesized that oxycline depth also plays a important role for anchovy: a shallow oxycline provides a large offshore low oxygen habitat to anchovy, allowing it access to the high concentrations of macrozooplankton at the shelf break and reducing competition for food and predation by 'expelling' species that cannot survive in low oxygen conditions.

Other marine species

Sardine (*Sardinops sargax*) has been very abundant and supported an important part of the pelagic fishery during the 1980s (Alheit and Ñiquen, 2004), but it has been almost completely absent from the NHCS since the early 2000s (Gutiérrez *et al.*, 2007). The crash of the sardine population can be attributed to the synergetic effect of trophic structure and oxygen; these environmental effects were most likely aggravated by overfishing (Bertrand *et al.*, 2011, 2004a).

Other pelagic species with important presence in the NHCS are chub mackerel (*Scomber japonicus*) and jack mackerel (*Trachurus murphyi*). Finally, three other species are likely to be encountered in large quantities particularly during population explosions: pelagic squat lobster (*Pleuroncodes monodon*), jumbo squid (*Dosidicus gigas*) and mesopelagic Panama lightfish (*Vinciguerria lucetia*).

Top predators

The top predators other than the jumbo squid, include fishermen, seabirds and pinnipeds. Pinnipeds are mainly South American sea lions (*Otaria flavescens*) and South American fur seals (*Arctocephalus australis*). The main populations of seabirds producing guano are the Peruvian booby (*Sula variegata*), Guanay Cormorant (*Phalacrocorax bougainvillii*) and less abundant Peruvian Pelican (*Pelecanus thagus*). They have shown remarkable levels of population, reaching approximately 25 million individuals in the 1950s. They differ in their foraging strategies. Pelicans forage mainly at night and seize prey close to the sea surface using a sit-and-wait strategy (Zavalaga *et al.*, 2011). Peruvian boobies have good flying skills (thus few constraints for horizontal displacement and more exploration capability), but do not dive deep ($\sim$ 5-10 m). By contrast, cormorants rely more on group strategies for locating fish aggregations, and on their excellent diving skills ($>$ 60 m; Weimerskirch *et al.*, 2012).

3.4 Multiple scale dynamics

The Humboldt Current system is subject to bottom-up forcing at seasonal, inter-annual, multi-decadal, centennial and higher scales. The time scales are associated with proportional length scales (Mann and Lazier, 2006, Fig. 3.8). In this section, we briefly describe the main features at inter-annual and seasonal scales.

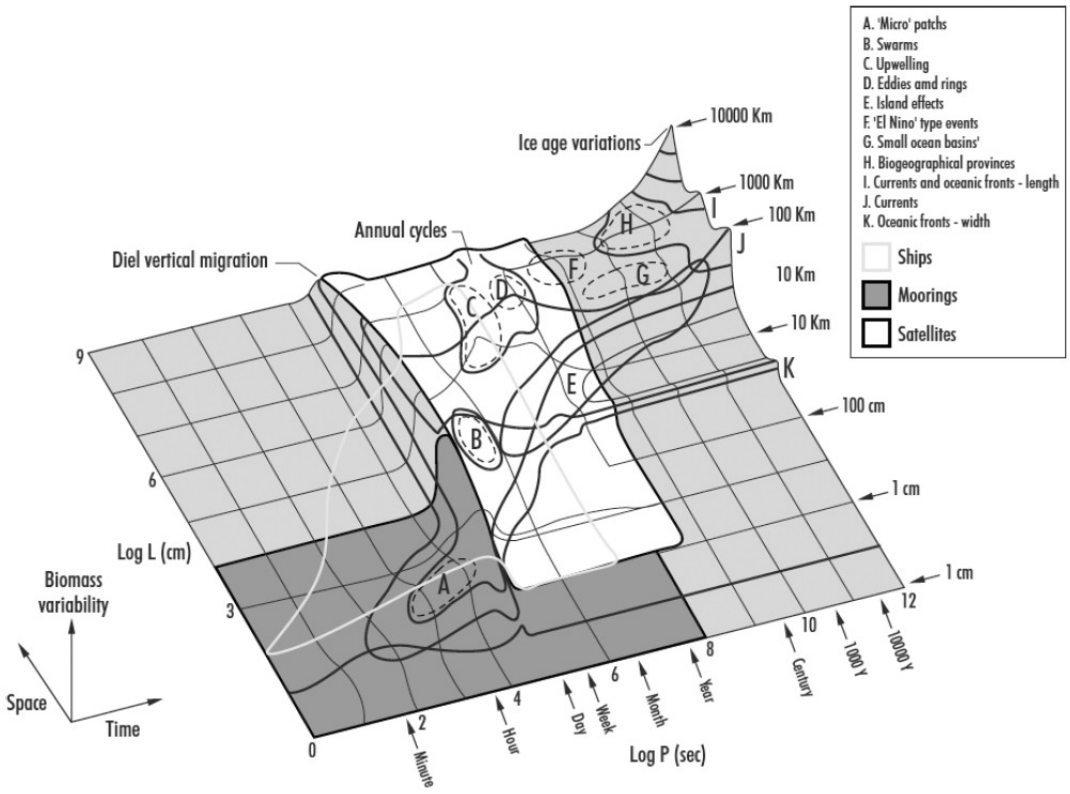


Figure 3.8: Stommel diagram of spatial and temporal scales of marine ecology dynamics and sources of data. Source: Kaiser (2005)

3.4.1 Centennial and multi-decadal variability

Large scale variability ($\sim$ 20000 last years) in terms of environmental conditions and fish abundance proxy have been recently studied (Salvatteci, 2012; Salvatteci *et al.*, 2012). Both components of the ecosystem presented very high variability at centennial and millennial scales. During the little ice age, there was little productivity; that shifted $\sim$ 1820 into better productivity conditions. Then, since ca.

1900, anchovy has reached its highest levels of productivity in this 20000 years. The high levels of anchovy productivity have been maintained except for the 1975-1995 time-period. In general, anchovy modulation appears to be strongly linked to primary productivity and oxygen. Moreover, the period of time where the fishery has been developed seems to be of exceptional abundance at the millennial scale.

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 Southern Oscillation (ENSO), but with longer duration: 20 to 30 years. During cooler periods, called La Vieja, the eastern Pacific exhibits a lower sea level slope and shallower thermocline, increasing the nutrient supply and productivity. This seems to favor the anchovy population (Chavez *et al.*, 2003), probably as a result of the expansion of the CCW (Bertrand *et al.*, 2004a; Swartzman *et al.*, 2008). During warmer periods, however, the thermocline deepens, the upwelling weakens and the productivity decreases, so that the range of habitat favorable to anchovy is dramatically reduced while the habitat favorable to sardine increases and spreads towards the continental shelf (Bertrand *et al.*, 2004a). However, this alternation in the abundance of anchovy and sardine has only been observed in the fishing records, covering only one cycle. When looking at longer records, such as paleo-fish scales from sediments, this alternation seems to occur only occasionally (Gutierrez *et al.*, 2009; Valdes *et al.*, 2008).

3.4.2 Inter-annual variability

The El Niño Southern Oscillation (ENSO) phenomenon is due to a periodic instability of the ocean-atmosphere dynamics in the Ocean Pacific basin. The frequency of ENSO is irregular and varies depending on the SST anomalies in the eastern equatorial Pacific or the Central Pacific (Eastern Pacific El Niño and Modoki El Niño, respectively Dewitte *et al.*, 2012). El Niño and La Niña events represent opposite cold and warm phases of the ENSO cycle, respectively. Extreme El Niño events have caused devastating floods in the East Pacific and droughts in the West Pacific. El Niño is Spanish for 'The Child', which refers to Jesus, because El Niño events near South America are usually noticed around Christmas.

Eastern Pacific El Niño conditions are generated by downwelling Kelvin Waves (KWs) that deepen the thermocline, making the coastal upwelling inefficient in

terms of nutrient enrichment, as the upwelled waters are warm and low in nutrients (Barber and Chavez, 1983, Fig. 3.5b). Conversely, La Niña conditions are generated by upwelling KWs that raise the thermocline and allow coastal upwelling to bring cold and nutrient-rich water towards the surface, making the cold coastal water to expand (Fig. 3.5c).

La Niña and, most of all, El Niño events condition the spatial organization of living organisms by modifying the volume of their favorable habitat. Under El Niño conditions, in the short term ($\sim$ 0-3 months), the extent of cold and nutrient-rich waters (CCW) is reduced and sea surface temperature increases in the coastal domain. With these changes, anchovy tends to distribute closer to the coast, remaining in the CCW (Bertrand *et al.*, 2004a; Swartzman *et al.*, 2008, Fig. 3.9), and deeper in the water column beneath the warm and less productive surface waters (Arntz *et al.*, 1996). Under these conditions, anchovy are highly concentrated in space (Gutiérrez *et al.*, 2007) and fishermen make lower catches and briefer trips (Bertrand *et al.*, 2008b). In the long term, however, it has been suggested that El Niño events contribute to maintaining the high fish production of the system by favoring fast-growing fish species like anchovy, which take advantage of the low predation pressure and rapidly increase their population and dominate the system (Bakun and Weeks, 2008)

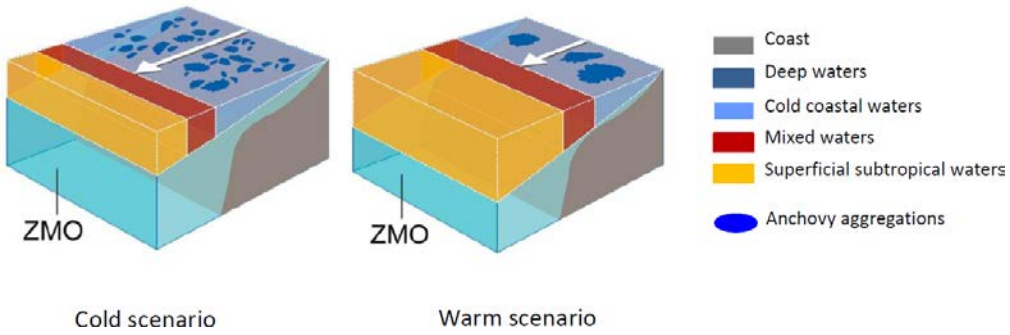


Figure 3.9: Water masses and anchovy distribution in cold and warm conditions. Source: S. Bertrand (pers.comm.)

The southern oscillation index (SOI) is a measure of the large-scale fluctuation 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 indicates an El Niño event while a positive index reveals a La Niña event (Fig.3.10).

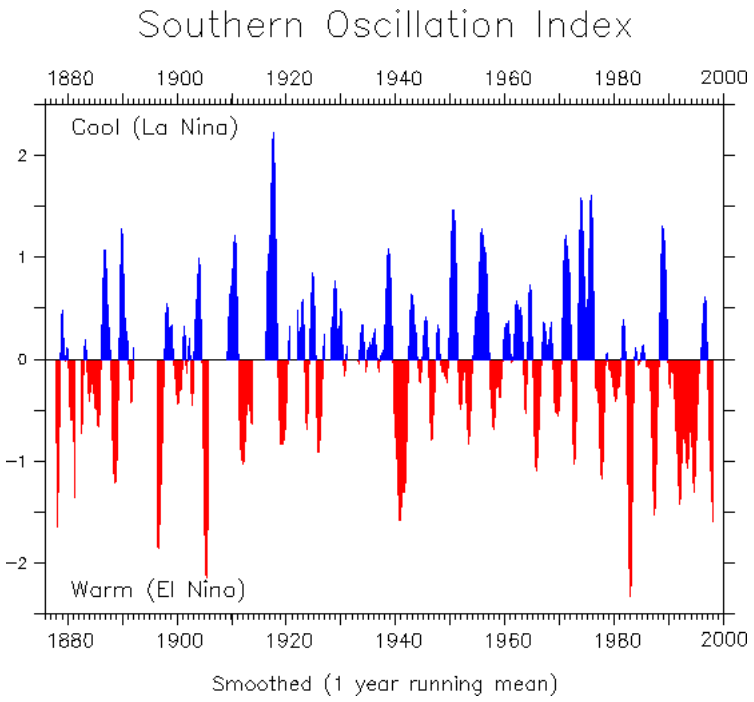


Figure 3.10: Southern Oscillation Index (SOI). Source: NOAA

3.4.3 Seasonal variability

In the Humboldt Current system, the upwelling is present all year round, but it varies in intensity between seasons. Seasonal fluctuations are also observed in other environmental variables such as the wind intensity, mixed layer depth, light, water masses, mesoscale activity, thermocline and so on. Although the highest upwelling intensity and nutrient supply occur during austral winter as a result of a higher wind intensity (Bakun and Mendelssohn, 1989), the system's maximum primary production occurs in spring (Echevin *et al.*, 2008; Thomas *et al.*, 2001) when a shallower mixed layer increases both the nutrient concentration and the light availability to phytoplankton (Echevin *et al.*, 2008).

3.5 Anchovy fishery context

3.5.1 History: the race for fish

The modern Peruvian fishery was developed during the 1940s and 1950s on species like bonito (*Sarda chilensis chilensis*) and tuna (mainly *Thunnus albacares*). It was

driven by the high demand for the liver oil of these species in the US market during World War II and later the Korean War (Chavez *et al.*, 2008). The anchovy fishery as well as the fishmeal production started in the 1950s (Aranda, 2009). In 1954, the Consejo de Investigaciones Hidrobiológicas (CIH) was created for coordinating hydrobiological studies with the goal of improving the use and sustainability of living marine resources. In 1959, the Instituto de Investigaciones de los Recursos Marinos (IREMAR) was created with programs dealing with oceanography, fisheries biology, biology of whales, fishery economics and fishery technology. Both entities merged in the Instituto del Mar del Perú (IMARPE) in 1964, which continues to conduct research on the NHCS and its fisheries today.

During the 1960s the fishing fleet grew steadily (Laws, 1997). The anchovy fishery continued to grow during the 1960s to a peak harvest of 12 million tonnes per year in 1970 accounting for 20% of the world catch (Chavez *et al.*, 2008). In 1972, the industry was hit by a particularly strong El Niño. This phenomenon together with overfishing produced the collapse of anchovy (Hilborn and Walters, 1992). As a consequence, the military government decided to nationalize the industry. In 1973, the state-owned company Pesca-Peru, decided to reduce the fishing and processing capacity and to forbid the building or renewal of fishing vessels (Laws, 1997). At that time, some boats targeted the sardine resource, other underexploited species like jack mackerel and horse mackerel, and the remains of the anchovy stock (Freón *et al.*, 2008).

In 1975, the fleet was privatized, while plants remained under state control. The anchovy stock started giving signs of recovery in 1981, but was hit again by a devastating El Niño in 1982-83 (Hilborn and Walters, 1992). Consequently, the population of anchovy was seriously depressed. While during El Niño 1972 and El Niño 1982-83 the anchovy experienced strong collapses, the sardine population and exploitation increased to become the most important pelagic fishery during the 1980s (Alheit and Ñiquen, 2004), reaching a peak of 3 million tonnes of landings in 1985 (Freón *et al.*, 2008).

The favorable long-term environmental conditions (La Vieja; Chavez *et al.*, 2003) prevailing since the beginning of the 1990s, however, led to a significant increase in anchovy abundance and catches during the 1990s. In contrast, sardine abundance and catches gradually decreased during this period, maybe as a result of unfavorable conditions and overfishing (Bertrand *et al.*, 2011, 2004a). A new administration

introduced neo-liberal economic policies and processing plants were privatized between 1992 and 2001 (CIDEF, 2002). Due to new policies and the recovery of the anchovy stock, the private sector found optimal conditions to invest in vessels and plant modernization and construction. As a result, the pelagic fleet capacity experienced a fast expansion (Aguilar Ibarra *et al.*, 2000). A strong El Niño occurred in 1997-98, but the anchovy experienced a rapid recovery; this event also marked the collapse of the sardine stock which has not yet recovered (Bertrand *et al.*, 2004a, Fig. 3.11).

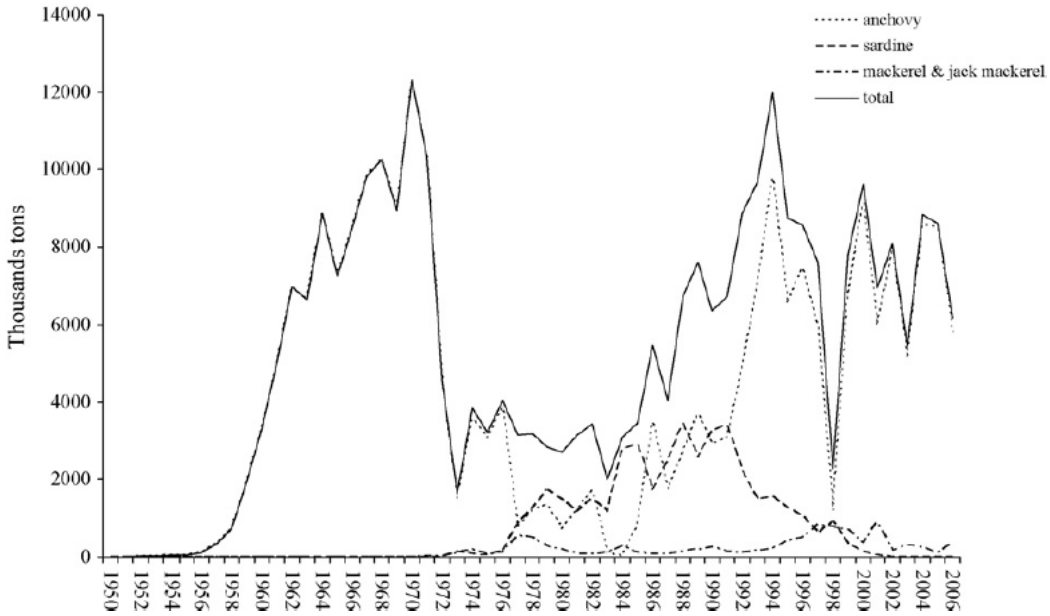


Figure 3.11: Landings of Peruvian pelagic fisheries 1950-2006. Source: PRODUCE (taken from Aranda (2009)).

Until 1998, there had been a single type of fishing vessel, all with steel hull but of different sizes ranging from 30 to 900 tonnes of holding capacity, mostly owned by large fishing companies. In that year, the government passed the Law 26920 which authorized owners of wooden vessels larger than $30m^3$ to join the pelagic industrial fleet, increasing the pressure on anchovy. In the 2000s, the investment and capacity increased, and annual landings of anchovy ranged from 5 to 9 millions tonnes (Aranda, 2009; Freón *et al.*, 2008, Fig. 3.11, 3.12). One of the most visible effects of overcapacity on fishing activity was a reduction in the annual number of fishing days. The fishing season of Peruvian anchovy decreased from $\sim$ 350 days in 1987 to $\sim$ 50 days in 2006 (Freón *et al.*, 2008, Fig. 3.13). In June 2008, the government decided to implement individual vessel quota allocations (IVQs) in the anchovy fishery. It aimed at stopping the race for fish, without allowing the transferability of rights.

The introduction of the IVQs had an immediate effect lengthening the annual fishing season and reducing the total number of operating fishing vessels (Tveteras *et al.*, 2011, Fig. 3.14, 3.15).

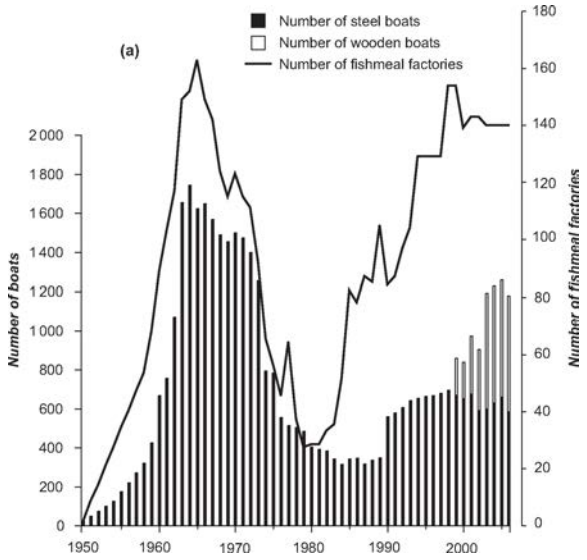


Figure 3.12: Number of purse-seiners and factories in the Peruvian Pelagic fisheries from 1950 to 2006. Source: Freón *et al.* (2008)

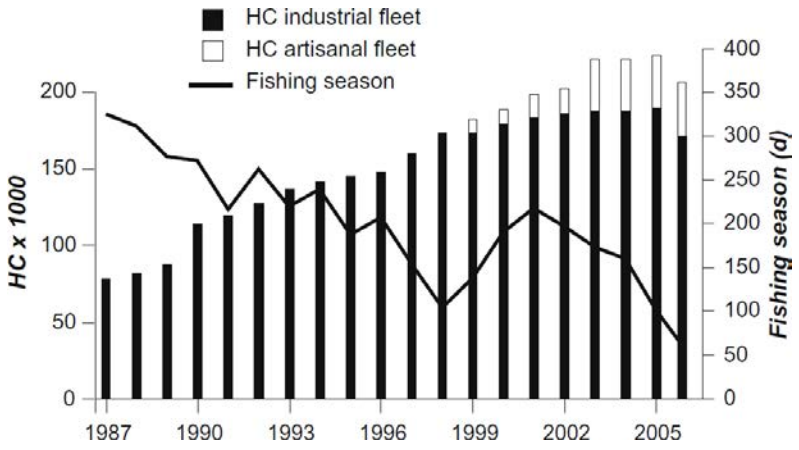
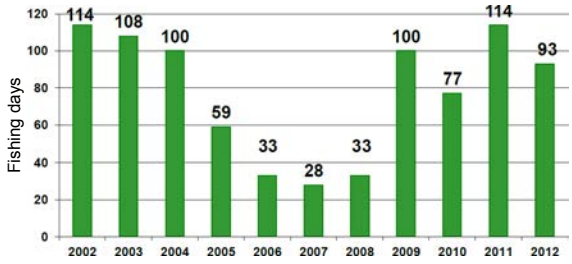


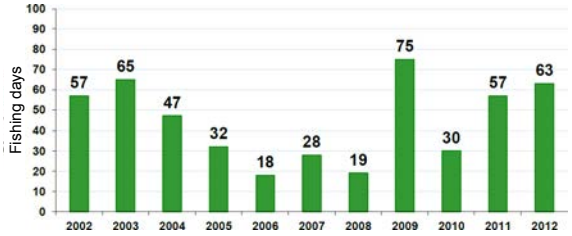
Figure 3.13: Holding capacity (HC) of the industrial fleet compared with the length of the fishing season from 1987 to 2005. Source: Freón *et al.* (2008)

3.5.2 Characteristics of the pelagic industrial fleet

The fishery is mono-specific, since $\sim$ 95% of the catches consist in anchovy (Freón *et al.*, 2008). During the last decade, the number of operating vessels reached a peak of $\sim$ 1200 vessels by day (Freón *et al.*, 2008, Fig. 3.12). The purse-seiner fleet is



(a) First fishing season of the year



(b) Second fishing season of the year

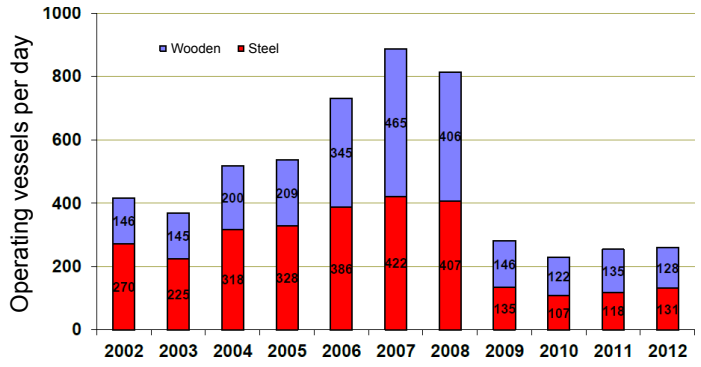
Figure 3.14: Length of the first (a) and second (b) fishing seasons of each year from 2002 to 2012 (north-center region). Source: IMARPE

classified into two segments: the steel and the wooden fleet (Fig. 3.16). The former comprises vessels mostly made of steel and larger than 120m^3 of fish-hold capacity. The wooden fleet is locally named the 'Viking' fleet, because of the shape of the hull; it comprises vessels ranging from 30 to 119m^3 of fish-hold capacity. In addition to the industrial fishery, there is a growing artisanal fishery (Estrella Arellano and Swartzman, 2010). Here we focus on the industrial fleet only. Due to the IVQ system, the number of operating vessels from both segments of the industrial fleet has been significantly reduced and may continue to do so (Fig. 3.15).

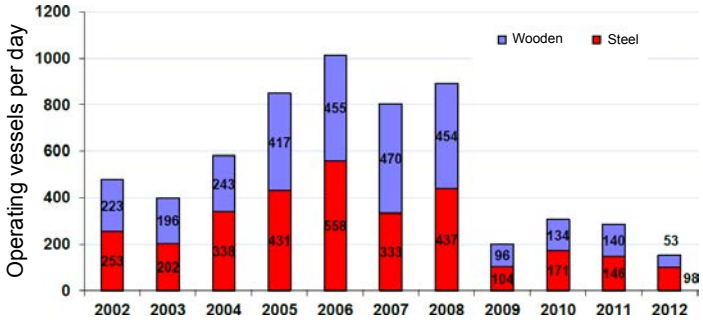
3.5.3 Fisheries economy

The fishery is the second sector of the Peruvian economy in terms of export revenues (it generated more than US$1 billion in export revenues in 2009; Tveteras *et al.*, 2011). It is an important source of employment and income for coastal areas (Aranda, 2009). It employs approximately 18000 fishermen that earn an average gross wage of US$4500 per fishing season, working on average 4 months per year (Aranda, 2009).

More than 99% of the anchovy landings of the fleet are processed into fishmeal and fish oil; although in the last years, anchovy processing for human consumption has started to take place (Freón *et al.*, 2008). Peruvian fishmeal and fish oil



(a) First fishing season of the year



(b) Second fishing season of the year

Figure 3.15: Number of operating vessels per day during the first (a) and second (b) fishing seasons of each year from 2002 to 2012 (north-center region). The red bars correspond to the steel fleet and the blue bars to the wooden fleet. Source: IMARPE

(a) Vessel from the steel fleet.

(b) Vessel from the wooden fleet

Figure 3.16: Examples of fishing vessels. Source: On-board observers program from IMARPE (left panel) and P. Fréon (right panel).

(a) 'Chata' (b) 'Chata' transporting anchovy

Figure 3.17: Operating 'Chata'. Source: http://tecnicanaval.blogspot.fr/2012/06/actividad-en-chata-transporte-de.html

production represent 25% and 30% of the world production of fishmeal and fish oil, respectively (IFFO, 2008). The industrial fleet supplies anchovy to 142 plants distributed at 24 locations along the Peruvian coast (Arias Schreiber *et al.*, 2011). Each company works with its own 'chatas' (Fig. 3.17). A 'chata' consits in a floating platform $\sim$ 20 m long, which supports a long pipeline that is used to transport anchovy directly from a fishing vessel offshore to the factory. The largest fishing companies simultaneously own fishing fleets and several fishmeal plants, located at different sites along the coast (Table 3.2). The highest processing power concentrates between latitudes 9°S and 11°S; though, in general, the distribution of processing capacity of each of those companies enables them to adapt to drastic changes in the distribution of anchovy fishing grounds (Arias Schreiber *et al.*, 2011).

In 2008, 40 fishing companies with authorization for fishmeal reduction were officially registered, of which seven were responsible for 68% of the total fishmeal production (Bendezú, 2008). These most productive companies owned 281 fishing vessels (equivalent to 54% of vessels holding capacity) and 72 processing factories (equivalent to 66% of the processing capacity; Bendezú, 2008). Stakeholders, i.e., fishing companies and boat owners connected to fishmeal production, are organized in fishing associations which are active in lobbying and information exchange with the Ministry of Production and the Working Group on Fisheries in the Congress (Aranda, 2009). The fishing associations do not have a direct involvement in decision-making but play an advisory role. The most representative fishing association, for the large-scale sector, is the National Society of Fisheries which integrates boat owners and vertically integrated companies. The wooden fleet is represented by the National Association of Boat-owners of the Law 26920.

Table 3.2: Latitudinal distribution of fishmeal processing capacity (tonnes/hour) from the largest fishing companies along the Peruvian coast in 2010.

		Fishing companies							
		Tecnología de Alimentos S.A.	Corporación Pesquera Inca S.A.	Pesquera Exalmar S.A.	CFG Investment S.A.C.	Pesquera Diamante S.A.	Pesquera Hayduk S.A.	Austral Group S.A.	Total (%)
Latitude	5°S	150	170			70	138		528 (9%)
	6°S								
	7°S	352	259	60	80		180		931 (15%)
	8°S								
	9°S	374	407	90	164	125	100	80	1340 (22%)
	10°S	80	142		76	80	60	113	551 (9%)
	11°S	391	80	134	80	202	120	100	1107 (18%)
	12°S								
	13°S	249		100	40	178	80	120	767 (13%)
	14°S								
	15°S	140							140 (2%)
	16°S	141			145	80			366 (6%)
	17°S	131	90				78	100	399 (7%)

Source: PRODUCE (taken from Arias Schreiber *et al.* (2011)).

3.5.4 Fishery management

Management policies for this fishery are differentiated according to two regions. In the north-center region of Peru (from the frontier with Ecuador at $\sim$ 3°S to 16°S), where most of the landings take place, industrial fishing is forbidden within the first 5 nm from the coast. Until 2008, total landings were limited by a total allowable catch (TAC). Once the TAC was reached, the Vice-Ministry of Fisheries ordered the closure of the fishing season, banning not only anchovy fishing but also fishmeal processing (Arias Schreiber *et al.*, 2011). From 2009, an IVQ system was introduced (Aranda, 2009). Under the TAC and now the IVQ system, effort is limited by fairly long fishing bans, decided on the basis of daily monitoring of the environment, the fish population and the fishery. In the southern region (from 16°S to the frontier with Chile), coastal restrictions vary from 1.5 to 3 nm from the coast. Here, management is also adaptive and based on daily monitoring of the ecosystem. Until early 2009, total landings in that area were not limited by a TAC and the fishery was open almost the whole year. Since July 2009, an IVQ system, independent from the one in the north-center region, was implemented.

In both regions, north-center and south, management is adaptive on short time scales (Chavez *et al.*, 2008). There is no mid-term management plan in operation for the anchovy fishery, as it is the case in many other commercial fisheries in Peru (e.g. demersal hake fishery since 2003, jumbo squid fishery, jack and chub mackerel since 2001). This reflects an intention of the government to avoid the use of legal instruments that could restrict or delay a rapid management decision process (Arias Schreiber *et al.*, 2011). Arias Schreiber *et al.* (2011) provide an example of rapid management of the anchovy fishery through the provisional closure of landing harbors when juveniles account for more than 10% of the catches: IMARPE delivers a report of high juvenile capture rates and the legal department of the Vice-Ministry of fisheries draws up the ministerial resolution for closing the port one day after. It is signed by the Minister of Production on that same day. The next day, the resolution is published in the daily official Peruvian newspaper 'El Peruano'. The port is closed right after the resolution has been published, on the same day. Two days elapsed from the delivery of IMARPE's report to the port closure.

3.5.5 Fishery monitoring

IMARPE is in charge of an intense ecosystem monitoring, within an ecosystem approach to fisheries (Fig. 3.18). It comprises satellite information on environmental conditions (e.g. sea surface temperature, Chlorophyll-a and sea level anomaly, among others) in daily and weekly resolutions. Fish population distribution and biomass are monitored through scientific acoustic surveys (two to three times a year). The fishing activity is supervised through landing statistics, Vessel Monitoring System (VMS) and on-board observers reports.

Remote Sensing and Geographic Information System

Monitoring of worldwide and local oceanographic conditions are done by real-time remote sensing of sea surface temperature, Chlorophyll-a, sea level anomaly, salinity, surface current speed and bathymetry, among others. Maps of oceanographic conditions are produced on daily, weekly, fortnightly and monthly bases.

Acoustic surveys

Since 1983, IMARPE has been conducting on average two acoustic surveys per year for monitoring fish population distribution and biomass. In each survey, from 1 to 3 scientific vessels can participate at the same time, splitting the total survey design among them. The surveys consist in parallel cross-shore transects of $\sim$ 100 nm long, with a $\sim$ 15 nm spacing. Simrad (kongsberg Maritime AS, Norway) scientific echosounders working at several frequencies are used to estimate biomasses (see Castillo *et al.*, 2009; Gutiérrez *et al.*, 2007; Simmonds *et al.*, 2009). An extensive midwater-trawl sampling completes the acoustic surveys for species identification. The nautical-area-backscattering coefficient (NASC, in $m^2.mn^{-2}$), an index of fish biomass (Simmonds and MacLennan, 2005), is recorded at each georeferenced elementary distance sampling unit (EDSU) of 1 nm. From NASC, indicators of biomass and spatial occupation are computed. In addition to the acoustic observations, fishing samples, water samples (Niskin bottles and CTD probes) and plankton (phytoplankton and zooplankton) samples are taken. Moreover, observations on seabirds and mammals are also registered. The biomass assessment is carried out during the survey and the final report is provided in the days following the end of the survey.

Eureka Operations

Since 1966, Eureka operations have been conducted (Villanueva, 1972). Eureka cruises are based on legal agreements by which the fishing companies allow IMARPE to use their fleet as scientific platforms in order to monitor biological conditions and anchovy stock in real time (Barange *et al.*, 2009). Scientific and technical staff from IMARPE are sent to the main fishing ports, from where anchovy vessels (between 25 and 50) cruise simultaneously to perform oceanographic sampling and carry out experimental fishing, following a survey track with a couple of parallel cross-shore transects of $\sim$ 100 nm long. For each Eureka operation, IMARPE elaborates a profile for the fishing vessels; then, among the vessels that fit the profile, a selection is made by lot. The cost of the cruises are completely covered by the fishing companies, thus notably reducing not only governmental expenses but also the delays involved in planning and implementing conventional large scale research cruises. The fishing companies involved are allowed to process the anchovy catches obtained from the experimental fishing to recover some of the associated financial costs.

Program of on-board observers

IMARPE runs a program of on-board observers for a sample of $\sim$ 25 vessels (the number varies each fishing season depending on the available funding). They are mostly concentrated in the anchovy fishery. Accepting observers on board is not a legal obligation for fishing companies. The observer program is therefore run on a voluntary basis by the vessels and relies on 'gentlemen's agreements' between IMARPE and the fishery.

This program started in 1974, after the fishery collapsed in 1972-1973. In the early 1980s, the observers program was discontinued until 1996. The program has been continuously evolving, in order to improve the quality and quantity of the information collected. From 1996 to 2007, observers mainly recorded in each fishing trip: the location and time of departure and arrival; the location and time of the different activities occurring during the trips (fishing, searching, cruising, drifting, helping other vessels, and receiving or giving fish to other vessel); total catch; observed cetaceans; and biological characteristics of fish samples from the catches. From 2007, they also recorded: the fishing company, type of fleet, holding capacity, cooling system, year of construction, vessel dimensions, net dimensions, engine, echo-sounders, sonars, discarded catch, offered catch, received catch, criteria for choosing the fishing zones, captain and years of experience, detection of top preda-

tors during the trip (species, date-time, location, number, etc.), among others.

Data from ports and coastal laboratories

IMARPE has seven coastal laboratories and sampling points in all landing ports of the industrial fleet. Daily samples from landings are taken for biological and size structure analyses. In addition, fish processing plants provide the coastal laboratories their daily records on supplying vessels and the quantities supplied.

Vessel monitoring system

A great source of information used by IMARPE for supervising the fishing activity is the Vessel Monitoring System (VMS). Industrial purse-seiners are legally obliged to use VMS tracking devices since 2000. In practice, while the steel fleet was almost entirely covered with VMS by 2000, the coverage of the wooden fleet has been much more gradual. Since 2000, vessel positions ($\pm$100 m of accuracy; $\sim$ 1 record per hour) for hundreds of thousands of fishing trips are thus available for scientific purposes. VMS data has been used for characterizing fishermen movement patterns (Bertrand *et al.*, 2007; Joo *et al.*, 2011), and their association with fish spatial patterns (Bertrand *et al.*, 2005, 2008c), other foragers (Bertrand *et al.*, 2012), and several ecosystem components (Bertrand *et al.*, 2008b).

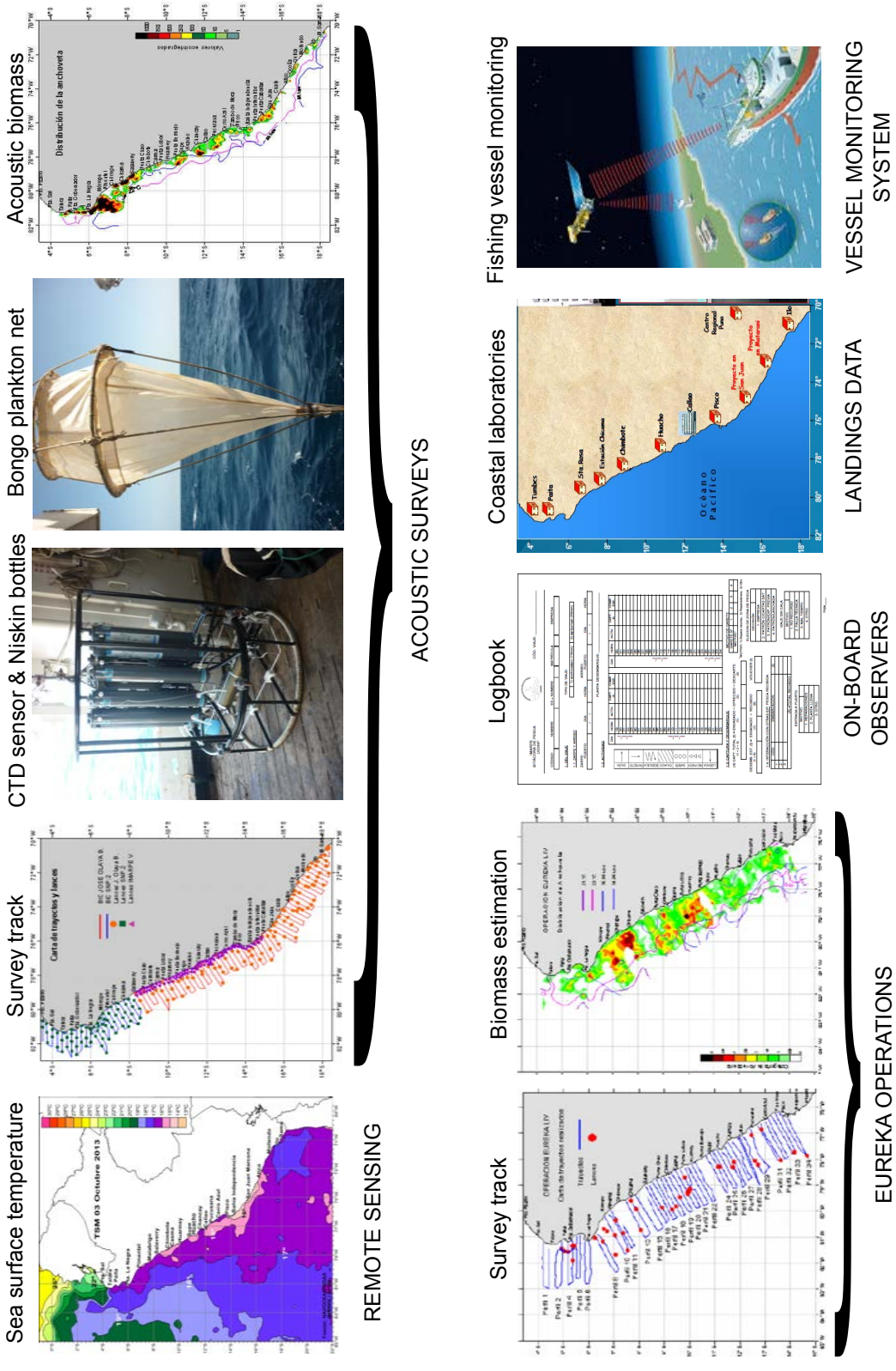


Figure 3.18: IMARPE's monitoring for an ecosystem approach to fisheries.

Chapter 4

Hidden Markov models: the best models for modeling behavioral modes?

"Remember that all models are wrong; the practical question is how wrong do they have to be to not be useful."

– George Box (*Empirical Model-Building and Response Surfaces*)

This chapter addresses an important methodological issue in movement analysis: how to identify the behavioral mode sequence associated with each movement path? Here, the movement paths correspond to fishing trips, and have been reconstructed from Vessel Monitoring System (VMS) data. The behavioral modes have been a priori defined: cruising, searching and fishing. We described in section 2.5 some of the main approaches for inferring behavioral modes in animal and human movement. Among the methods used in a priori defined behavioral modes, Markovian models, discriminative models (e.g. artificial neural networks and support vector machines), movelets and wavelets can be used for identifying multiple behavioral modes. However, large time series are needed for the two latter methods. As it is not the case, we compare here Markovian (Hidden Semi-Markov and Markov models) and discriminative models (random forests, artificial neural networks and support vector machines) using a groundtruthed dataset.

In section 3.5.5, we described the tools used by IMARPE (Peruvian Marine Research Institute or Instituto del Mar del Perú in Spanish) for monitoring the fishery. VMS positioning records are available for dozens of thousands of anchovy fishing trips per year. In turn, sequences of behavioral modes are observed for a sample

($\sim$ 1%) of those trips from on-board observers data. From those two datasets, a groundtruthed dataset, corresponding to one year (2008), was built ($>$ 200 fishing tracks with known behavioral modes). This dataset was used for comparing the performance of the inference models. Then, after the best model was chosen, it was used for (1) evaluating its performance with the groundtruthed dataset of all the other years, and (2) inferring the unknown behavioral modes in the remaining 99% of the fishing trips (2000-2009).

The chapter is presented under the form of a scientific article (Joo *et al.*, 2013). It presents the model performance and comparison in a general framework for foraging movement, though applied to the Peruvian anchovy fishery. The only changes introduced here in respect to the published version, are some additional results and less description of the data, since it was already described in section 3.5.5 and appendix A.

4.1 Introduction

Movement paths result from the interaction between the behavior of an organism and the spatial structuring patterns of its environment (Bergman *et al.*, 2000; Johnson *et al.*, 1992; Nams, 1996; Nathan *et al.*, 2008; With, 1994). Those paths result from the succession of distinct types of behavioral modes (e.g., traveling from one area to another, searching for cues or preys, pursuing and eating a prey), each one associated with the fulfillment of a particular goal. The knowledge of these modes provides rich information on the processes underlying movement, but they are not directly accessible through the sole observation of the sequence of positions recorded by GPS or other position-logging artifacts. The inference of the behavioral modes from movement paths remains a challenging issue in the emerging field of movement ecology (Patterson *et al.*, 2008).

Hidden Markov models (HMMs) have become increasingly popular to address this issue (for examples in classifying activities such as foraging, searching, encamping, cruising, migrating and bedding, see Bestley *et al.* (2010); Dean *et al.* (2012); Franke *et al.* (2004, 2006); Hart *et al.* (2010); Jonsen *et al.* (2007); Langrock *et al.* (2012); Patterson *et al.* (2009); Pedersen *et al.* (2011); Peel and Good (2011); Vermard *et al.* (2010); Walker and Bez (2010); in navigation strategies, see Guilford *et al.* (2004); Lau *et al.* (2006); Roberts *et al.* (2004); and in types of movement orientation, see Mann *et al.* (2013b)). HMMs rely on probabilistic inference of the

behavioral modes, stated as hidden states, from the *in situ* observed series. Those series are typically sequences of positions or associated features such as distances, speeds or turning angles along the movement paths (Langrock *et al.*, 2012). The key feature of HMMs is to account for the temporal dynamics of the behavioral modes, mostly based on state transitions between steps (two consecutive positions define a step). Such first-order HMMs comprise computationally efficient inference procedures (Gimpel and Rudoy, 2008; Rabiner, 1989). However, it may be unrealistic to consider that a forager takes a decision about changing its behavioral mode at each step, and regardless of any behavior dating from more than one step back. In this respect, hidden semi-Markov models (HSMMs), recently investigated in movement ecology (Langrock *et al.*, 2012), may be more appealing. While HMMs characterize behavior at the step scale, HSMMs characterize behavior at the segment scale; a segment is composed of consecutive steps associated with a same state. HSMMs do account for transitions between consecutive but distinct states and for durations of state segments corresponding to one behavioral mode.

For most living organisms studied in ecology, groundtruthed datasets – samples of tracks or positions for which behavioral modes are known – are hardly available. Therefore, inference issues are generally stated within a non-supervised framework. Furthermore, rigorous model validation (e.g. by cross-validation as in Hijmans (2012); Tan *et al.* (2006)) cannot be performed. Model validation mainly relies on some expert-driven evaluation of the ecological or behavioral plausibility of the behavioral modes inferred. Fishermen had long been the only foragers whose true behavioral modes were available. Actually, on-board observers can provide direct observations of the vessels' activities during fishing trips, allowing for model validation (Walker and Bez, 2010).

In such supervised settings, Markovian models may be applied. Nevertheless, alternative models may also be considered, particularly because Markovian models are limited for handling multiple observed variables. Choosing and fitting the most appropriate multivariate distribution may be delicate. A simplifying hypothesis is commonly adopted to solve this issue: observed variables conditioned on states are assumed mutually independent, so the multivariate distribution becomes the product of the univariate conditional distributions of each variable. By contrast, discriminative models, such as random forests (RFs), artificial neural networks (ANNs) and support vector machines (SVMs), provide robust solutions for non-linear discrimination in high-dimensional spaces. They have been shown to be highly efficient for

a wide range of applications (Byun and Lee, 2002; Cutler *et al.*, 2007; Hastie *et al.*, 2009; Mountrakis *et al.*, 2011; Zhang, 2000). Their availability in several softwares without the need of strong computational skills makes them attractive for applications to ecological datasets (Crisci *et al.*, 2012; Olden *et al.*, 2008). This includes a few studies dedicated to behavioral modes (Bertrand *et al.*, 2008c; Joo *et al.*, 2011; Morales *et al.*, 2005). This context of technological advances for data collection enables a wide range of supervised models. Hence, evaluating and comparing models accuracy for inferring behavioral modes becomes necessary.

Here, we examine the foraging movement of 50 Peruvian purse-seiners targeting anchovy in 2008. More than 200 of their fishing trips were documented by a Vessel Monitoring System (VMS) and their behavioral modes simultaneously recorded by on-board observers. This unique and large groundtruthed dataset allows performing, via cross-validation, a comprehensive evaluation and comparison of Markovian (HMMs and HSMMs) and discriminative models (random forests, artificial neural networks and support vector machines) for inferring the behavioral modes taken during fishing trips. We show that HSMMs provide the most accurate inference of the behavioral modes with 80% of global accuracy. We also show via simulation that this result could be greatly reinforced with position records of higher frequency. Finally, we use HSMMs for inferring the associated behavioral modes in all the fishing trips done by the Peruvian anchovy fishing fleet from 2000 to 2009 (more than 300000 fishing trips).

4.2 Materials and methods

VMS positions ($\pm$100m of accuracy; $\sim$ 1 record per hour) from the whole Peruvian industrial fishing fleet ($>$ 1000 vessels) are available since 2000. Although most records are given according to one-hour intervals, some irregularities (e.g. 0.17, 0.99, 1.2) seldom occur. Since there is no straightforward optimal interpolation method for these cases (Langrock *et al.*, 2012), we work with the records as they are. Therefore, the considered VMS data consist in tracks (i.e., series of positions) with non-regular steps. For each VMS track, several observed variables are computed at each step: speed (sp), heading (θ), changes of speed and turning angles between the previous and the current step (Δsp_{-1} and $\Delta\theta_{-1}$) and between the current and the next step (Δsp_{+1} and $\Delta\theta_{+1}$).

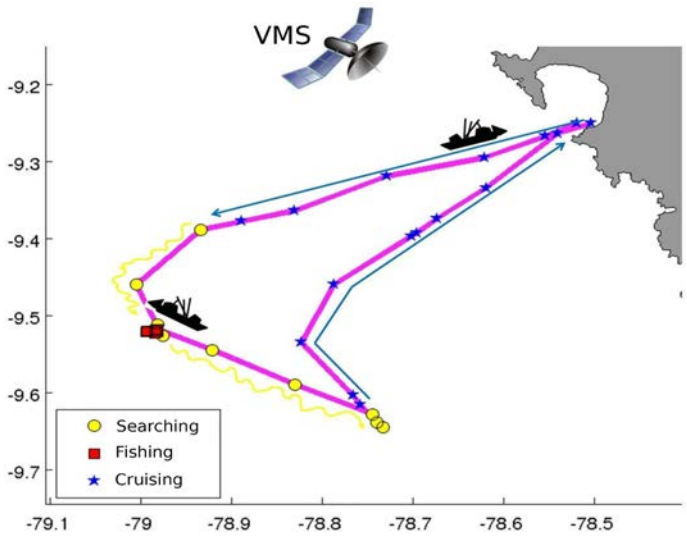


Figure 4.1: Fishing trip with VMS records and their corresponding behavioral modes.

In addition, IMARPE runs a program of on-board observers for a $\sim 1\%$ sample of the fishing trips. They record the location and time of the different behavioral modes occurring during the trips: fishing, searching, cruising (i.e. traveling following a predetermined course), drifting, helping other vessels, and receiving or giving fish to other vessel. For the remaining 99% of the fishing trips, behavioral modes are unknown.

Based on the criteria described in Bertrand *et al.* (2007, 2005) and Joo *et al.* (2011), a groundtruthed dataset gathering tracking data and their corresponding behavioral modes is built (see Appendix A for more details). Overall we consider three behavioral modes, fishing, searching and cruising. Fishing trips involving 'helping', 'receiving/giving' and 'drifting' modes are discarded, due to the low number of occurrences of these modes. Together they represent less than 6% of the groundtruthed dataset. We work with a dataset corresponding to 2008, consisting of 242 fishing trips ($\sim$ 36000 fishing trips were performed in total in 2008). Figure 4.1 shows an example of a trip with each VMS record associated with a behavioral mode.

For hidden state inference, two different approaches are investigated and evaluated. Markovian models, which take into account the sequential nature of data; and discriminative models, remarkably popular in the pattern recognition and machine learning domain (He *et al.*, 2008; Mjolsness and DeCoste, 2001; Nallapati, 2004).

Henceforth, we will denote by S_t the state variable at time t taking a discrete value s_t, which encodes a behavioral mode (fishing, searching or cruising). A state sequence starting at time 0 and ending at T is then denoted by $S_0^T \equiv S_0, S_1, \ldots, S_T$, taking discrete values s_0^T. Likewise, X_0^T represents the sequence of continuous observed variables taking values x_0^T. Under the two approaches, the goal is to infer s_0^T.

We perform a quantitative evaluation of the models performance using a classic cross-validation procedure. It proceeds as follows. The groundtruthed dataset is split into two sub-samples. The first partition is used for training the models, i.e., learning from the data and estimating the parameters. The second partition is used for validating the models, i.e., evaluating model performance. Training and validation partitions gather each 50% of the original sample of trips and are built by repeated random sub-sampling (20 repetitions). This parameter setting provides a trade-off between the performance evaluation and computational efficiency.

4.2.1 Markovian models

HMM: HMMs are the classic models for inferring hidden state sequences from observed variables (Bengio, 1999). A HMM combines the two following processes. An underlying first-order Markov process of the hidden state sequence, where the probability of currently being at state s_t only depends on the immediately preceding state s_{t-1}. And a state-dependent observation process, where the probability of $X_t = x_t$ only depends on the current state s_t and not on previous states or observations. Assuming homogeneity, a HMM can be fully characterized by (1) the initial probabilities $\pi_i = P(S_0 = i)$, (2) the transition probabilities $p_{ij} = P(S_t = j \mid S_{t-1} = i)$, and (3) the state-dependent observation probability density functions (pdfs) $b_j(x_t) = f(x_t \mid S_t = j)$, where $f(x_t \mid S_t = j)$ denotes the conditional pdf of X_t at x_t given $S_t = j$. When observations are multivariate, under mutual independence

$$b_j(x_t) = \prod_{e=1}^{E} b_{je}(x_{te})$$

where E is the number of observed variables included in the model. The likelihood of a HMM can be written as

$$P(S_0^T = s_0^T, X_0^T = x_0^T) = \pi_{s_0} b_{s_0}(x_0) \left[\prod_{t=1}^{T} p_{s_{t-1} s_t} b_{s_t}(x_t) \right]$$

In our case study, several observed variables are available (sp, θ, Δsp_{-1}, $\Delta\theta_{-1}$, Δsp_{+1} and $\Delta\theta_{+1}$). Over all possible combinations of observed variables, the subset (combination) of variables giving the highest state-inference accuracy is chosen –the computation of accuracy as well as other performance indicators are described in section 'Indicators of model performance'. For each observed variable, we test several probability distributions based on a supervised maximum likelihood (ML) fit. When ML estimation cannot be derived analytically, a numerical optimization is used. Goodness-of-fit (GOF) is tested using the robust Cramér-von Mises statistic (Schwarz, 1978). In cases where two or more distributions provided significant fits, the AIC criterion (Akaike, 1981) is used for selection among them. All fishing trips start in cruising mode, so initial probabilities are set to one for cruising and zero for the other states. Given the training partition, the ML estimation of the transition probabilities resorts to computing the relative frequencies of the transitions between successive states (Dietterich, 2002). Using all these elements, the inference of the sequence of hidden states s_0^T is done by global decoding via the Viterbi algorithm (Rabiner, 1989). Hidden Markov Model toolbox for Matlab™(Murphy, 1998) is used.

HSMM: A first-order Markov state process may not be, however, the most natural choice for the interpretation of movement patterns. It implicitly assumes that time spent at a given state is distributed according to a geometric distribution. This distribution is memoryless; it means that at a given time t, the waiting time for switching from one state to a distinct state is independent from the time already spent in the former state. However, in practice, a forager's behavior – and more specifically, a fisherman's behavior – is not memoryless. A semi-Markov process may therefore be more suitable. It explicitly models the state duration distribution and may consider any distribution function. HSMMs are thus generalizations of HMMs. They combine two processes: a state-dependent observation process as in HMMs, and an underlying semi-Markov state process. A semi-Markov process is determined by the duration distributions $d_j(u) = P(S_{t+u+1} \neq j, S_{t+2}^{t+u} = j \mid S_{t+1} = j, S_t \neq j)$ and transition probabilities between distinct states $p_{ij} = P(S_t = j \mid S_t \neq i, S_{t-1} = i)$. For the last visited state, a survival function of the duration is used: $D_j(u) = \sum_{v \leqslant u} d_j(v)$. The likelihood of a HSMM (Guédon, 2003) can be written as

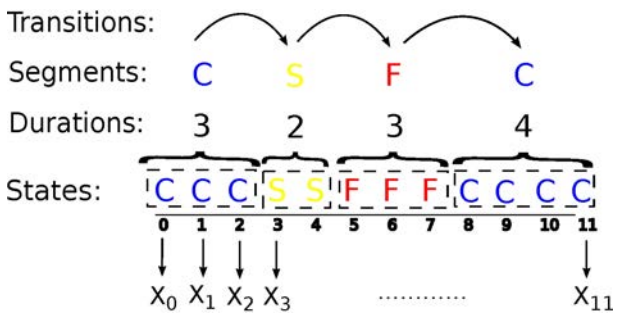


Figure 4.2: Schematic representation of a HSMM. At each step, an observed feature X is related to a state, which encodes a behavioral mode (C: cruising, F: fishing, S: searching). The state process is modelled at the segment scale and it is characterized by durations and transitions as shown above.

$$P(S_0^T = s_0^T, X_0^T = x_0^T) = \pi_{s_0} d_{s_0}(u_0) \left[\prod_{r=1}^{R-1} p_{s_{r-1}s_r} d_{s_r}(u_r) \right] p_{s_{R-1}s_R} D(u_R) \times I\left(\sum_{r=0}^{R} u_r = T \right) \left[\prod_{t=0}^{T} b_{s_t}(x_t) \right]$$

where $R+1$ is the number of visited states, u_r is the duration at state s_r, and $I()$ denotes the indicator function.

Therefore, compared to HMMs, HSMMs provide a model of the state process at a higher scale: the segment scale (Fig. 4.2; Dong and He (2007); Yu (2010)). This segment scale is potentially more relevant for interpreting and discriminating distinct behavioral modes in foraging movement (in our case, the activities made throughout a fishing trip).

The selection of observed variables, the fit of state-dependent observed variable distributions and the estimation of transition probabilities (between distinct states) follow the same criteria as for HMMs. Although state durations are inherently discrete, continuous distributions provide flexibility under certain irregularities on the frequencies of positioning of satellite records. They enable the incorporation of those data directly into the model. Extensive literature on the use of continuous distributions for modeling duration is available (e.g., Beyreuther and Wassermann (2011); Dong and He (2007); Levinson (1986); Yu (2010)). Here we examine seven continuous probability distributions for modeling the duration of each of the three behavioral modes. Their parameters are estimated by maximum likelihood using the training dataset. Then, GOF is tested using Cramér-von Mises statistic and AIC

criterion is used for selection among distributions not rejected by the test. Using all these elements, the inference of the sequence of hidden states s_0^T is done by global decoding via the forward-backward Viterbi algorithm (Guédon, 2007). A description of the main features of several extensions of the Viterbi algorithm for HSMMs is given in Appendix B (the one we use is described in section B.5).

4.2.2 Discriminative models

Discriminative models are alternative approaches for inferring behavioral states within recorded trajectories. In contrast to Markovian approaches, discriminative models do not rely on the explicit modeling of the joint likelihood of observation and state sequences. The inference of the behavioral mode sequence s_0^T is stated as a classification issue, i.e. the determination of the class (behavioral mode) attached to any position along the trajectory. Within a supervised framework, discriminative models learn a classification rule to predict a class from an observed vector x_t. Random forests (Breiman, 2001), support vector machines (Burges, 1998) and artificial neural networks (Warner and Misra, 1996) are among the state-of-the-art techniques in the machine learning domain (Hastie *et al.*, 2009). These models differ in the way classification rules are stated and learned. For SVMs, the goal is to maximize the margin around the hyperplane that separates classes. For ANNs, the objective is to minimize the classification error. And for RFs, discrimination is achieved by the simultaneous minimization of the within-group variances and maximization of the between-group variances. The relative performances between these methods are application-dependent and vary according to the structure of the observation space (Meyer *et al.*, 2003). A key feature of discriminative models is that they do not require any assumption on the nature of the observed variables, their distributions or covariances. To prevent over-fitting during the learning stage, a cross-validation procedure can be applied. Still, it requires sufficiently large and representative groundtruthed datasets.

As for HMMs and HSMMs, the subset of observed variables giving the highest inference accuracy is selected. The selected subsets may differ among the three discriminative models. Architecture and parametrization of each discriminative model is described below.

RFs: A random forest involves a set of N decision trees. A decision tree discrim-

inates patterns recursively in a tree-like structure. At each tree node, m variables are randomly selected among the subset of observed variables. Data are split following certain conditions on those m variables, so that within-group variance is minimized and between-group variance is maximized. For each observed vector x_t, a tree's output is its classification in a behavioral mode. Consequently, a random forest's output is the statistical mode of the classification outputs of N trees. We test $N = \{50, 100, 500, 1000\}$ and $m = \{1, 2, \ldots, Y\}$, where Y is the size of the subset of observed variables. The Matlab™implementation of the random forest library (Jaiantilal, 2009) is used.

SVMs: Support vector machines are based on linear discrimination. A Gaussian kernel is used here for mapping the originally observed vectors into a new space in which classes (i.e., behavioral modes) may be linearly separated. Tested values for the scale parameter of the Gaussian kernel are $\{10^{-4}, 10^{-3}, 0.1, 0.5\}$. SVMs also involve a regularization parameter C. Increasing the value of C increases the cost of misclassifying points and decreases generalization power of the model. We test $C = \{0.1, 1, 10, 100\}$. The Matlab™implementation of the Libsvm library (Chang and Lin, 2011) is used.

ANNs: Multilayer perceptrons (MLPs) are the most widely used architectures of ANNs. Neurons are organized in layers. The first layer is composed of the observed variables and the last layer is composed of the model classification output. Between those first and last layers, one or more hidden layers can exist. Here, we use a MLP with one hidden layer as in Joo *et al.* (2011). Considered options for the number of hidden neurons range from one to ten. The Matlab™neural network toolbox is used for the analysis.

For each discriminative model, we determine the optimal parameter setting according to the classification accuracy.

4.2.3 Indicators of model performance

Overall, we aim at accurately reconstructing the sequence of states associated with each fishing trip. We consider two scales of analysis. First, we evaluate the accuracy of the inference at the step scale, and define the accuracy indicator as the percentage of individual steps where the inferred states correspond to the real ones. Second, we

assess model performance at the segment scale (Fig. 4.2), which best characterizes behavioral modes. We use three indicators for each behavioral mode:

- The segment-level precision, defined as the percentage of inferred segments where the inferred behavioral mode corresponds to the true one.
- The segment-level recall, defined as the percentage of real segments where the true mode is correctly inferred.
- The F-measure or F1, which combines precision and recall performances (McSherry and Najork, 2008). It is defined as the harmonic mean of precision and recall, and reported here in terms of percentage similarly to precision and recall indicators.

Accuracy, precision, recall and F1 are standard performance evaluation measures in supervised contexts (Kohavi and Provost, 1998). Beyond these performance measures, we also investigate the extent to which the considered models deliver a relevant global characterization of behavioral patterns, particularly regarding the shape of the distributions of the behavioral mode durations. In this respect, we define a fourth indicator at the segment scale, called duration. This auxiliary indicator is computed as the mean squared difference between the empirical cumulative distribution functions of both real and inferred mode durations. Its values range from 0 to 1, where 0 refers to an error-free inference.

Formulas for the computation of all these indicators are shown in Table 4.1. Further details as well as an illustrative example on the computation of accuracy, precision, recall and F1 are described in Appendix C.

After the best model is chosen, we use that model and its corresponding set of observed variables for inferring the behavioral modes to all the available VMS tracks from 2000 to 2009. Indicators of performance are computed for groundtruthed subsets of tracks for each year.

4.3 Results

The selected distributions for the state-dependent observation process (for HMMs and HSMMs) and for the duration of the states (for HSMMs) are shown in Table 4.2.

Table 4.1: Indicators of model performance.

Scale	Indicator
Step	$\text{Accuracy} = \frac{\text{inferred} \cap \text{true}}{\#\text{states}} \times 100\%$
Segment	$\text{Recall} = \frac{\text{inferred} \cap \text{true}}{\#\text{true}} \times 100\%$
	$\text{Precision} = \frac{\text{inferred} \cap \text{true}}{\#\text{inferred}} \times 100\%$
	$\text{F1} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \times 100\%$
	$\text{Duration} = \frac{\sum_{i=1}^{n} (F_i - G_i)^2}{n}$

Notes: F and G represent empirical cumulative distributions for the real and inferred durations of a given behavioral mode, respectively.

Table 4.2: Distribution for each observed variable and duration conditioned on states.

Observed Variable	Searching	Fishing	Cruising
sp	generalized Pareto	generalized extreme value	Gaussian mixture
θ	uniform	wrapped Cauchy	Laplace-Gaussian mixture
$\Delta\theta_{-1}$	Kumaraswamy	uniform	loglogistic
$\Delta\theta_{+1}$	Beta	uniform	loglogistic
Δsp_{-1}	Laplace	Gaussian mixture	Student's t
Δsp_{+1}	Laplace	Gumbel	Student's t
Duration	generalized extreme value	lognormal	generalized extreme value

Notes: When Beta and Kumaraswamy distributions are used, data is transformed to scale from 0 to 1.

For evaluating and comparing the two Markovian and the three discriminative models, we selected, for each model, the subset of observed variables which led to the greatest inference performance in terms of accuracy rate. Performance indicators at step and segment scales are reported for each of these models (Table 4.3). All models infer states with an accuracy greater than 75%. By a small though significant difference ($p < 10^{-5}$ in paired-sample randomness tests; Siegel, 1956), the HSMM's accuracy is the highest.

Table 4.3: Performance of all models for their corresponding best subsets of observed variables.

Mode	Model	HSMM	HMM	SVM	RF	ANN
	Subset	sp, Δsp_{+1}	sp, Δsp_{+1}	$sp, \Delta sp_{-1}$, Δsp_{+1}	$sp, \Delta sp_{-1}$, $\Delta sp_{+1}, \theta, \Delta\theta_{+1}$	$sp, \Delta sp_{-1}$, $\Delta sp_{+1}, \Delta\theta_{+1}$
	Accuracy	**80.3**%	79.1%	79.0%	76.4%	79.2%
F	Recall	86.6%	84.7%	88.3%	85.5%	88.5%
	Precision	69.3%	69.4%	65.8%	64.5%	65.7%
	F1	**77.0**%	76.3%	75.4%	73.5%	75.4%
	Duration	4	14	17	29	22
S	Recall	67.5%	64.9%	56.3%	62.0%	59.5%
	Precision	56.9%	56.1%	57.6%	47.5%	54.3%
	F1	**66.7**%	60.2%	56.9%	53.7%	56.9%
	Duration	7	4	33	41	26
C	Recall	91.0%	87.4%	89.6%	76.9%	88.7%
	Precision	87.3%	86.4%	71.9%	72.0%	75.2%
	F1	**89.1**%	86.9%	79.8%	74.3%	82.1%
	Duration	1	2	7	25	7

Notes: In bold, the highest values of accuracy and F1. F: fishing; S: searching; C: cruising. Duration values are scaled by (10^{-4}).

Regarding behavioral modes, cruising seems to be the easiest mode to identify. All models show greater F1 scores for the cruising mode (between 74% and 89%). Likewise, the greatest recall and precision values correspond to cruising for all models. Relevant F1 scores are also reached for fishing mode inference (between 73% and 77%). By contrast, the identification of the searching mode appears difficult for all models (F1 between 54% and 67%). This behavioral mode involves relatively large confusion rates with both fishing and cruising modes (between 15% and 19% of the searching states are classified as fishing, and between 25% and 34% are classified as cruising, among all models).

For each behavioral mode, the HSMM outperforms all the other models (greatest F1 scores of 77%, 67% and 89% for fishing, searching and cruising, respectively). The second best model is the HMM. Differences between F1 scores of the HSMM and the HMM are significant for all behavioral modes ($p < 10^{-5}$ in all cases). Among the discriminative models, the ANN is the best model, followed closely by the SVM.

The analysis of the distribution of the inferred durations for each behavioral mode leads to similar conclusions. In Figure 4.3 it can be observed that all three discriminative models show higher empirical densities for low duration values than

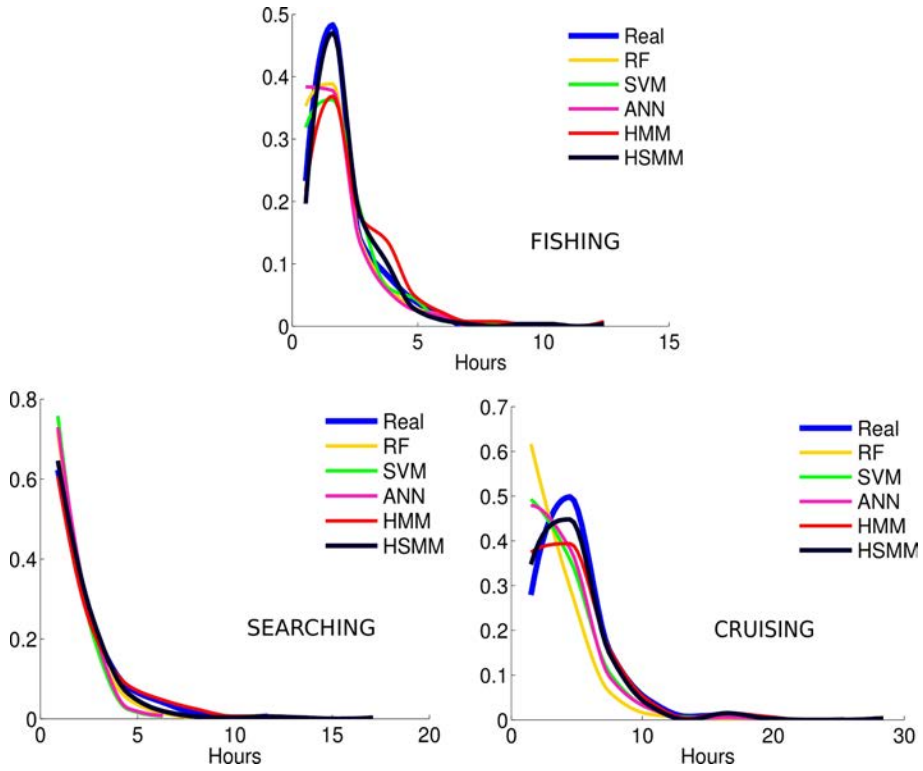


Figure 4.3: Distribution of the duration of each behavioral mode. For each model, an empirical distribution of the duration of each mode is estimated based on the duration of all inferred segments encoding the mode. RF: random forest. SVM: support vector machine. ANN: artificial neural network. HMM: hidden Markov model. HSMM: hidden semi-Markov model. Real: known behavioral modes.

the Markovian models and the groundtruth. Discriminative models, i.e. RF, SVM and ANN, which do not consider state transitions nor durations, tend to under-estimate the duration of modes due to over-segmentation. By contrast, the Markovian models, particularly HSMM, provide more accurate estimates of these durations. Whereas the distribution of the durations for fishing and cruising modes are clearly better represented with the HSMM (duration statistics of 4×10^{-4} and 1×10^{-4} for fishing and cruising, respectively; Table 4.3), the HMM gives slightly better results for the searching mode (duration of 1×10^{-4}).

The over-segmentation problem is illustrated for one trajectory sample when comparing the sequences of behavioral modes inferred by the HSMM and the RF with the true sequence of modes (Fig. 4.4). There is strong over-segmentation in the sequences inferred by the RF, leading to under-estimation of the duration of the segments. By contrast, the HSMM achieves relevant representation of the mode sequences through time (Fig. 4.4, low panel) and thus also through space (Fig. 4.4, right panel).

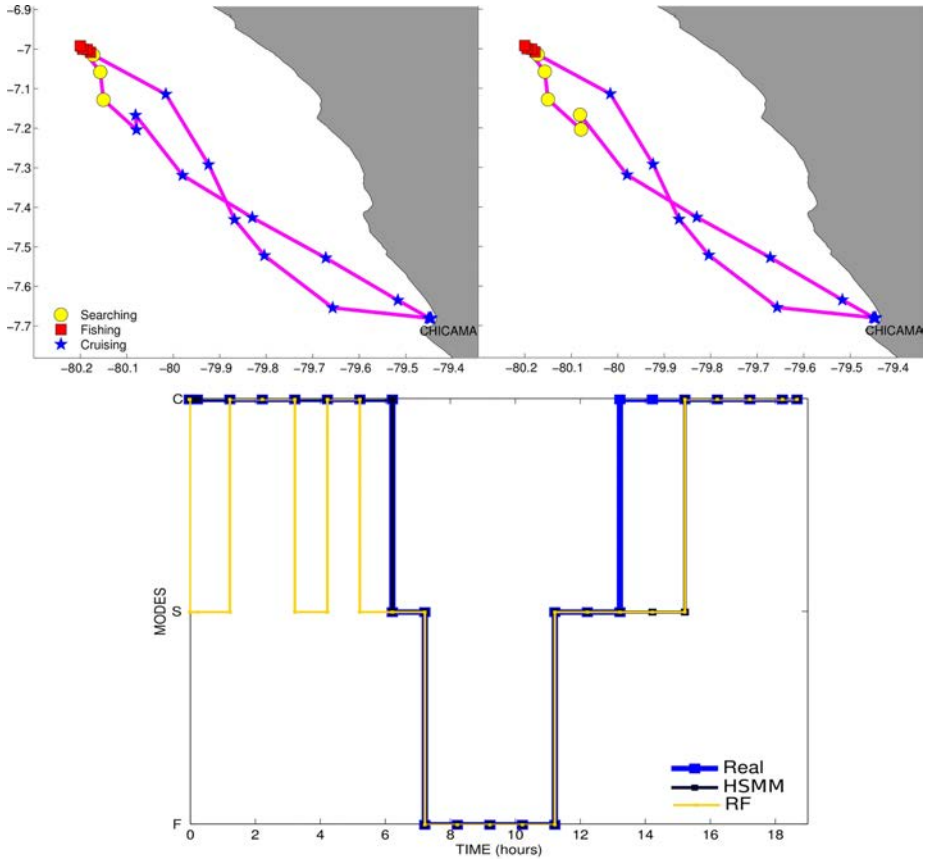


Figure 4.4: A fishing trajectory. Left upper panel: track with real behavioral modes. Right upper panel: track with inferred modes using the HSMM. Lower panel: temporal representation of the behavioral mode sequences, real and inferred, where 0 in the x-axis represents the beginning of the trip.

Regarding computational cost, we compare all five models in Table 4.3 for one replica where 121 tracks were randomly selected for training and the remaining 121 for validation. The HMM shows the lowest computational time (16.78 seconds), followed by the RF and the SVM models (22.09 and 23.07 seconds, respectively). Next it is the HSMM (64.04 seconds) and finally the most expensive one is the ANN (140.14 seconds). For the HMM and the HSMM, the computational time comprised the estimation of the probability density function parameters and Viterbi algorithm application. For the SVM, the RF and the ANN, it comprised the optimal parameter setting, as described in the Methods section. The high computational cost of the ANN could be greatly affected by the call to a graphical interface as automatically performed by the Neural Network toolbox of Matlab™. This computational analysis should only be regarded in relative terms. Optimized implementations of these models could be expected to provide important computational gains (by a factor of 10 or more).

Table 4.4: Performance of HSMMs for all years. Accuracy (in general), F1-scores for each behavioral mode and number of fishing trips with groundtruthed data

	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009
Accuracy	81.1%	78.4%	80.0%	74.5%	79.4%	81.1%	82.1%	77.7%	80.3%	73.6%
Fishing	77.0%	78.9%	79.6%	83.3%	79.6%	82.8%	84.6%	74.5%	77.0%	83.8%
Searching	58.9%	58.2%	62.3%	54.1%	63.2%	60.2%	58.3%	55.3%	66.7%	54.8%
Cruising	92.7%	88.1%	88.7%	81.0%	86.9%	87.9%	86.0%	87.4%	89.1%	83.6%
♯ trips	132	109	356	193	265	309	155	127	242	182

Since HSMMs outperform all the other models, they are then used for inferring the associated behavioral modes to all the available VMS tracks from 2000 to 2009. We use sp and Δsp_{+1} as observed variables since they show the best performance for the 2008 dataset. All subsets of observed variables were also tested when using HSMMs for 2000, and sp and Δsp_{+1} gave the best performance, too. Indicators of performance were computed over the groundtruthed subsets of tracks available for each year (Table 4.4). HSMM performance is similar among all years.

4.4 Discussion

With a representative groundtruthed dataset composed of 242 fishing trips, we perform a comprehensive cross-validation evaluation of different Markovian and discriminative models for inferring behavioral modes from trajectory data. Our results show that the HSMM is the best model and enlighten several critical issues.

4.4.1 State dynamics are key information

Markovian models have the strength of considering the sequential nature of the data: state transitions are explicitly modeled and the sequence of states is inferred as the most likely sequence given the performed trajectory. However, they present limitations for incorporating the information contained in the observed variables, especially in cases of non-Gaussian multivariate observation spaces. Practical applications of Markovian models often involve simplifications such as independence and/or Gaussianity assumptions for modeling the multivariate distribution of the observed features given the behavioral modes. In contrast, discriminative models state the inference of behavioral modes as a classification issue. They use powerful non-linear and multivariate classification rules. At the step scale, the HSMM surpassed the discriminative models by small differences (+1% of accuracy with respect to the ANN and the SVM, and +4% with respect to the RF; Table 4.3). At the segment scale, the surpassing performance of the HSMM was clearer (differences in F1 scores between +1.6% and +9.8% regarding both the ANN and the SVM, and between +3.5% and 14.8% regarding the RF; Table 4.3). This evidences that the information contained in the state sequence is key for accurately inferring the behavioral modes.

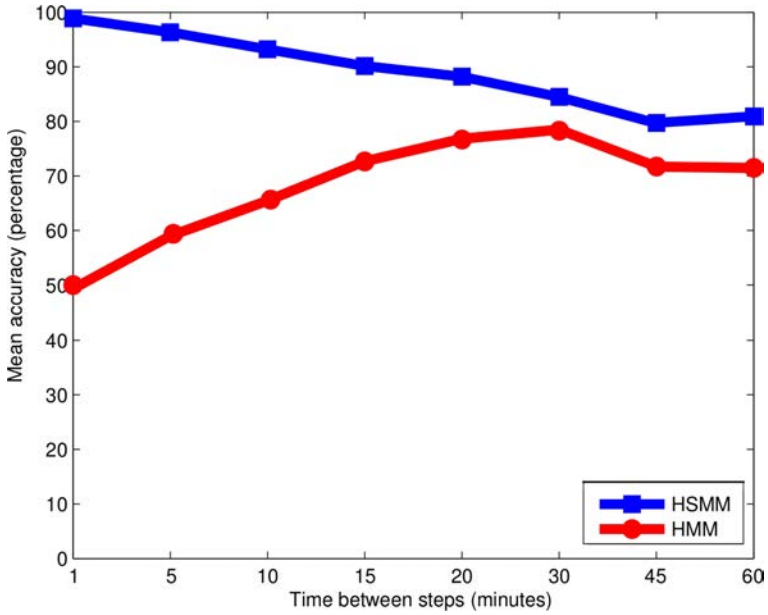


Figure 4.5: Mean accuracy for simulated sequences for different sampling rates using HSMM and HMM.

4.4.2 HSMMs are recommended for behavioral mode inference

To our knowledge, our study presents the first application of HSMMs to fishing tracks using groundtruthed data on behavioral modes. For this study case, with steps of $\sim$ 1 hour, the HSMM performed slightly better than the HMM. A simulation study on high-resolution data (one-second steps) is described in Appendix D. We applied HSMMs and HMMs to sub-sampled versions of these sequences. The performance of each model was assessed by the mean accuracy (MA), which is the average of the accuracy for each behavioral mode (Fig. 4.5). For one-minute steps, the HMM performed very poorly, whereas for 30-minute steps it was by far more relevant (50% vs 78% of MA). By contrast, MA rates for the HSMM remained above 80% for all time steps. The HSMM actually benefited from high-resolution sequences – when available – to significantly improve inference performance (100% of MA for one-minute steps). These additional results clearly illustrate that the relevance of the first-order Markov state process embedded in HMMs greatly depends on the time steps of the trajectory data. By contrast, we show that the relevance of the HSMM does not decrease with smaller time steps. Likewise, Whitehead and Jonsen (2013) showed that reducing time steps severely decreased the performance of first-order Markov processes for estimating animal spatial distributions from tracking data.

Alternatively, higher-order (n^{th}-order) hidden Markov models account for additional complexity in the dynamics of the state sequence. They comprise a n^{th} memory, i.e. the state value s_t depends on the state values taken at the n preceding states. They implicitly involve more general distributions on state segment durations than geometrical distributions. Therefore, they should outperform first-order HMMs for high-resolution sequences. However, in most practical problems the choice of the order of the hidden Markov model is not obvious and depends on both the time resolution of the data and the characteristic durations of the state segments. In addition, they are computationally expensive. HSMMs avoid the problem of choosing and fixing an order for the Markovian process. By considering transitions between distinct state segments and distributions on their durations, HSMMs model the scale of a homogeneous behavioral mode. By considering any distribution for modeling duration probability, HSMMs explicitly model the time an individual stays in a behavioral mode, rather than simply accepting the geometric decay of the duration distribution imposed by standard first-order HMMs (Langrock *et al.*, 2012). Moreover, by considering continuous distributions, HSMMs can directly incorporate tracking data involving some cases with different time steps.

Overall the great flexibility of HSMMs makes them particularly attractive for the analysis of fishing movement patterns.

4.4.3 Real behavioral modes and the relevance of model validation

The technological and methodological advances enable access to larger amounts of data and lead to continuously elaborating and applying new flexible modeling approaches for individual's movement (Jonsen *et al.*, 2013). While following this trend, model validation and evaluation are often disregarded. Sacks and Ylvisaker (2012) discuss this issue as a challenge in the future of statistics in general. It is also a challenge in movement modeling, particularly due to the conceptual and practical difficulties for obtaining groundtruthed data on individual's behavior.

Hence, when validating models with groundtruthed data, not only the models should be discussed but the data as well. In this work, we had access to a groundtruthed dataset, where behavioral modes were not chosen by us. Instead, they were previously defined by the fishermen themselves together with the on-

board observers. This meant that states were not chosen in a way that they would be *a priori* easily recognizable (based on path geometry). On the other hand, it gave a great opportunity for evaluating the models performance for inferring real and complex behavioral mode sequences.

We reported 80% of global accuracy and 77%, 67% and 89% of F1 for fishing, searching and cruising, respectively, using the fitted HSMM model. Whereas the general performance is satisfactory, the searching mode appears difficult to identify. It might be explained by the nature of this behavioral mode. Interviewed fishermen anticipated that geometrical patterns in their tracks related to searching might vary greatly depending on several factors, especially whether or not they presume the inspected zone to be of high prey density. According to the fishermen, observed patterns for fishing and cruising are more stable. The low F1 score for searching may also be due to the time resolution of the data. As for fishing, the activity lasts ~ 2 hours on average. However, 30-minute searching modes between two fishing modes were also reported by on-board observers. Such short state segments result in mixed signatures at the one-hour steps of the VMS data and can hardly be analyzed. Higher-resolution tracking data should clearly contribute to a better identification of such searching modes, and would decrease the confusion rates with fishing and cruising; thus improve the inference accuracy of all behavioral modes. Moreover, as shown by the simulation study, HSMMs would increase their inference power if data resolution increases.

4.4.4 Beyond validation: inference in supervised and semi-supervised contexts

In supervised contexts, inferring behavioral modes is not only useful for achieving model validation. Supervised contexts do not necessarily imply that groundtruthed data on the behavioral modes of the whole population of tracks are available. Known behavioral modes may only be available for a subset of the tracks. For some fisheries, there may not be enough resources for on-board observers to register activities from all fishing trips of the entire population of vessels with tracking devices. For the Peruvian anchovy fishing fleet, more than 30000 fishing trips are tracked by VMS per year, but behavioral modes of only ~ 300 of those trips are registered by on-board observers. Thus, models trained and validated over the groundtruthed samples could be used for inferring behavioral modes over the remaining tracks.

On the other hand, the non-supervised observed data could be used for updating the trained and validated models. In the machine learning domain, this is generally referred to as a semi-supervised setting. During the last years, numerous semi-supervised strategies have been proposed (see Chapelle *et al.* (2006) for an extensive classification and revision). Among them, Markovian models naturally extend from the supervised case to the semi-supervised one, using the EM algorithm (Dempster *et al.*, 1977). This appears as a particularly promising research direction for ecological studies, including the estimation of the resources (i.e. number of on-board observers and analyses) to be allocated for gathering an optimal groundtruthed dataset.

4.4.5 Modeling extensions for improving inference power

We have shown and discussed the advantages of Markovian models for taking into account the sequential nature of the data, while discriminative models typically achieve an independent inference of each state. Introducing past information on the observed variables may improve the inference performance of the discriminative models. We tested this possibility by introducing the immediate past values of the observed variables as new observed variables for the discriminative models. That meant adding four observed variables: speed at the previous step (sp_{-1}), heading at the previous step (θ_{-1}), change of speed between the two previous steps (Δsp_{-2}) and turning angle between the two previous steps ($\Delta\theta_{-2}$). The immediate past values of Δsp_{+1} and $\Delta\theta_{+1}$ are Δsp_{-1} and $\Delta\theta_{-1}$, respectively. As indicated in the Methods section, for each model, from all the possible combinations of observed variables, we retained the subset of variables giving the greatest accuracy rate. Only for ANNs, a different subset of variables (sp,sp_{-1},Δsp_{-1},Δsp_{+1},θ,θ_{-1},$\Delta\theta_{-1}$,$\Delta\theta_{+1}$) gave a higher accuracy. The new subset of observed variables involves the subset of variables from Table 4.3 plus four more observed variables. It improves inference of cruising modes (+0.5% in F1) and the general accuracy of the ANN model (+0.3%), although it decreases the performance over fishing and searching modes (-0.8% and -0.9% in F1, respectively).

Of course, more memory (past and future) in the observed variables could be added. But then, we would come across with the same memory-order dilemma than the one discussed for states in HMMs. Moreover, when we consider n^{th} order past (or future) of an observed variable, the first (or last) n records will have missing

values. This could be particularly annoying for classification using discriminative models. Another possibility would be to incorporate binary probabilities of the past states (i.e. presence or absence of a behavioral mode in the past states) for incremental training of discriminative models (Heskes and Wigerinck, 1996). Incremental training involves training the model one time-step at a time, updating the model at each step. Nonetheless, this may result in over-fitting and large generalization errors. Besides, the direct application of this strategy may lead to drift effect. It means that inference at time t may be biased as it is driven by the effect of the inference at time $t-1$. By contrast, Markovian models rely on a global inference, i.e. retrieving the state sequence that maximizes the posterior likelihood given the observed series. This global inference involves a forward-backward procedure which guarantees that the inference of any given state equally depends on past and future features along the trajectory.

Hence, combining the Markovian setting, which accounts for the sequential nature of the states, and the discriminative setting, which can achieve improved classification performance in high-dimensional non-Gaussian observation spaces, seems highly appealing. Such hybrid models have been investigated for different applications, especially speech recognition (e.g.Bourlard and Morgan (1994, 1998); Ganapathiraju *et al.* (2000); Stadermann and Rigoll (2004)). They are stated as Markovian models that rely on the definition of an observation likelihood from the output of the chosen discriminative model (e.g. the discrimination SVM function for hybrid SVM-Markov models; Ganapathiraju *et al.* (2000)). However, the parametrization of the observation likelihood and the training of the hybrid model remain complex issues, as we show in Appendix E.

Another attractive extension would be to model the observation process at the segment scale, i.e. at the same scale than that of the semi-Markov state process. That way, at each segment, one observation feature would be related to one state segment, which at the same time, would depend on the immediately preceding state segment. This modeling approach presents some potential advantages: it would imply modeling at the behavioral mode scale not only the state process but the observation process as well, and it could significantly improve the robustness to the presence of low-informative observation features.

The incorporation of informative priors could also play an important role in improving behavioral mode inference. For instance, fishermen may know *a priori*

that the probability of fishing success increases/decreases with daylight. Since this knowledge affects their behavior, hour-dependent state transition priors can be incorporated to the model. Likewise, priors on competition/association, as well as local climate conditions restricting mode transitions and durations could also be introduced in the model. A preliminary analysis on the inclusion of covariates into the model is shown in Appendix E.

4.4.6 Synthesis

We have shown a pioneer evaluation and comparison of Markovian and discriminative models for inferring behavioral modes within movement tracks in a supervised framework. The surpassing performance of Markovian models over the discriminative models highlights the importance of modeling state dynamics for accurately inferring the behavioral mode sequences. HMMs have been the most common approach in movement ecology. However, semi-Markov processes represent better the behavioral mode sequences than first-order Markov processes, since they explicitly model state duration and consider transitions at a segment scale. The HSMM performance on the groundtruthed dataset is slightly better than that of the HMM. As discussed above, this result responds to the nature of these particular behavioral modes as well as to the low resolution of the data. The $\sim$ 1 hour time steps are slightly below the characteristic durations of fishing and searching segments. Hence, regarding time steps, it is a favorable scenario for HMM. Through a simulation experiment, it was shown that increasing time resolution may decrease the accuracy obtained with HMMs and conversely increase the accuracy of HSMM inference. In foraging movement analysis, where (1) each type of behavior contained in a track is typically characterized by a distinct duration, (2) tracking data are increasingly available at high resolutions, and (3) irregularity in sampling rates is not uncommon, we highly recommend the use of HSMMs.

For the purposes of this thesis, we will use HSMMs for inferring behavioral modes on the VMS tracks of the whole fishing fleet from 2000 to 2009.

Chapter 5

Classifying fishing trip patterns of Peruvian anchovy purse-seiners

"In nature it is not convenient to consider every difference that is in things, and divide them into distinct classes: this will run us into particulars, and we shall be able to establish no general truth (..) The collection of several things into several classes gives the mind more general and larger views."

– John Locke (*An Essay Concerning Human Understanding*)

5.1 Introduction

Understanding fishermen spatial behavior and effort is essential for fisheries management (Garcia and Cochrane, 2005; Wilen, 2004). Fishermen spatial behavior responds to external factors (e.g., biotic and abiotic conditions, management rules, economic stimulus) and 'internal' factors (e.g., skippers' skill and personality, characteristics of the vessels). A fishing trip, as a behavioral unit, can reveal fishermen strategies that respond to both external and internal factors. However, at the scale of a fishing trip, the characterization of movement patterns has been only used for identifying different métiers (Russo *et al.*, 2011b).

The Peruvian anchovy fishery (*Engraulis ringens*), the world's largest monospecific fishery, provides a great opportunity for analyzing fishing trip strategies, which are not subject to a métier effect. In the Northern Humboldt Current system, fishermen trips are subject to several sources of variability: (1) the highly variable environmental conditions; (2) the sensitive and adaptive response of anchovy to the environmental conditions; (3) the race for fish until 2008 and the individual quota system since 2009; (4) the adaptive management, with distinct policies for the north-

center (*NC*; from 3°S to 16°S) and south regions (*S*; from 16°S to the frontier with Chile). Regarding the latter, there are several differences in the management policies between the *NC* and *S* regions: catch quotas are computed independently for each region; fishing is opened for many more days at the south; there is a ban for fishing within the first 5 nm from the coast at the *NC*, while at the *S* it varies between 1.5 nm and 3 nm from the coast. These differences in policies can cause contrasted fishing patterns among regions. Trip patterns should also respond to more intrinsic characteristics, such as skippers skills and personalities, and the intrinsic characteristics of the vessels.

Patterns of fishermen trajectories can be investigated from Vessel Monitoring System (VMS) data. From time series of vessel positions, it is possible to obtain fishing trip descriptors such as duration, distance traveled, maximum distance to the coast and time spent fishing, among others. In Peru, VMS is mandatory for the industrial fleet since 2000. In practice, while the steel fleet (vessels made of steel and with at least 120 m^3 of fish-hold capacity; Aranda, 2009) was almost entirely covered with VMS by 2000, the coverage of the wooden fleet (wooden-hulled and between 30 m^3 and 119 m^3 of fish-hold capacity) was much more gradual. In Joo *et al.* (in review), we analyzed how global fishermen trip patterns (i.e., average trip descriptors) are shaped by the environmental and biotic conditions occurring at a fishing season scale. Here, we study the fishing trip as a behavioral unit, and not an 'average trip' for a given fishing season. Therefore, we analyze fishing trips from different fishing seasons and years mixed all together, isolating them from their time period. By doing so, we 'weaken' the environment and fish stock effects in fishermen behavior, what allows a closer examination of the intrinsic variability (vessel and skipper characteristics).

In this work, we explore the patterns of fishermen behavior using a dataset of 352711 fishing trips monitored during the 2000-2009 decade, from which a set of descriptors were computed from VMS data. By means of a hierarchical cluster analysis, we study how the trips associate in different groups without establishing a priori the number of clusters. All the trips from the 10 years are analyzed together. If an ecological condition had a very strong effect in fishing trip patterns during a given year or group of years, then the trips from that year or those years will naturally associate with a specific cluster. We hypothesize that:
(1) Since anchovy is relatively accessible near the coast, most trips are short and stay near the coast, while a small group of high risk takers or stochasts (Allen and

McGlade, 1986) go farther away into unexplored zones; the patterns of this stochasts should stand out.
(2) Technology, manifested in the physical characteristics of the vessels and their equipment, conditions fishermen strategies (Stouten *et al.*, 2011), which should be expressed through fishermen patterns. The only information of vessels in our dataset is the type of fleet segment; we thus expect to find differentiated behaviors between the two segments.
(3) Differences in management conditions between north-center and south regions off Peru cause contrasted fishing trip patterns.
(4) The end of the race for fish in 2008, must have changed fishermen behavior and strategies at sea, so fishing trips from 2009 should differentiate from the others.
Therefore, we expect that fishing trip groups will disentangle risk takers from followers, *NC* from *S* trips, steel from wooden vessels, and trips under TAC (total allowable catch) system from the ones under IVQ system.

This chapter is organized as follows. In the next section, we present the fishing trip descriptors and the statistical methodology based on principal component and hierarchical clustering analyses (PCA and HCA, respectively). In section 3, we describe the multivariate fishing trip behavior via PCA and use the principal components for HCA. After choosing the number of clusters, we examine the pertinence of our hypotheses.

5.2 Materials and methods

5.2.1 Fisheries management, fleet and data

For this study, VMS positioning records from the Peruvian anchovy industrial fleet corresponding to the decade $2000 - 2009$ are used ($\pm$100m of accuracy; $\sim$ 1 record per hour; Fig. 5.1). Pre-processing of VMS data are performed based on the criteria and algorithms described in Bertrand *et al.* (2007, 2005); Joo *et al.* (2011) and are detailed in appendix A. The number of fishing trips per year kept in the dataset is shown in Table 5.1.

For each fishing trip, we first computed the following 11 global metrics: the duration (Dur), the total distance traveled (Dist), the maximum distance from the coast (Max.DC), the maximum and the minimum latitudes attained during the trip

(Lat.Max and Lat.Min, respectively), and the maximum and minimum longitudes attained during the trip (Lon.Max and Lon.Min, respectively). Longitudes and latitudes are all located in the Western and Southern hemispheres, so they take negative values.

Another source of information regarding fishing trip patterns is IMARPE's (Peruvian Marine Research Institute) program of on-board observers. For 1% of the fishing trips, on-board observers record the location and time of three main activities (i.e., behavioral modes) occurring during the trips: fishing, searching and cruising. In order to infer the behavioral modes for the remaining 99% of the VMS-tracked fishing trips, a supervised hidden semi-Markov model was trained and validated using the on-board observer dataset (Chapter 4). This model reached a mean accuracy of 80% in the determination of the correct behavioral modes from the VMS data. From the reconstructed sequences of behavioral modes, we compute four additional features: the time spent searching (Time.Searching), fishing (Time.Fishing) and cruising (Time.Cruising), the proportion of time spent on each of those modes with respect to the duration of the trip (Prop.Searching, Prop.Fishing and Prop.Cruising, respectively) and the time spent from the beginning of the trip until the first fishing set (Bef.Fishing). Hence, we had a total a total of 14 metrics describing fishing trips.

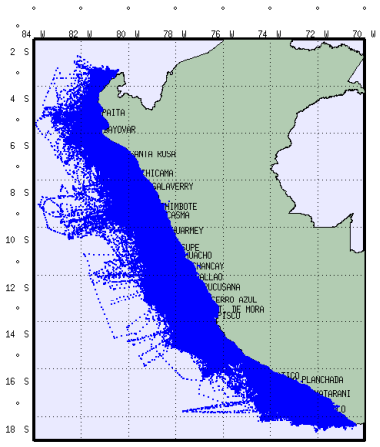


Figure 5.1: VMS positioning records from 2000 to 2009.

5.2.2 Statistical analyses

For the considered multivariate analyses, all variables are standardized. Then, we carry out a PCA (Pearson, 1901) before the clustering. Using a PCA before a clus-

Table 5.1: Fishing trips per year.

	Number of trips	Percentage of trips (from the total of trips in the dataset)	Percentage of wooden-vessel trips (from the trips occurred in each year)
2000	28807	8.2	0.5
2001	26176	7.4	1.1
2002	44429	12.6	0.7
2003	30764	8.7	1.6
2004	43648	12.4	8.9
2005	41518	11.8	27.7
2006	31018	8.8	37.6
2007	30606	8.7	42.0
2008	36143	10.2	42.5
2009	39602	11.2	53.2
Total	352711	100.0	22.0

ter analysis has several attractive advantages (Deporte *et al.*, 2012; Husson *et al.*, 2010): (1) removing the noise in the data by eliminating the last dimensions; (2) retaining only the components that we know how to interpret; (3) learning more about the interaction between variables; and (4) reducing the dimensionality of the initial data matrix which makes the cluster analysis less computationally expensive (although in our case study the dimensionality is not too high to be computationally expensive).

Clustering is carried out over the first principal components (Husson *et al.*, 2010). Cluster analysis is a method for grouping sets of features into groups or clusters in such a way that the features in the same cluster are more similar to each other than to those in other clusters. It is an unsupervised classification technique, since the clusters are not defined a priori (Hastie *et al.*, 2009). There are two main types of clustering: hierarchical and flat or partitioning. In partitioning clustering, clusters are independent of each other; i.e. there is no particular structure or organization within them or between them. Moreover, the number of clusters has to be given as an input to the analysis. Here we use hierarchical clustering. This method organizes partitions in a dendrogram (i.e., a tree structure) and partitions can be seen at different levels of granularities (i.e. refine/coarsen clusters) using different numbers of clusters, which provides a better understanding of the data. For merging clusters and constructing the dendrogram, we use Ward's minimum variance method (John-

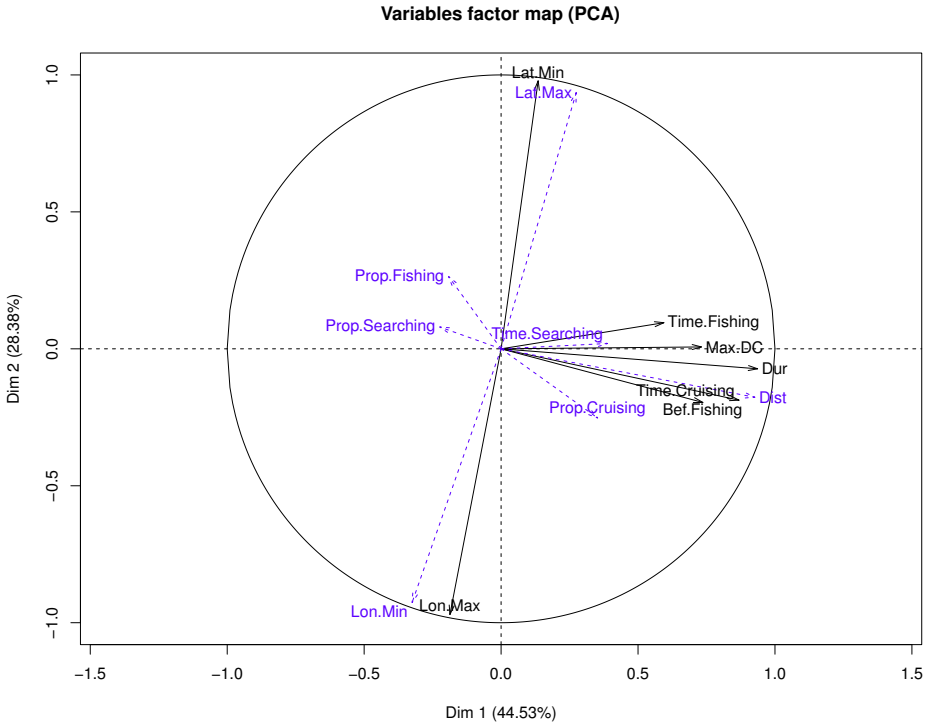
son and Wichern, 2007) and the Euclidean distance function.

Statistical analyses were performed with R software (R Core Team, 2013). FactoMineR package (Husson *et al.*, 2013) was used for PCA and Rclusterpp (Linderman, 2012) for hierarchical clustering.

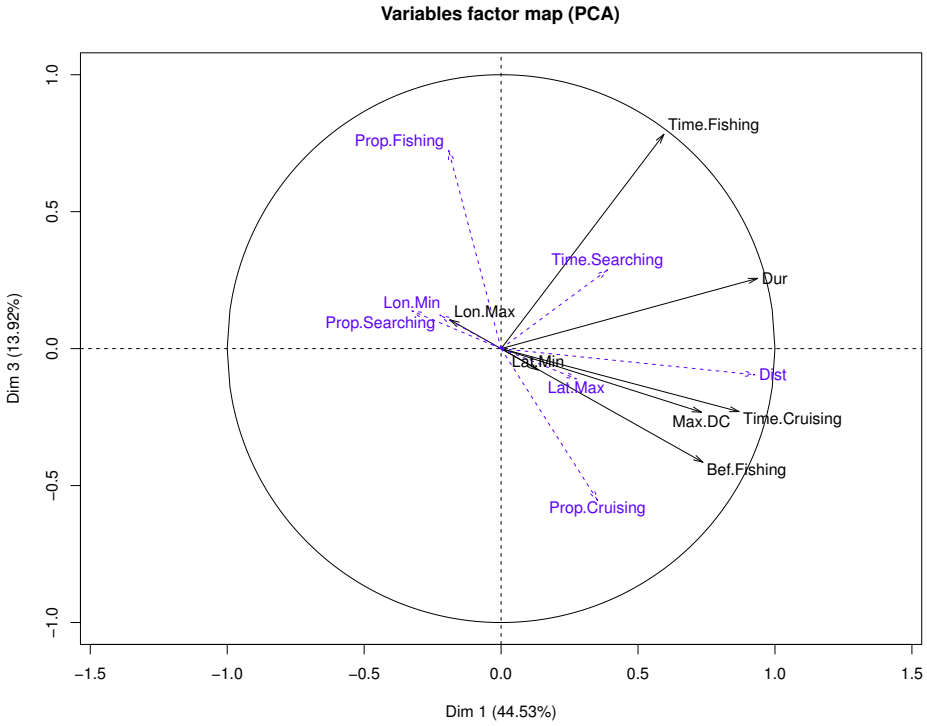
5.3 Results

For the PCA, not all the variables described above are used as active variables. Dist, Lat.Max and Lon.Min are discarded (i.e., only used as supplementary variables) since they are highly correlated to other active variables (Dur and Time.Cruising, Lat.Min, and Lon.Max, respectively). We prefer to use absolute time in each activity rather than proportions because absolute times are more correlated to the principal components; so Prop.Searching, Prop.Fishing and Prop.Cruising are also discarded. Moreover, we also discard Time.Searching since the sum of Time.Cruising, Time.Fishing and Time. Searching for each trip is equal to Dur (no variable should be expressed as a linear combination of others). Overall, we use Dur, Lat.Min, Lon.Max, Max.DC, Time.Fishing, Time.Cruising and Bef.Fishing. The other variables are still used as supplementary variables and projected into the PCA space (Fig. 5.2). The first component, which accounts for 45% of the variance, is strongly correlated with Dur, Dist, Time.Cruising, Bef.Fishing and Max.DC (Table 5.2). It refers to long and distant trips. The second component, accounting for 28% of the variance, relates to geographical location. High values on this axis correspond to northern fishing trips (larger latitude values are closer to zero or the Equator; and smaller longitude values are farther from Greenwich). The third component, which accounts for 14% of the variance, can be interpreted as a fishing effort proxy. High values in this axis are associated with much time spent fishing (in absolute time and in proportion of the duration of the trip). Overall, the three components explain 87% of the variance.

Cluster analysis is then performed over the scores corresponding to the three components. Based on the computed dendrogram (Fig. 5.3a) and the explained variance (i.e., the percentage of total variance represented by the between-cluster variance, Husson *et al.*, 2010, Fig. 5.3b) the 4-cluster structure is kept; it represents 61% of the explained variance. Boxplots of each descriptor are computed for each cluster (Fig. 5.4). The first cluster is composed of 15% of the fishing trips. It is



(a) Components 1 and 2



(b) Components 1 and 3

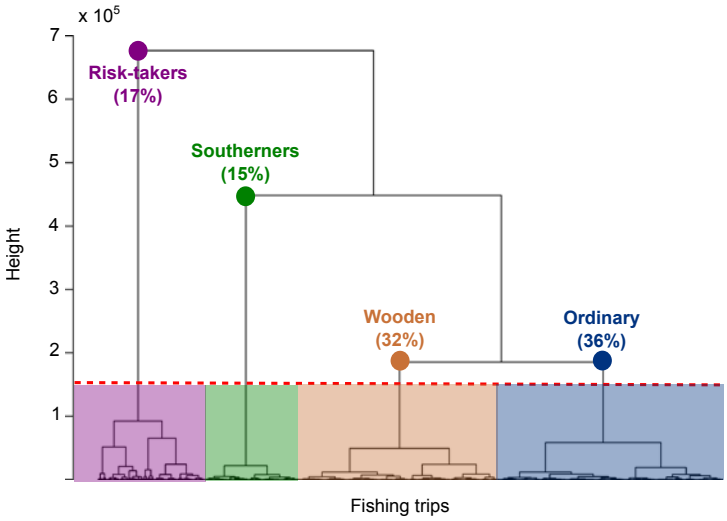
Figure 5.2: Projected variables in the principal component space. Active variables are in black and supplementary variables in blue. Upper panel: First and second components. Lower panel: First and third component.

Table 5.2: Significant correlations between the variables and the principal components ($p < 0.05$). Supplementary variables are in italic. Correlations above 0.70 are in bold.

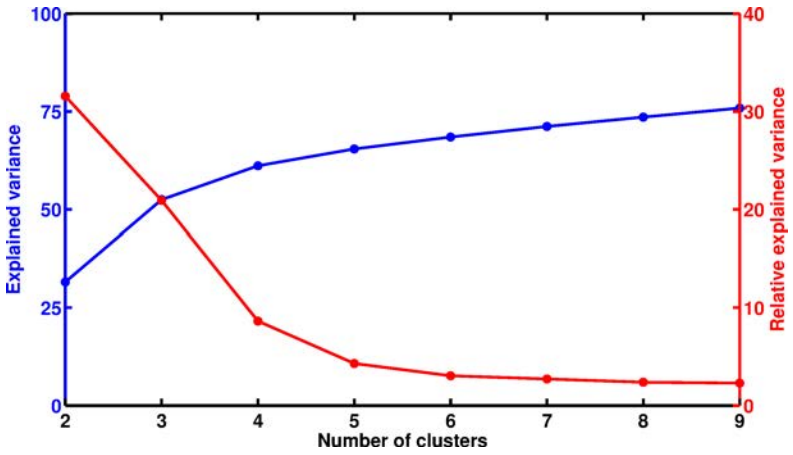
Variables	PC1 (45%)	PC2 (28%)	PC3 (14%)
Dur	**0.94**	−0.07	0.26
Dist	**0.93**	−0.18	−0.10
Time.Cruising	**0.87**	−0.19	−0.23
Bef.Fishing	**0.74**	−0.20	−0.41
Max.DC	**0.73**	0.01	−0.23
Time.Fishing	0.59	0.10	**0.78**
Time.Searching	0.39	0.02	0.29
Prop.Cruising	0.35	−0.25	−0.55
Lat.Max	0.27	**0.94**	−0.11
Lat.Min	0.14	**0.98**	−0.08
Lon.Max	−0.19	**-0.97**	0.11
Prop.Fishing	−0.19	0.26	**0.72**
Prop.Searching	−0.22	0.08	0.12
Lon.Min	−0.33	**-0.93**	0.14

mostly associated to fishing trips from the Southern region (almost 100% of southern trips; see Table 5.3) which remain near to the coast. The second cluster is composed of 17% of the fishing trips. It contains the longest and most distant trips by far. The highest fishing, searching and cruising absolute times are associated to this cluster. Trips of this second cluster were mostly concentrated at the central latitudes (between 8°S and 12°S), though they also operated in all the other latitudinal zones off Peru. The third cluster is composed of 32% of the fishing trips, and is mostly related to short and inshore trips. Those trips have the highest proportions of time spent fishing and searching and the lowest in cruising. They mostly operated between latitudes 7°S and 9°S. 75% of the fishing trips corresponding to wooden vessels are contained in this cluster (Table 5.4). The fourth cluster is composed of 36% of the fishing trips. Trips associated with this cluster have higher distance traveled, durations and maximum distance from the coast than the first and third clusters. Most of those fishing trips operated between latitudes 9°S and 13°S.

Each cluster contains fishing trips from all years (Table 5.3). Regarding how each year's trips are distributed in the clusters (Table 5.4), it should be noticed that cluster 4 contains higher proportions of fishing trips from 2000 to 2005 than the



(a) Dendrogram



(b) Variance plot

Figure 5.3: Hierarchical cluster analysis results. Upper panel: Dendrogram. Dotted line represents the pruning level. For each cluster, an associated label and percentage of fishing trips are shown. The height corresponds to the increase within-cluster computed in Ward's method (Husson *et al.*, 2010). Lower panel: variance plot. For each number of clusters (x-axis), the left y-axis represents the percentage of explained variance, and the right y-axis represents the increase in the percentage of explained variance when passing in relation to the preceding number of clusters.

other clusters (between 38% and 45% for each year). Likewise, cluster 3 contains the highest proportions of fishing trips from 2006 to 2009 (between 40% and 49% for each year). Those $2006 - 2009$ fishing trips constitute 56% of the trips in the cluster.

Table 5.3: Characteristics of clusters. Percentage of fishing trips corresponding to each year, fleet and region, for each cluster.

		Clusters			
		1	2	3	4
Years	2000	5.1	10.3	6.6	9.9
	2001	4.9	9.5	7.0	7.9
	2002	17.1	13.3	6.2	16.0
	2003	5.3	18.5	4.4	9.4
	2004	13.1	15.0	8.0	14.7
	2005	14.4	7.5	11.6	12.8
	2006	11.1	6.3	11.2	6.8
	2007	10.4	6.4	12.0	6.0
	2008	11.7	5.6	15.8	6.8
	2009	6.7	7.5	17.2	9.6
	Total	100.0	100.0	100.0	100.0
Fleet	Steel	90.4	93.6	48.3	91.8
	Wooden	9.6	6.4	51.7	8.1
	Total	100.0	100.0	100.0	100.0
Region	North-center	0.3	90.9	100.0	100.0
	South	99.7	9.1	0.0	0.0
	Total	100.0	100.0	100.0	100.0

Regarding the hypotheses made at the beginning of this work, we expected to disentangle risk takers from followers, *NC* from *S* trips, steel and wooden vessels, and trips under TAC system from the ones under IVQ system. Through PCA and HCA we identified 4 clusters. One cluster contained the longest trips and was associated with risk takers. Another cluster corresponded to trips in the Southern region. A third cluster was mostly associated with wooden vessels; while the last and larger cluster, which was label 'ordinary' (Fig. 5.3a), did not present any extraordinary characteristic: their trips were made by steel vessels in the *NC* region (mostly at the center), without outstanding duration, distances for steel vessels (they made longer trips than wooden vessels). No cluster grouped 2009 fishing trips, so no evidence in change of behavior as a response to the IVQ system was found.

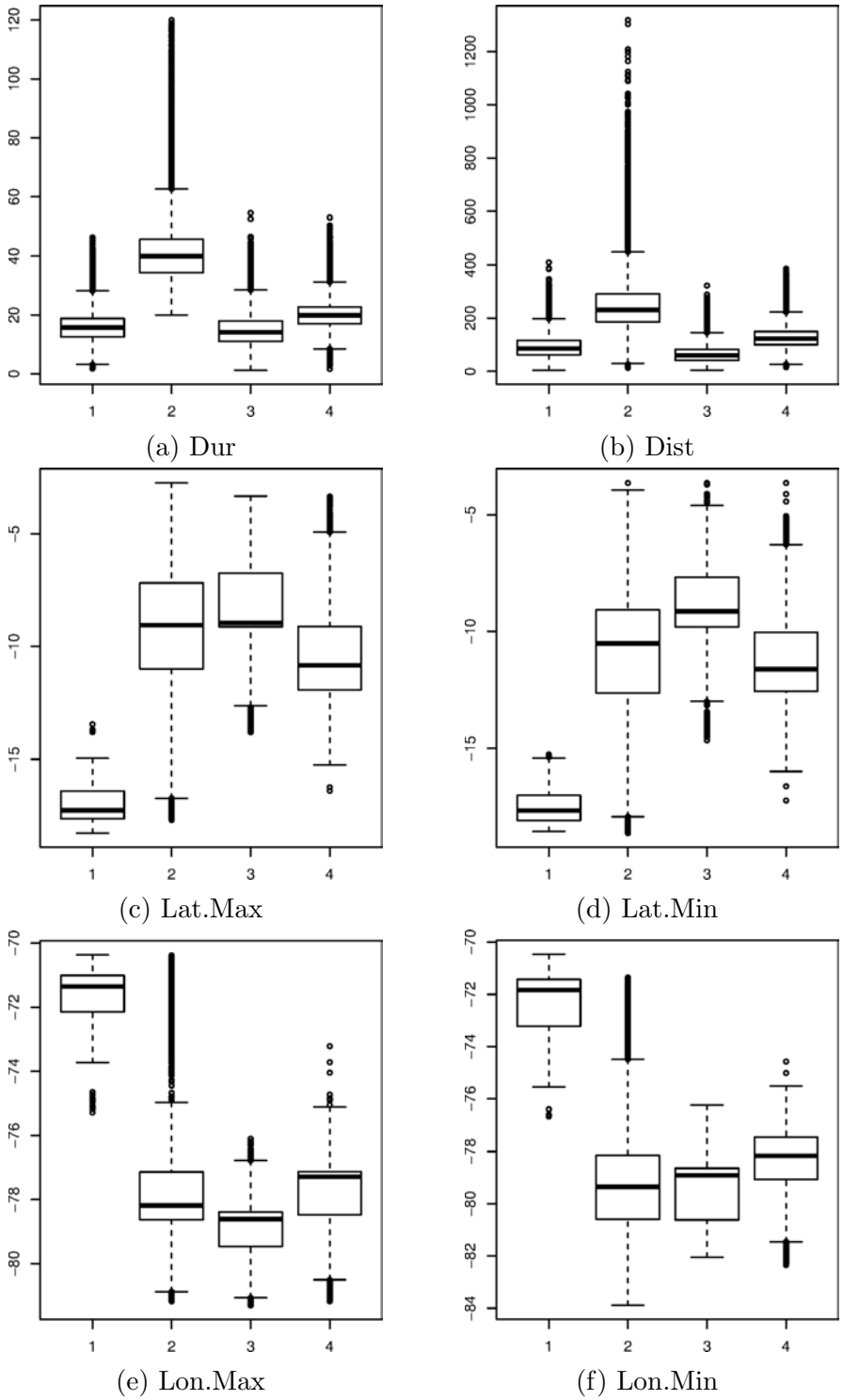
Table 5.4: Characteristics of each year, fleet and region. Percentage of fishing trips associated to each cluster. For each line, the highest percentage of trips associated to a cluster is in bold.

		Clusters				
		1	2	3	4	Total
Years	2000	9.6	21.6	25.7	**43.1**	100.0
	2001	10.1	21.8	30.0	**38.1**	100.0
	2002	20.9	18.1	15.7	**45.3**	100.0
	2003	9.3	36.4	16.0	**38.3**	100.0
	2004	16.3	20.8	20.8	**42.2**	100.0
	2005	18.8	10.9	31.6	**38.6**	100.0
	2006	19.4	12.3	**40.6**	27.7	100.0
	2007	18.4	12.7	**44.4**	24.5	100.0
	2008	17.5	9.4	**49.3**	23.7	100.0
	2009	9.2	11.4	**49.1**	30.3	100.0
Fleet	Steel	17.8	20.5	19.8	**41.9**	100.0
	Wooden	6.7	5.0	**75.2**	13.2	100.0
Region	North-center	0.1	18.7	38.5	**42.8**	100.0
	South	**90.8**	9.2	0.0	0.0	100.0

5.4 Discussion

In this work, we used a large dataset composed of 352711 fishing tracks, for studying how fishing trips grouped in different types of behavior, and identifying the main factors conditioning those behaviors. We used HCA over 3 principal components, which were in turn computed on several movement descriptors of the trips. The first component was associated with long distant trips, the second one was related to geographical location and a third one was associated with time spent fishing. Four clusters were identified based on those principal components and were labeled 'risk takers', 'Southerners', 'wooden' and 'ordinary' (Fig. 5.3a). Those clusters were consistent with 3 out of 4 of the hypotheses proposed at the beginning of the chapter. We will briefly discuss the implications of each of those clusters.

The first cluster to be dissociated from the population of tracks is the one of the risk takers. The risk-taker cluster is associated with long distant trips (Fig. 5.5b). In the Peruvian anchovy fishery, fishing trips typically last about 24 hours, mostly because vessels can only land once every 24 hours. In addition, since anchovy prefers cold coastal waters (Swartzman *et al.*, 2008), in general, fishing vessels do not need



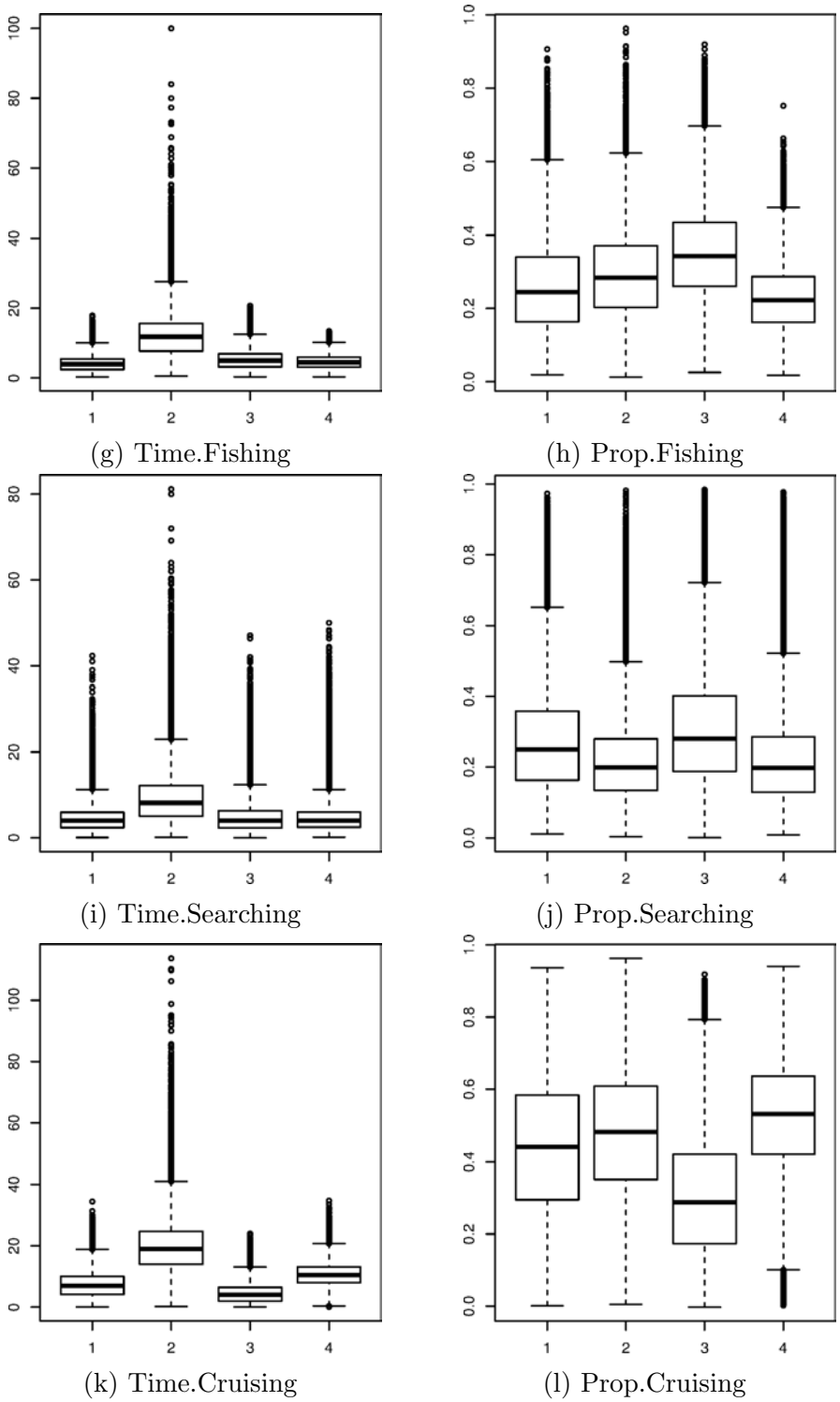
(a) Dur (b) Dist

(c) Lat.Max (d) Lat.Min

(e) Lon.Max (f) Lon.Min

Figure 5.4: Boxplots for each fishing trip descriptor and each cluster.

to go too far from the coast to fish. Long distant offshore trips may be made by risk takers. Allen and McGlade (1986) define them as skippers whose behavior is poorly driven by information – of where the large patches of fish are or of where the vessels are concentrated. They search 'randomly' or, more precisely, according



(g) Time.Fishing (h) Prop.Fishing

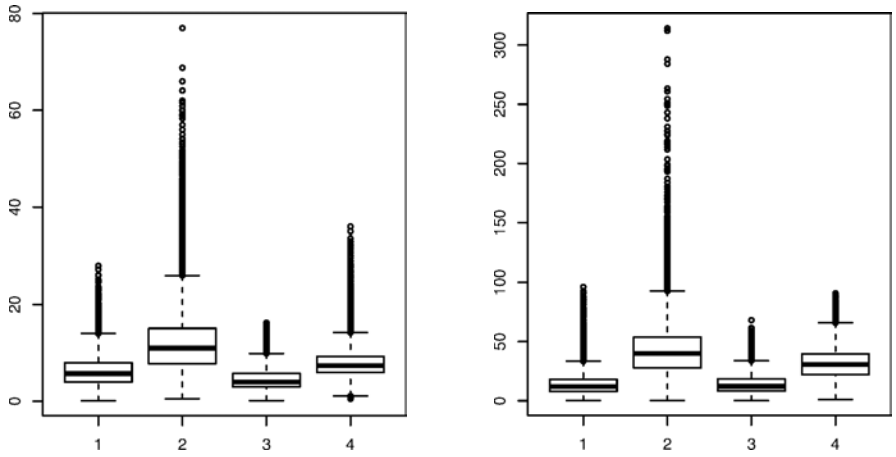
(i) Time.Searching (j) Prop.Searching

(k) Time.Cruising (l) Prop.Cruising

Figure 5.4: Boxplots for each fishing trip descriptor and each cluster.

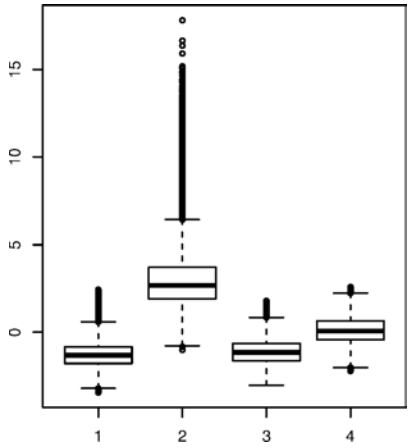
to some personal scheme of knowledge. Because they take more risks, they explore less visited zones and might discover new fish patches.

Their behavioral patterns strongly differ from the ones of the 'cartesian' skip-

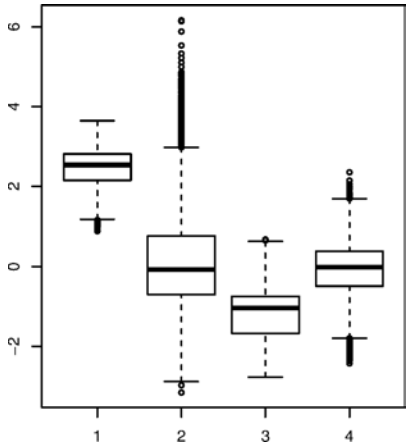


(m) Bef.Fishing

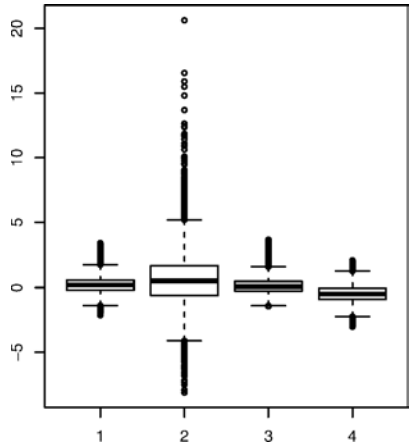
(n) Max.DC



(o) Principal component 1



(p) Principal component 2



(q) Principal component 3

Figure 5.4: Boxplots for each fishing trip descriptor and each cluster.

pers. The latter, unwilling to take risks, go to the zones promising the best known return, based on information on fish patches or satisfactory catches by other vessels. For that reason, 'cartesians' are also called followers. Risk takers are mostly from the steel fleet, since those vessels are better equipped for longer trips. Still, a small

proportion of 'risky' trips were performed by wooden vessels.

Long distant trips may also refer to rare cases of low accessibility to anchovy near the coast, due to either environmental conditions, or local depletion due to fishing pressure over the past days in the more coastal zones. Rather than willingly risk takers, these trips would correspond to 'forced visitors'. For better disentangling risk takers from visitors, the proportion of risk-taking trips by vessel should be computed, since trips from risk takers should often correspond to this cluster, while only a few trips from the visitors would be associated with this cluster. This is a feasible though difficult task, due to the current practice of skippers of changing the name of their vessels through the years.

'Southerners' was the second cluster to be formed. It is related to the particular management policy for that region, with very short fishing closures and less strict coastal restrictions for fishing (between 1.5 nm and 3 nm). For those reasons, fishing trips in this region typically last shorter and do not go very far from the coast (Fig. 5.5a). The percentage of fishing trips associated with this cluster, 15%, matches the average percentage of fishing trips performed in the south region between 2000 and 2008 (also 15%), computed from landing data (M. Bouchon, pers.comm.).

The 'wooden' cluster comprises more than 75% of the wooden-vessel trips. This cluster is mostly associated with short coastal trips (Fig. 5.5c). It also has a high percentage of trips from $2006-2009$. As mentioned in the introduction of this chapter, wooden vessels have been gradually introduced into the VMS. A large increase in the number of wooden-vessel trips monitored by VMS occurred in 2005 and the trend continued until 2009 (Table 5.5). Reports from IMARPE indicate that there was also a large increase in the number of wooden vessels operating per day at the end of 2005, and that this trend continued until the end of 2008 (IMARPE, 2013); so it is not just a VMS-coverage artifact. In 2009, the IVQ system was introduced, and the sizes of both steel and wooden fleets significantly decreased (the total number of fishing vessels operating by day decreased in 72%, Bouchon *et al.*, 2010b). The concentration of fishing trips of this cluster between 7°S and 9°S matches the spatial distribution of wooden vessels in the last years (M. Gutierrez, pers.comm.). This clustering identification supports the evidence shown by Bertrand (2013) that wooden and steel vessels present different spatial behaviors. The moves from multiple trajectories from steel and wooden vessels were modeled using random walks. Steel vessels showed more diffusive behavior and explored wider areas. Conversely,

wooden vessels showed less diffusion and explored smaller areas. Our results confirm that steel and wooden vessels deploy different movement strategies.

The fourth cluster is the largest one and is mostly composed of steel-vessel and *NC* trips. This cluster refers to 'ordinary' fishing trips (Fig. 5.5d). In relation to all the fishing trips in the dataset, trips in this cluster show average patterns.

It was also expected to find a cluster composed of 2009 trips, corresponding to the IVQ system. The fact that those trips were mostly distributed in two clusters ('wooden' and 'ordinary') could imply that there were no great changes in fishermen movement patterns and strategies when the race for fish went to an end. Interviews with fishermen and fleet managers suggest that the introduction of the IVQ system destabilized fishermen and companies strategies, and that they have been adapting progressively to the new context. Consequently, more years of VMS data would be required to accurately characterize changes in fishermen movement patterns with the IVQ system.

In this work, we explored fishing trip heterogeneity and found that fishermen personality, fleet segmentation and management regions conditioned 4 main fishing trip strategies: (1) risk-taking exploration, (2) short Southern trips, (3) follower wooden trips, (4) follower steel (ordinary) trips. These findings are key for understanding fishermen spatial behavior and effort.

From a methodological perspective, the HCA approach could also be used for outlier detection. The criteria used for keeping only anchovy fishing trips in the dataset is not free of errors (appendix A.1). In Figure 5.1, a few of the positioning records seem to correspond to jack mackerel or prospection trips. Anchovy represents by far the main landing in the Peruvian purse-seiner fisheries ($\sim$ 95% of the total of anchovy, sardine, Jack mackerel and Chub mackerel landings tonnes in 2007; Montecino and Lange, 2009). Thus, even without the pre-treatment algorithm (appendix A.1), only a small amount of non-anchovy trips would be mistaken as anchovy fishing trips. However, for a better identification of non-anchovy trips, it could be appealing to explore the 17% of the trips that belong to cluster 2 and separate them into sub-clusters, where one of them could be the outlier-non-anchovy-trip cluster. Overall, the methodology presented allows investigating fishing trip patterns at several levels of details (e.g., large clusters discriminating between the most distinct patterns of behavior, or small clusters for identifying outliers).

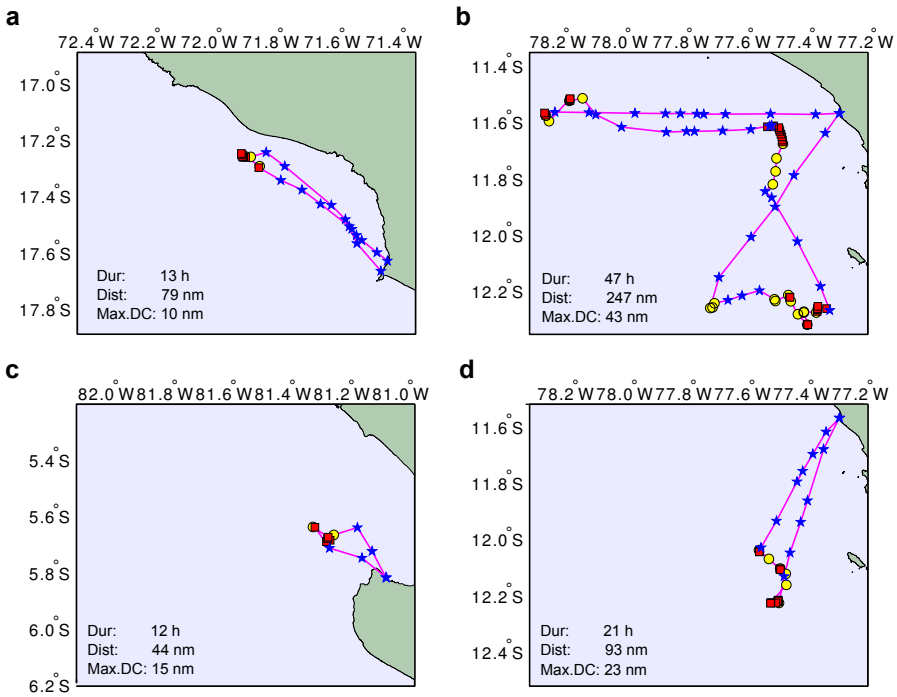


Figure 5.5: A fishing trip example associated to the first cluster (a), the second cluster (b), the third cluster (c) and the fourth cluster (d). Red squares represent fishing positioning records, yellow circles represent searching positioning records and blue stars represent cruising positioning records.

Chapter 6

Ecosystem scenarios shape fishermen spatial behavior

"L'homme est la seule créature qui refuse d'être ce qu'elle est."

– Albert Camus (*L'Homme Révolté*)

6.1 Introduction

Marine ecosystems are highly structured in space. Pelagic ecosystems, where habitats are made of constantly moving water masses, are also highly dynamic in time (Freón and Misund, 1999; Levin, 1992). Because those natural systems tend to be out of balance (Pimm, 1991), it is fundamental to consider their variability in space and time to understand the relative contributions of bottom-up and top-down controls in their functioning (Gripenberg and Roslin, 2007; Matson and Hunter, 1992; Polishchuk *et al.*, 2013). Also, since each organism tends to feed on smaller organisms, the rapid environmental variations tend to smooth out along trophic levels (Mann and Lazier, 2006).

Limitations on observation data to examine such processes have been largely overcome in the last decades thanks to new technologies and ecosystem models. There are now, in several ecosystems, spatially explicit data available at fine time resolutions on abiotic factors, prey and predator distribution and abundance (Bograd *et al.*, 2010; Boyd *et al.*, 2004; Costa, 1993; Decker and O'Dor, 2003; Rutz and Hays, 2009).

A few recent works undertook the study of the linkages between environmental conditions, prey distribution and predators behavior in time and space. Croll *et al.*

(2005) analyzed temporal lags in bottom-up physical and biological process and observed that they have possibly translated into spatial mismatches when relating whale abundance and physical forcing. They used relationships between upwelling, primary production and euphausiid prey and whale foraging behavior for characterizing Monterey Bay (California) as an important foraging area for blue whales. Hazen *et al.* (2011) found that both environment factors (i.e. bottom depth, salinity, temperature and turbulence level) and prey density were important for modeling foraging activity of beaked whale in the Tongue of the Ocean, Bahamas. Santora *et al.* (2012) showed significant coupling between physics (e.g. pycnocline depth, salinity, temperature and wind, among others), primary productivity, micronekton abundance and top predator counts in the Central California Current region.

Those studies identified critical processes and key areas that can be directly useful in conservation, marine spatial planning, or in the management of human activities in marine ecosystems. Nevertheless, while fishing is one of the main anthropogenic activities at sea, competing directly with natural predators, none of those studies on spatial ecology from physics to top predators contemplate fishermen as part of the system. Admittedly, fishermen are peculiar top predators because they rely on technology and are driven by a distinct currency than that of natural predators. Nonetheless, in pelagic ecosystems water masses and fish schools are constantly moving (Peraltilla and Bertrand, 2013; Swartzman *et al.*, 2008). Thus, fishermen do face the same uncertainty on prey localization as natural predators do and their spatial behavior reflects their need for solving the same challenge. In that sense, it has been shown that fishermen in the Peruvian anchovy fishery deploy similar spatial foraging strategies to those of other animal predators (Bertrand *et al.*, 2007).

The Northern Humboldt Current System (NHCS) off Peru provides a great opportunity for studying the association between the dynamics of ecosystem components in a bottom-up-controlled exploited ecosystem (Ayón *et al.*, 2008b; Bertrand *et al.*, 2008b). First, the NHCS is submitted to an intense regional climatic variability at a variety of spatio-temporal scales (Chavez *et al.*, 2008). The resulting environmental scenarios directly determine the extent of the 3D anchovy habitat (Bertrand *et al.*, 2011, 2004a), which in turn conditions fish availability for the main predators in this system, the fishermen (Bertrand *et al.*, 2008b). Second, the NHCS produces more fish per unit area than any other region in the world oceans and sustains the world's largest monospecific fishery (Peruvian anchovy or anchoveta,

Engraulis ringens). To cope with the intense climatic variability, the management of the anchovy fishery is adaptive, i.e. catch limits are re-assessed every $\sim$ 6 months and opening and closure periods decided on the basis of daily monitoring of the ecosystem, the fish population and the fishery (Chavez *et al.*, 2008). The Peruvian Marine Research Institute (IMARPE) is in charge of this intense monitoring. It comprises satellite information on environmental conditions (e.g. sea surface temperature, Chlorophyll-a and sea level anomaly, among others) in daily and weekly resolutions. Fish population distribution and biomass are monitored through scientific acoustic surveys (two to three times a year). The fishing activity is supervised through landing statistics, Vessel Monitoring System (VMS) and on-board observers reports (Bertrand *et al.*, 2008c; Joo *et al.*, 2011). The amount of available data makes the NHCS a highly appealing ecosystem for analyzing its ecological dynamics. In this highly-variable and data-rich ecosystem, Bertrand *et al.* (2008b) analyzed how large scale oceanic forcing, via Kelvin waves, affected the coastal ecosystem (from oceanography to fishermen). They pioneered the incorporation of fishermen as a top predator for studying ecological dynamics and proposed contrasting scenarios of coastal oceanography, anchovy distribution and fishing activity, during the passage of coastally trapped upwelling and downwelling Kelvin waves.

The present work focuses more closely on the coastal processes in the NHCS, and on the spatial response of fishermen to varying environmental and anchovy conditions. In particular, we explore and quantify the associations between the dynamics of three ecosystem compartments: environmental conditions (Environment), anchovy biomass and distribution (Anchovy) and fishermen spatial behavior (Fishermen), for a ten-year period ($2000 - 2009$). Data on Environment and Anchovy were issued from acoustic surveys performed by IMARPE and satellite observations. Data on Fishermen were based on VMS data ($\sim$ 90000 fishing trips from 2000 to 2009), processed with a state-space model so that the nature of the behavior in which fishermen are engaged is known at each position (Joo *et al.*, 2013). Since environmental fluctuations tend to smooth out along trophic levels (Mann and Lazier, 2006), we hypothesized that, from the three associations (i.e. Environment-Anchovy, Anchovy-Fishermen and Environment-Fishermen), Anchovy-Fishermen should be the strongest since it is a direct prey-predator relationship. Conversely, Environment-Fishermen is expected to display the weakest association. The studied decade does not encompass strong ENSO (El Niño Southern Oscillation) events. Season is thus expected to be the major scale of variability for environmental conditions. We tested for differences between two seasonal modes, summer and spring/winter,

within each ecological compartment and analyzed trends separately for each seasonal mode. We finally propose ecosystem scenarios based on the linkages found between the three compartments and discuss the potential use of fishermen spatial behavior as ecosystem indicator.

6.2 Materials and methods

We focused on different time-periods from 2000 to 2009 such that concomitant data on the environment, fisheries acoustic and VMS were available. In total 16 time-periods were available, 6 in austral summer and 10 in austral winter/spring (Table 6.1).

Table 6.1: Number of fishing trips corresponding to each time-period.

Time-period	Fishing trips
2000/06 – 07	5839
2000/10 – 11	5750
2001/03 – 04	7012
2001/07 – 08	865
2001/10 – 11	1612
2002/02 – 03	1368
2002/10 – 11	6409
2003/10 – 12	6262
2004/11 – 12	8983
2005/11 – 12	15252
2006/02 – 04	2990
2006/11 – 12	8593
2007/02 – 04	2395
2008/02 – 04	4611
2008/11 – 12	8604
2009/02 – 04	3151

6.2.1 Environmental data

For each time-period, we produced a description of the environment as detailed in Table 6.2. We used Sea surface temperature (SST) from the AVHRR sensor of NOAA satellites from 2000 to 2009. The chlorophyll-a (CHL) satellite data from

2000 to 2007 were obtained from the SeaWiFS sensor and from 2008 to 2009 from the MODIS sensor. The MODIS chlorophyll-a data were previously corrected on a monthly basis to match the SeaWiFS data by using the common period between both sensors ($2002 - 2008$). All satellite data, initially in a spatial resolution of 4 km and a weekly time resolution, were averaged by month and over the whole study area. Then, for obtaining one representative value per time-period, an average was computed for each of these variables.

To account for the vertical distribution of the oxygen minimum zone, a critical parameter of the NHCS (Bertrand *et al.*, 2011), we computed the depth of the oxycline (OXY), i.e. the depth at which the dissolved oxygen equals 2 $ml.l^{-1}$. All measurements were made from Niskin bottles and CTD data sampled during scientific surveys run by IMARPE. For each time-period, the mean oxycline depth was computed as a weighted average of monthly values, taking into account the number of observations per month.

6.2.2 Anchovy data

Since 1983, IMARPE has been conducting on average two acoustic surveys per year for monitoring fish population distribution and biomass. These surveys consist of parallel cross-shore transects of $\sim$ 100 nm long, with a $\sim$ 15 nm spacing. Simrad (kongsberg Maritime AS, Norway) scientific echosounders working at several frequencies are used to estimate biomasses (see Castillo *et al.*, 2009; Gutiérrez *et al.*, 2007; Simmonds *et al.*, 2009). An extensive midwater-trawl sampling completes the acoustic surveys for species identification. The nautical-area-backscattering coefficient (NASC, in $m^2.mn^{-2}$), an index of fish biomass (Simmonds and MacLennan, 2005), is recorded at each georeferenced elementary distance sampling unit (EDSU) of 1 nm. From each acoustic survey, we extracted five descriptors of the spatial distribution of anchovy (Table 6.2): (i) the mean anchovy NASC, used as an index of fish biomass (s_A); (ii) an index of local biomass (s_A^+), i.e. the mean anchovy NASC for ESDU with presence of fish; (iii) an index of spatial occupation (ISO) i.e. the percentage of ESDU with anchovy; and (iv) the center of gravity and the inertia of the distance to the coast of anchovy NASC (DC and I, respectively).

Table 6.2: Description of Environment, Anchovy and Fishermen variables.

Ecosystem compartment	Variable acronym	Description	Data type	Scale of data
Environment	SST	Averaged sea surface temperature from 3°S to 16°S and from the shore up to the 2500 m isobath (in °C)	AHVRR	Monthly data
	CHL	Averaged Chlorophyll-a from 3°S to 16°S and up to the 2500 m isobath (in mg.m^{-3})	SEAWIFS and MODIS	Monthly data
	OXY	Average of oxycline depth from 7°S to 16°S (in m)	CTD and NISKIN	Scientific survey
Anchovy	s_A	Index of global fish acoustic biomass equal to $\log((\sum_{i=1}^{n} \text{NASC}_i/n)+1)$ (in m^2.nm^{-2})	Output from acoustic surveys (Bertrand et al., 2004)	Acoustic survey
	s_A^+	Index of local fish biomass equal to $\log((\sum_{i=1}^{n} \text{NASC}_i/n)+1)$, $\forall \text{NASC} > 0$ (in m^2.nm^{-2})	Output from acoustic surveys (Gutierrez et al., 2007)	Acoustic survey
	ISO	Index of spatial occupation. Percentage of ESDU where anchovy is present (NASC > 0)	Output from acoustic surveys (Gutierrez et al., 2007)	Acoustic survey
	DC	Center of gravity of the distance to the coast of the acoustically observed anchovy (in km) $\text{DC} = \sum_{i=1}^{n} (\text{dc}_i \log(\text{NASC}_i+1)) / \sum_{i=1}^{n} \log(\text{NASC}_i+1)$	Output from acoustic surveys (Gutierrez et al., 2007)	Acoustic survey
	I	Inertia of DC (in km) $\text{I} = \sum_{i=1}^{n} (\text{dc}_i - \text{DC})^2 \log(\text{NASC}_i+1) / \sum_{i=1}^{n} \log(\text{NASC}_i+1)$	Output from acoustic surveys	Acoustic survey
Fishermen	Dur	Fishing trip duration (in hours)	Computed from VMS data	Fishing trip
	Dist	Distance traveled during a fishing trip (in nm)	Computed from VMS data	Fishing trip
	Max.DC	Maximum distance to the coast during a fishing trip (in nm)	Computed from VMS data	Fishing trip
	Searching	Proportion of fishing trip duration spent searching	Model output from VMS and observers data (Joo et al., in press)	Fishing trip
	Fishing	Proportion of fishing trip duration spent fishing	Model output from VMS and observers data (Joo et al., in press)	Fishing trip
	Cruising	Proportion of fishing trip duration spent cruising	Model output from VMS and observers data (Joo et al., in press)	Fishing trip
	Bef.Fishing	Absolute time from the beginning of the trip until the first fishing set (in hours)	Model output from VMS and observers data (Joo et al., in press)	Fishing trip
	k	Shape parameter from random walk modeling	Model output from VMS data **(Bertrand et al. in review)**	Vessel moves by season
	sigma	Scale parameter from random walk modeling	Model output from VMS data **(Bertrand et al. in review)**	Vessel moves by season

Notes: NASC_i: nautical-area-backscattering coefficient at i^{th} georeferenced elementary distance sampling unit (ESDU). dc_i: distance to the coast at the i^{th} ESDU.

6.2.3 Fishermen spatial behavior

The fisheries management and VMS data have been described in the preceding chapters. For this work, only fishing trips from the industrial fleet in the north-center (3°S to 16°S) Peruvian coast were considered (the southern region was excluded from the analysis because of its lesser contribution to the catches and a different management policy). We report in Table 1 the number of fishing trips documented for each time-period.

For each fishing trip, the following global indicators were computed: the duration (Dur), the total distance traveled (Dist) and the maximum distance from the coast (Max.DC). Additionally we used two metrics, k and sigma, that characterize the geometry of the fishermen movements (Bertrand *et al.*, in review). They are derived from the fit of a Generalized Pareto random walk model to the move length distribution displayed in fishing trip trajectories. This random walk model involves two parameters of interest: a shape parameter k that characterizes the movement diffusion (finite tails with normal diffusion for $k < 0$, light tails with normal diffusion for $k \in [0; 0.5]$, and heavy tails with super diffusion for $k > 0.5$), and a scale parameter sigma (where $sigma > 0$) that characterizes the average length of the moves used to search for fish. Super diffusion is characterized by the mixing of rare, long and relatively straight movements with movement bouts composed of short moves and high turning rates. Descriptors k and sigma were computed for each vessel from the fishing trips performed within each time-period.

As in the preceding chapter, we also computed indicators on fishing activities inferred with a hidden semi-Markov model trained on logbook data (Joo *et al.*, 2013). Those indicators were: the proportion of time spent searching (Searching), fishing (Fishing) and cruising (Cruising), and the time spent from the beginning of the trip until the first fishing set (Bef.Fishing).

6.2.4 Statistical analyses

A synoptic sketch of the statistical analyses between the ecosystem compartments is reported in Figure 6.1. Univariate statistics were computed on each variable for testing for differences between seasonal modes, characterizing overall tendencies and potentially remarkable time-periods. The Wilcoxon rank sum test (Wilcoxon, 1945) was performed for testing mean differences between summer and winter/spring

periods. Trends were computed using least trimmed squares robust regression (Rousseeuw, 1984).

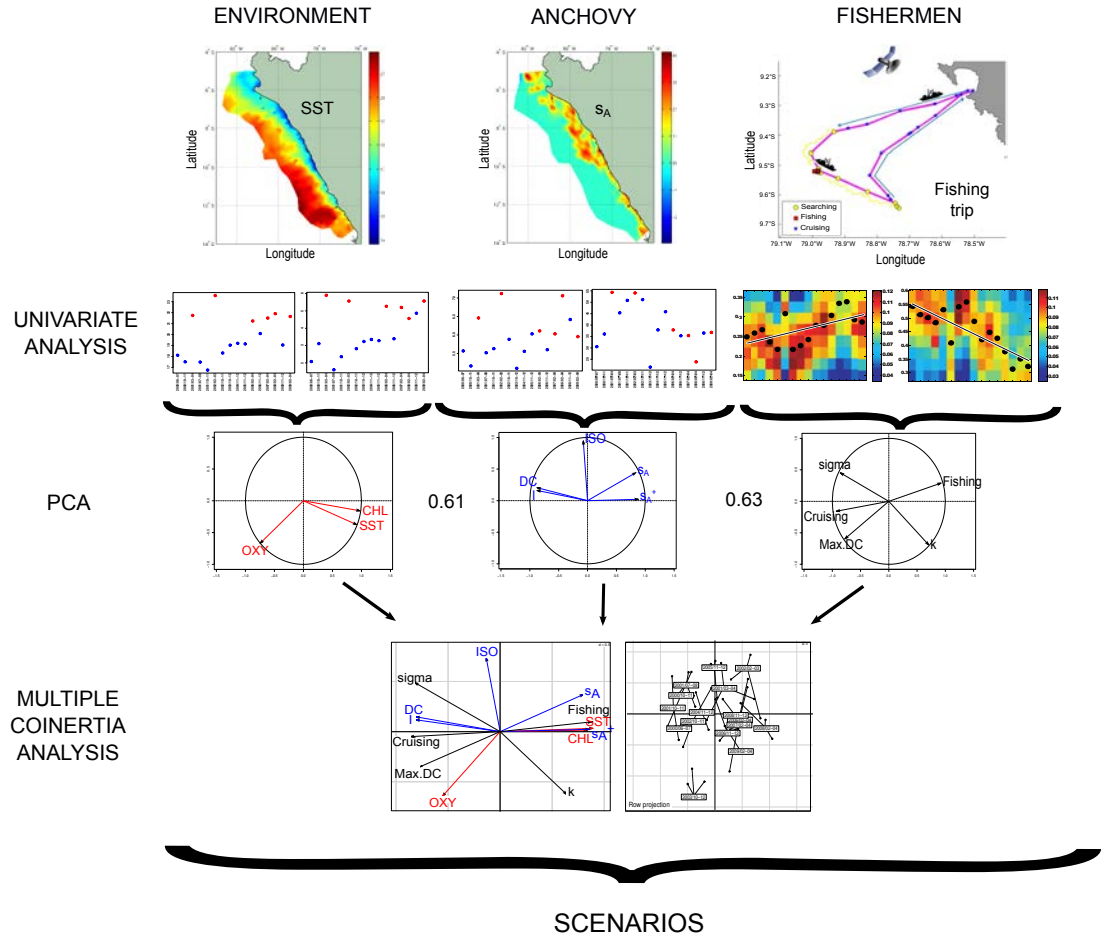


Figure 6.1: Synoptic sketch of the statistical analyses.

For multivariate analyses, all variables were standardized. We carried out a PCA (Principal Component Analysis; Pearson, 1901) for each compartment at the scale of the time-periods to identify the main modes of variability and distinct scenarios. For quantifying the linear association between the first principal component of Environment, Anchovy and Fishermen, we computed the Pearson product-moment correlation coefficient between each pair of first components. In addition, in order to quantify the overall association between each pair of compartments (Environment and Anchovy, Anchovy and Fishermen, and Environment and Fishermen) we used RV coefficients (Escoufier, 1973). The RV coefficient is a multivariate generalization of the squared Pearson correlation coefficient. RV values range between 0 and 1, where 0 means no association and 1 means perfect association. The significance of the RV coefficient was estimated by means of a permutation test (Heo and Gabriel,

1998).

Finally, we performed a multiple coinertia analysis (Chessel and Hanafi, 1996) between the three PCAs. The multiple coinertia analysis is a generalization of the coinertia analysis (Dolédec and Chessel, 1994; Dray *et al.*, 2003) for k tables (k > 2). The coinertia analysis is a multivariate method for coupling two tables sharing the same rows (individuals or variables). The tables must be previously analyzed by separate inertia analyses (for instance, PCAs). The multiple coinertia analysis aims at (i) calculating a synthetic table representing the common structure of the PCAs, and (ii) comparing each PCA to this synthesis. RV coefficients measure the association of each PCA and the synthetic table.

Statistical analyses were performed with R software (R Core Team, 2013). Robustbase package (Rousseeuw *et al.*, 2012) was used for least trimmed regression, FactoMineR package (Husson *et al.*, 2013) was used for PCA and ade4 (Dray *et al.*, 2007) for multiple coinertia analyses.

6.3 Results

6.3.1 Univariate analyses

Series of environmental conditions are shown in Figure 6.2a,b,c. SST and CHL values were higher in summer than in winter/spring periods ($p < 0.001$ for both variables) and exhibited increasing trends ($p < 0.01$ for both variables) on winter/spring periods. For OXY no significant differences between seasons were observed and no trend was observed either. Still, the October - December 2003 time-period (denoted 2003/10-12 hereafter) was characterized by a remarkably deep oxycline.

Fish distribution patterns also changed in time (Fig. 6.2d,e,f,g,h). The s_A^+ and s_A indexes differed significantly among seasonal modes ($p < 0.01$) and an increasing trend ($p < 0.01$) was observed for s_A^+ in winter/spring periods. Two periods presenting the highest s_A (and s_A^+) values lied out of the trend: 2008/02-04 and 2002/02-03. The former period, 2008/02-04, along with 2003/10-12, were also characterized by the lowest ISO. Finally, 2000/06-07, 2000/10-11 and 2001/10-11 had the highest DC and I.

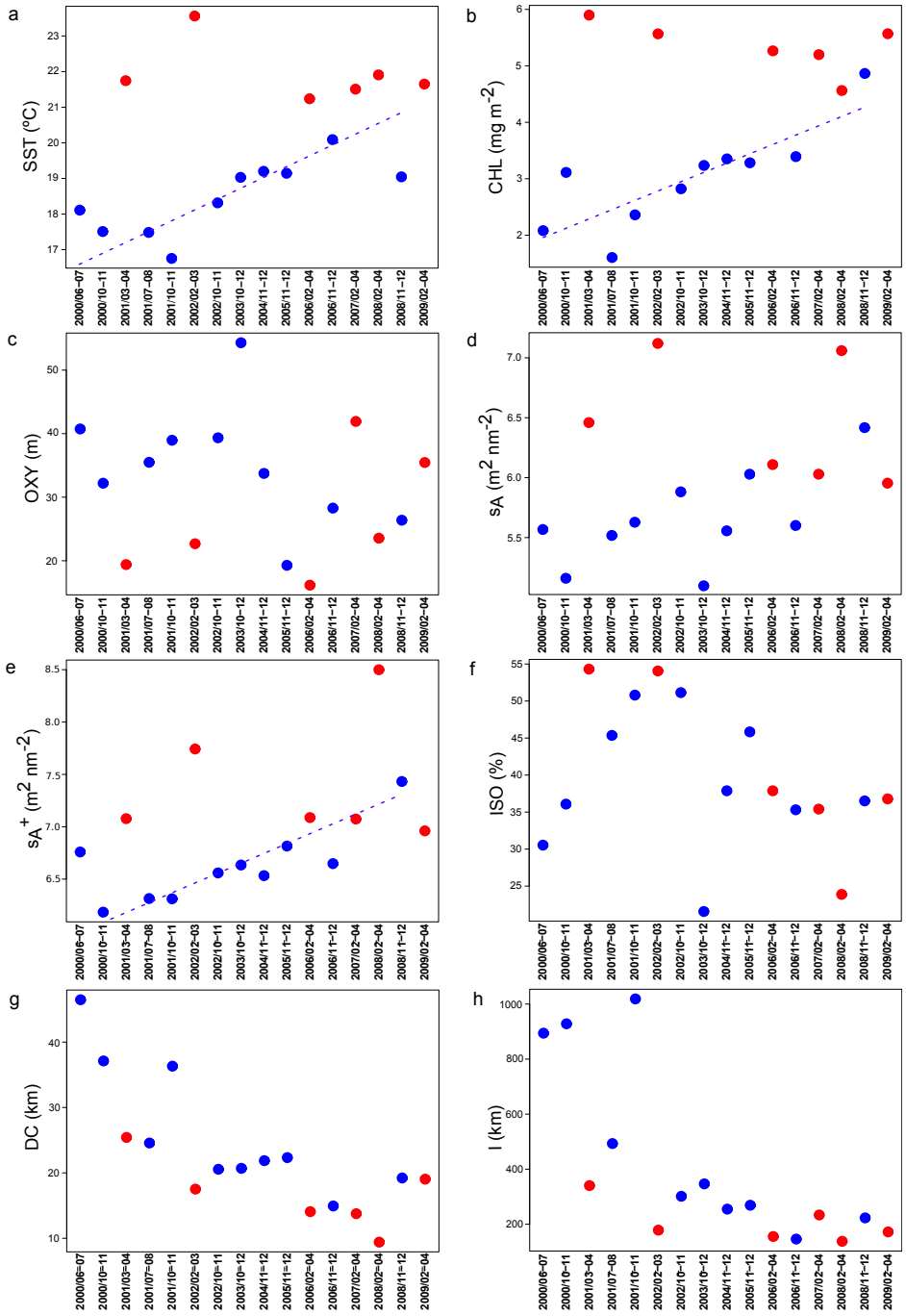


Figure 6.2: Univariate series of Environment and Anchovy variables: sea surface temperature or SST (a), chlorophyll-a or CHL (b), oxycline depth or OXY (c), global biomass or s_A (d), local biomass or s_A^+ (e), index of spatial occupation or ISO (f), center of gravity of distance from the coast or DC (g), inertia of DC or I (h). Blue points correspond to winter/spring time-periods, and red points correspond to summer periods. Blue dashed lines indicate the significant winter/spring trends. Here, the time-periods at the x axes are plotted at regular steps, but the trends were actually fitted considering the real intervals between time-periods.

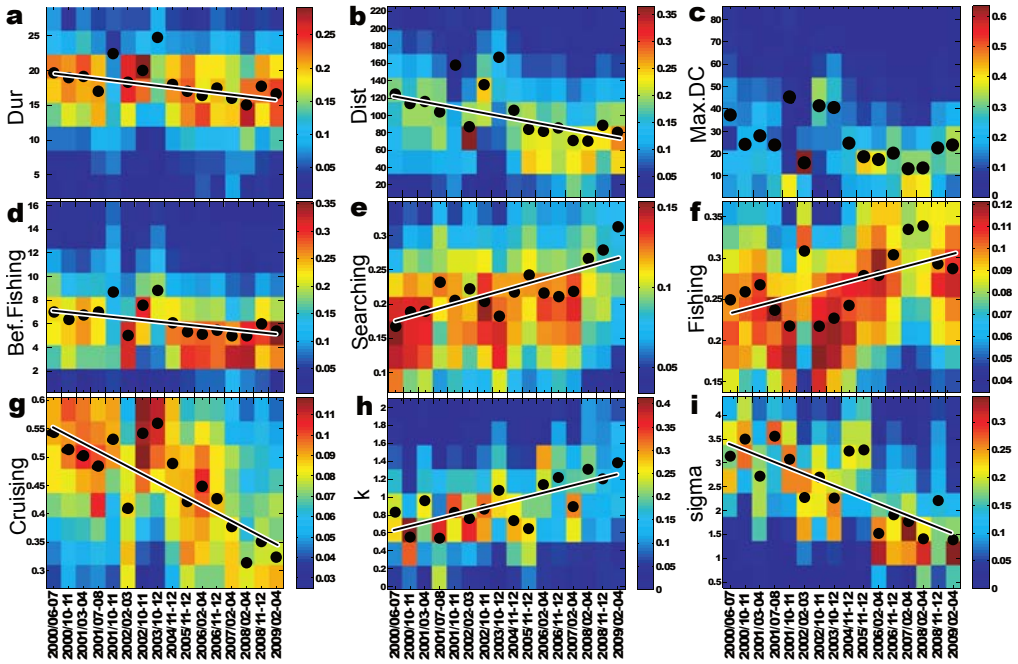


Figure 6.3: Marginal distributions and median by time-period for each Fishermen variable: fishing trip duration (a), distance traveled (b), maximum distance from the coast (c), absolute time from the beginning of the trip until the first fishing set (d), proportion of trip duration spent searching (e), proportion of trip duration spent fishing (f), proportion of trip duration spent cruising (g), shape parameter from random walk modeling (h), scale parameter from random walk modeling (i). Blue solid lines indicate significant trends ($p < 0.01$) over the whole studied period. Here, the time-periods at the x axes are plotted at regular steps, but the trends were actually fitted considering the real intervals between time-periods. Complete ranges of values are shown in Figure 6.4; here, a zoom is made over the values associated with the highest marginal distributions.

Since hundreds of fishing trips were comprised in each time-period (Table 6.1), we analyzed the temporal evolution of both the median values and the marginal distribution of each Fishermen variable. In order to facilitate visualization, a zoom was made on the range of values associated with the highest marginals (Fig. 6.3). The complete range of values is available in Figure 6.4.

Only Cruising showed significant differences between seasonal modes ($p < 0.01$). Therefore, for Fishermen variables robust trends were computed for the whole series without discriminating between seasonal modes. Negative trends were observed for Dur, Dist, Bef.Fishing, Cruising and sigma ($p < 0.0001$). By contrast, Searching, Fishing and k reported increasing trends ($p < 0.0001$). Indeed, medians of 2001/10-11, 2003/10-12 and to a lesser extent 2002/10-11 presented the highest values associated with Dur, Dist, Max.DC, Bef.Fishing and Cruising. 2001/10-11 and 2003/10-12

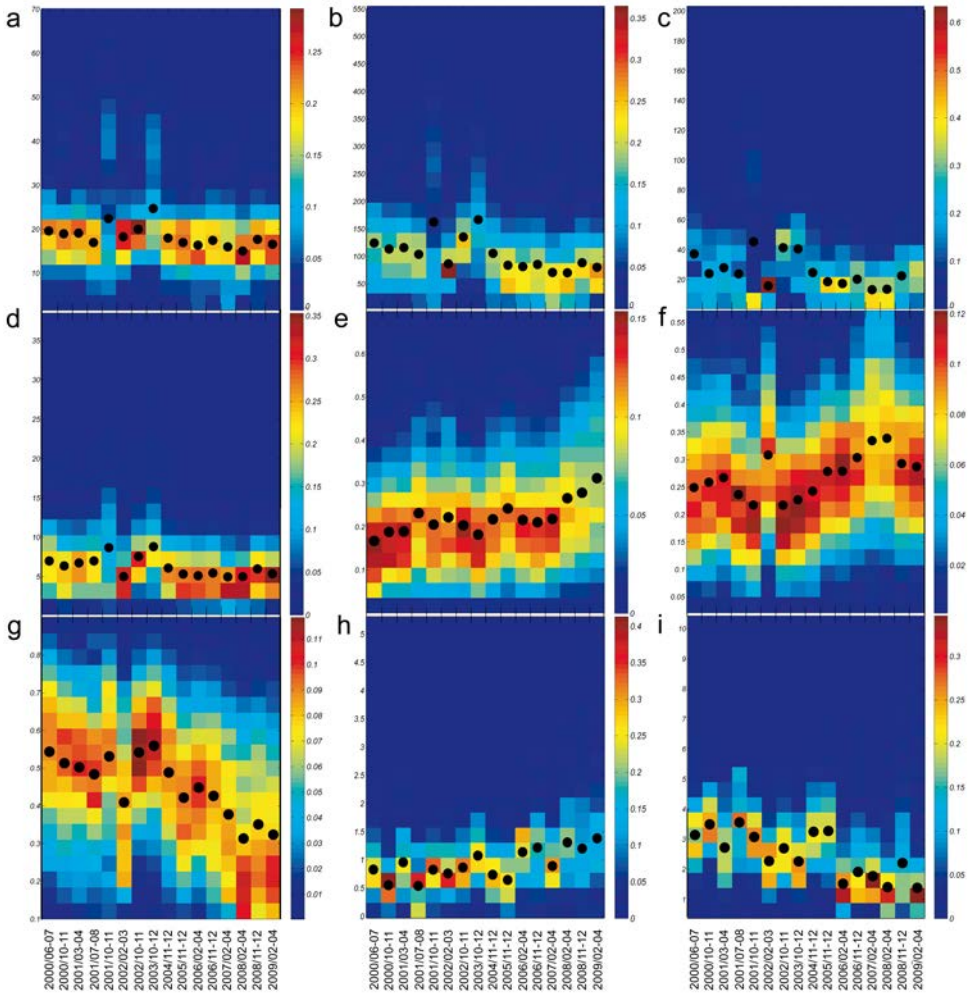


Figure 6.4: Marginal distributions and median by time-period for each Fishermen variable: fishing trip duration (a), distance traveled (b), maximum distance from the coast (c), absolute time from the beginning of the trip until the first fishing set (d), proportion of trip duration spent searching (e), proportion of trip duration spent fishing (f), proportion of trip duration spent cruising (g), shape parameter from random walk modeling (h), scale parameter from random walk modeling (i).

presented the highest variability for Dur, Dist, Max.DC, Bef.Fishing (marginal distributions in Fig. 6.3a,b,c,d). Conversely, in 2002/02-03, Dist, Max.DC, Bef.Fishing and Cruising were low and Fishing was high.

6.3.2 Multivariate analyses

In order to simplify the interpretation of the analyses, we do not present here the graphical representations of the spaces of individuals and variables of the PCAs. Significant correlations between variables and PCAs are shown in Tables 6.3, 6.4, 6.5. We also present the scores of each time-period for the main principal components (Fig. 6.5). For each score time-series, we provide an arbitrary classification from (−) to (+) indicating how those scores impacted fishermen activity (Fig. 6.5).

Table 6.3: Significant correlations between Environment variables and their principal components ($p < 0.05$).

Variables	PC1 (75%)	PC2 (21%)
CHL	0.93	–
SST	0.94	–
OXY	−0.71	−0.70

Table 6.4: Significant correlations between Anchovy variables and their principal components ($p < 0.05$).

Variables	PC1 (60%)	PC2 (23%)
s_A^+	0.88	–
s_A	0.84	–
DC	−0.87	–
I	−0.87	–
ISO	–	0.94

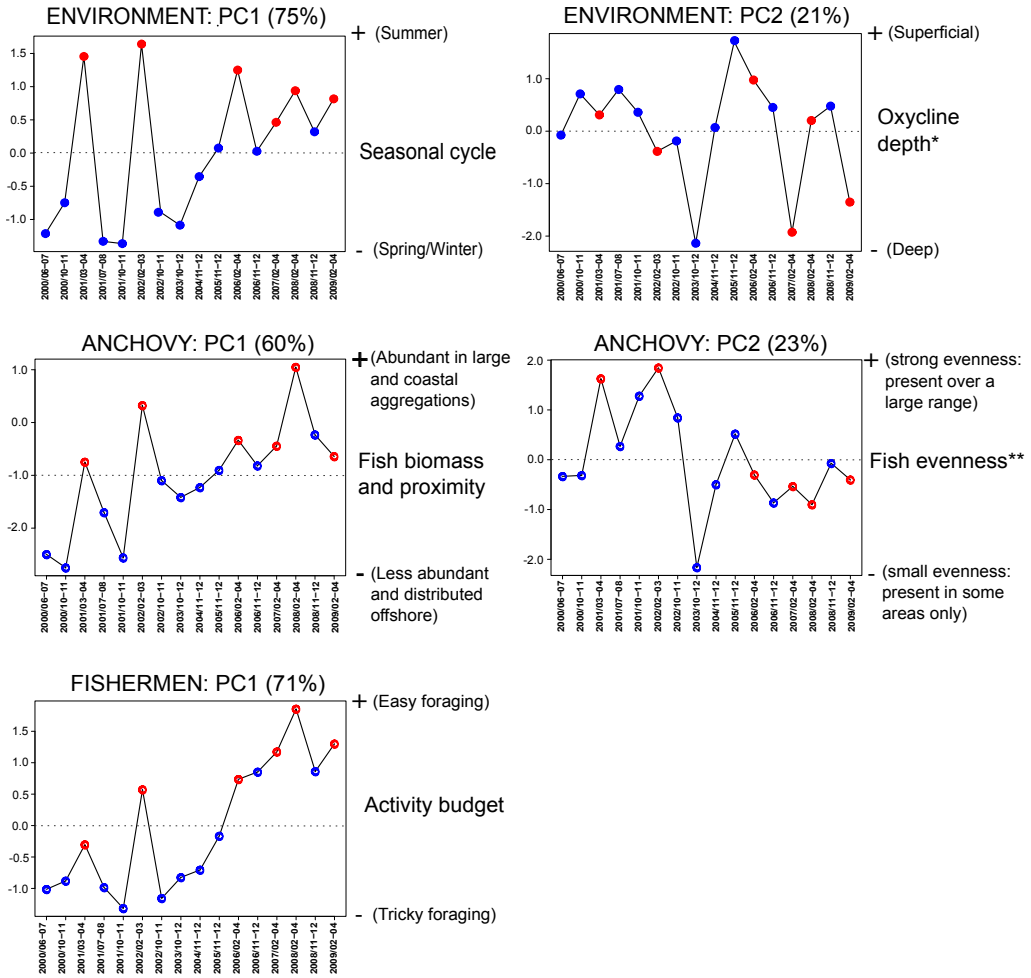


Figure 6.5: Time series of PCA scores for each ecosystem compartment. Percentage of explained variance are in parentheses at the top of each subfigure. Red points correspond to summer time-periods and blue points to winter and spring periods. At the right side of each subfigure, a label is given to each principal component. (+) and (-) symbols indicate favorable and unfavorable conditions for Fishermen, respectively. *Oxycline depth follows the seasonal cycling. But some specific events such as downwelling Kelvin wave could deepen the oxycline. **This component is only useful when comparing similar scenarios in terms of s_A, s_A^+ and DC.

Table 6.5: Significant correlations between Fishermen variables and their principal components ($p < 0.05$).

Variables	PC1 (71%)	PC2 (23%)
Fishing	0.92	–
k	0.70	−0.69
Searching	0.67	–
Dur	−0.70	–
Max.DC	−0.78	−0.59
Bef.Fishing	−0.81	–
Dist	−0.83	–
sigma	−0.86	–
Cruising	−0.93	–

Environmental conditions

In the Environment PCA, the first axis, which accounted for 75% of the variance, was correlated with CHL and SST (positively), and with OXY (negatively, Table 6.3). These relations mainly refer to the seasonal cycles. During summer, waters are warmer (higher SST), and due to a reduced cloud cover, more light penetrates into the water column, what positively impacts the productivity at sea (higher CHL; Echevin *et al.*, 2008). Thermocline and oxycline are also shallower during summer (lower OXY; Gutiérrez *et al.*, 2011). The opposite phenomenon occurs during winter. Scores reflected the seasonality: all summer periods had higher scores than winter and spring periods (Fig. 6.5; $p < 0.001$ for differences in scores between seasonal modes). The second component represented 21% of the variance and was significantly correlated with OXY (Table 6.3). This component did not reflect a significant seasonality. Although oxycline depth is related to the seasonal cycle, specific large scale forcing events such as downwelling Kelvin Waves can deepen the oxycline (Gutiérrez *et al.*, 2008). Within the studied period, 2003/10-12 and 2007/02-04 had the lowest scores (deepest oxycline depth) for this component (Fig. 6.5).

Anchovy distribution

The first principal component, accounting for 60% of the variance, was positively correlated with s_A and s_A^+, and negatively correlated with DC and I (Table 6.4). This axis can be interpreted as an index of fish biomass and proximity to the coast.

Scores reflected, although with less intensity, the seasonality observed in the first axis of the Environment PCA (Fig. 6.5; $p < 0.01$). A significant increasing trend in time was observed for this first component ($p < 0.01$) through a linear regression. The second component, explaining 23% of the variance, was significantly correlated to ISO (Table 6.4). It can be seen as fish evenness, and it needs to be interpreted considering the scenarios in terms of s_A, s_A^+ and DC: high evenness is favorable to fishermen activity if the global biomass is high, since there is lot of anchovy everywhere. On the opposite, if the biomass is low a high evenness means highly dispersed anchovy (low s_A^+), which are difficult to catch (Bertrand *et al.*, 2004b). The latter scenario corresponds to 2003/10-12, which had by far the lowest score (Fig. 6.5).

Fishermen spatial behavior

Due to the relatively high number of variables for the small number of individuals, we discarded from the multivariate analyses the variables presenting very strong correlations ($r > 0.9$; i.e. Dist., Bef.Fishing and Dur, all of them strongly correlated with Cruising). We also discarded Searching since the proportions of the time spent Cruising, Fishing and Searching sum to one (no variable should be expressed as a linear combination of others). Overall, we used Max.DC, Fishing, Cruising, k and sigma. The other variables were still used as supplementary variables and projected in the PCA space (Fig. 6.6). The first axis, explaining 71% of the variance (Table 6.5), can be viewed as fishermen activity budget. High values on this activity budget axis correspond to easy foraging (more time is allocated to searching and fishing than to cruising), while low values indicate tricky foraging (more time is dedicated to cruising, longer distances are traveled, more time is spent before the first fishing set, farther distances from the coast are traveled). Through a linear regression, scores showed an increasing trend (towards easy foraging) with time ($p < 0.0001$), taking negative values until 2005/11-12 (except 2002/02-03) and positive then after (Fig. 6.5). The second component, explaining 23% of the variance (Table 6.5), had significant correlations with only two variables (k and Max.DC).

Environment vs. Anchovy vs. Fishermen

Regarding the association between the first principal components only, the activity budget component of Fishermen PCA was strongly correlated with the fish biomass and proximity component of Anchovy PCA (0.83; $p < 0.0001$). The latter was also

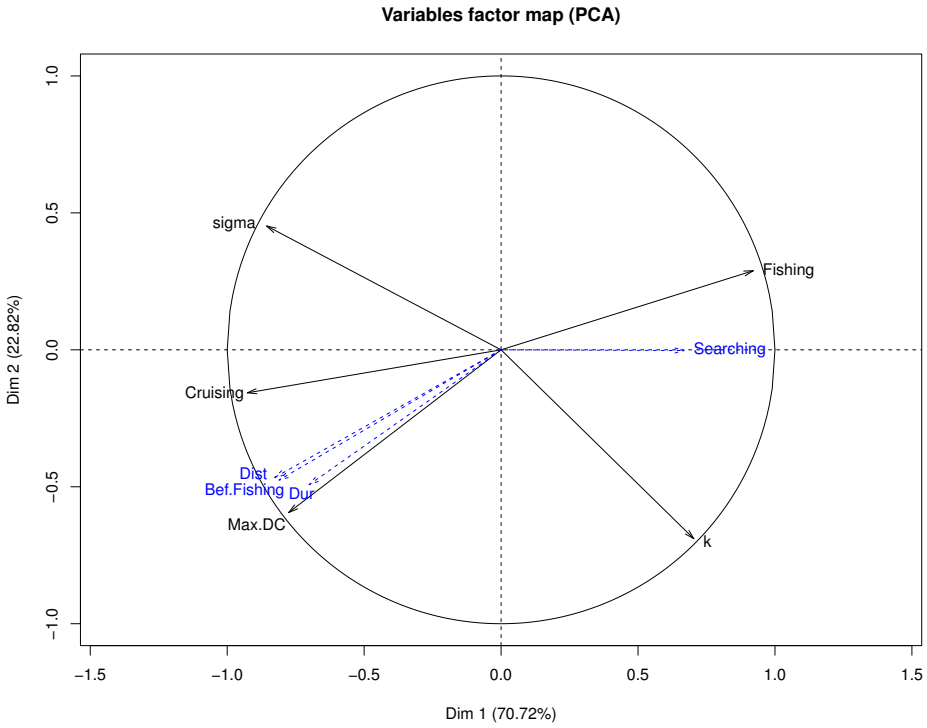


Figure 6.6: PCA of Fishermen. Components 1 and 2

correlated with the seasonal cycle component of Environment PCA (0.81; $p < 0.001$). The seasonal cycle was also correlated with activity budget, though more loosely (0.76; $p < 0.001$). As we hypothesized and consistent with the principal component correlations, RV coefficients were likewise larger for more direct relationships, i.e. Environment and Anchovy (0.61; $p < 0.001$) and Anchovy and Fishermen (0.63; $p < 0.001$) than for the less direct one: Environment and Fishermen (0.56; $p < 0.01$). Those significant associations provide thus a powerful way to quantify the strength of the relationships between ecosystem compartments (Fig. 6.7).

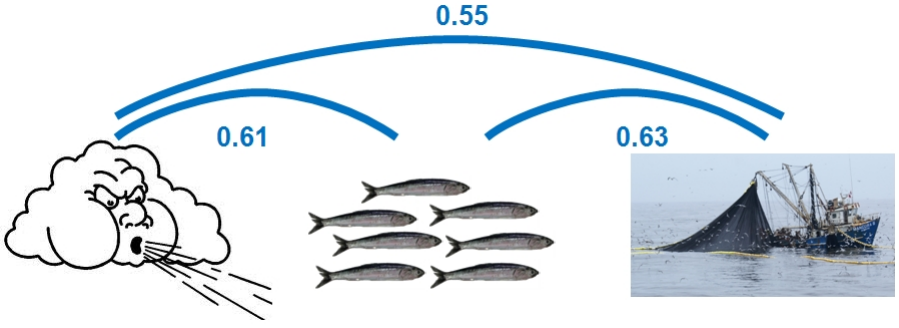


Figure 6.7: RV coefficients for each pair of ecosystem compartments: Environment-Anchovy, Anchovy-Fishermen and Environment-Fishermen

A multiple coinertia analysis was applied to Environment, Anchovy and Fishermen PCAs (Fig. 6.8). The multiple coinertia space was composed of two axes, accounting for 73% and 17% of the variance each. RV coefficients between the synthetic table and each of the compartments were 0.67, 0.73 and 0.75, respectively ($p < 0.0001$ for all cases). High SST and CHL values were strongly related to high s_A^+ (local biomass) and Fishing, and low Cruising, DC and I (distance from the coast of anchovy and its inertia). To a lesser extent, high SST and CHL were also related to high s_A and low Max.DC (distance to the coast of fishermen). Deep oxycline (OXY) was partly associated with low s_A and ISO (anchovy spatial occupation), and high Max.DC. Low ISO was also associated with high k (diffusivity of fishermen movements). Regarding the time-periods, some of the most remarkable ones already depicted in the preceding analyses were: 2008/02-04, 2002/02-03, 2003/10-12 and 2001/10-11. 2002/02-03 was characterized by a superficial OXY, very high ISO and s_A, and to a lesser extent, high SST, CHL, s_A^+ and Fishing. 2008/02-04 was characterized by the highest SST, CHL, s_A^+ and Fishing, and low DC, I, Cruising and Max.DC. 2003/10-12, on the contrary, corresponded to the deepest OXY, the lowest ISO, low s_A, the highest Cruising, high Max.DC and k. 2001/10-11 corresponded to the lowest SST, the highest I, a very high ISO, the highest Max.DC and low Fishing.

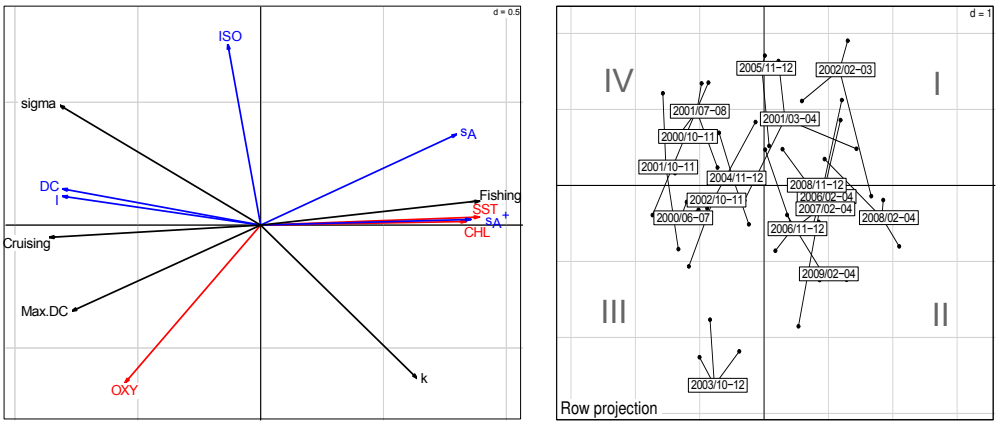


Figure 6.8: Multiple coinertia analysis between Environment, Anchovy and Fishermen compartments. Left panel: projection of variables in the coinertia space. The definition of each variable is in Table 6.2. Fishermen variables are in black, Anchovy in blue and Environment in red. Right panel: projection of time-periods in the coinertia space. Each period has three associated points, corresponding to the position of the period described by each compartment. The position of the label of each period corresponds to its location according to the synthetic table. Roman numerals indicate the different quadrants of scenarios to which the time-periods are associated.

6.4 Discussion

6.4.1 Environment, Anchovy, Fishermen and the bottom-up transfer

In this study, we explored and quantified the associations between the dynamics of three ecosystem compartments (Environment, Anchovy and Fishermen) and showed that (i) ecosystem scenarios – abiotic and biotic conditions – do shape fishing spatial behavior, and (ii) the association is stronger for direct links (Environment-Anchovy and Anchovy-Fishermen; Fig. 6.7). This is consistent with a bottom-up transfer, which was already evidenced by Bertrand *et al.* (2008b) when analyzing the effect of upwelling and downwelling Kelvin waves on several descriptors of the Peruvian coastal oceanography, anchovy distribution and fishing fleet behavior. They characterized contrasting ecological scenarios that corresponded to strong El Niño events (warm scenarios) and to average upwelling periods (cold scenarios).

Unlike Bertrand *et al.* (2008b), we focus on the coastal processes of the NHCS, and our studied period, $2000 - 2009$, does not encompass strong El Niño events, but only 'weak' events (2002-03, 2004-05 and 2009-10). These events correspond to 'El Niño Modoki' or 'Central Pacific El Niño' that are distinct from the extraordinary warm events such as the ones from 1982-83 and 1997-98 (Takahashi *et al.*, 2011). When Central Pacific El Niño occurs, the conditions off Peru can even be slightly cooler than average (Dewitte *et al.*, 2012), far from the archetypal El Niño effects on the Peruvian ecosystem (Alheit and Ñiquen, 2004; Barber and Chavez, 1983; Bertrand *et al.*, 2004a). The studied period therefore lies within the dominant ecosystem state corresponding to an average 'upwelling dominant scenario' without extreme El Niño events (Bertrand *et al.*, 2008b). In this context, the seasonal scale becomes the scale of greatest environmental variability.

An important contribution of our work concerns the Fishermen compartment. Based on statistics computed from an unprecedented rich dataset (spatially-explicit data from $\sim$ 90000 fishing trips), we showed that Fishermen spatial behavior can be seen as a function of anchovy biomass and distribution, and environmental conditions. Fishermen-Anchovy association (0.63) gave evidence that, despite the available technology and all the economic issues involved, uncertainty on prey location is a major driver of fishermen spatial behavior. That confirms that, as a consequence, fishermen deploy spatial foraging strategies comparable to that of other

animal predators (Bertrand *et al.*, 2007). Fishermen-Environment association was not at all despicable (0.56), showing that fishermen behavior is sensitive to changes in environmental conditions, most likely through transfer via Anchovy (RV = 0.61 for Environment-Anchovy).

The correlations coefficients between the first principal components had the same ranks of associations than the RV coefficients, but stronger (75%, 60% and 71% for Environment, Anchovy and Fishermen, respectively). Looking closely at the trends of those first principal components (Fig. 6.5), the strong seasonality observed for Environment (seasonal cycle component) is less pronounced for Anchovy (fish biomass and proximity) and almost not noticeable for Fishermen (activity budget). Conversely, the marked increasing trend observed for Fishermen, is blurred for Anchovy and then for Environment by the seasonal variability. Our results constitute evidence of high-frequency environmental variations being smoothed out through trophic levels (Di Lorenzo and Ohman, 2013; Mann and Lazier, 2006).

How environmental fluctuations influence population dynamics and animal behavior are questions that have been examined in several studies. McManus and Woodson (2012) showed that the more organism size and motility increase, the weaker are the relations between organism behavior and physical processes. Other studies (e.g. Greenman and Benton, 2003; Miramontes and Rohani, 1998; Petchey, 2000) showed transfer of environmental variability in ex situ and theoretical populations. Rouyer *et al.* (2012) showed, using empirical data, that mortality of larger fish lead to more prevalent short-term fluctuations in fish populations. However, to our knowledge, no work had shown empirical evidence of smoothed out transfer through several ecosystem compartments, from the environment to top predators. Here we provide – for the first time – empirical evidence of (1) bottom-up and smoothed out transfer of the high-frequency environmental fluctuations through the main compartments of an ecosystem (i.e. seasonal effect through Environment-Anchovy-Fishermen); and (2) magnification of low-frequency fluctuations by trophic transfer (i.e. increasing trend through Fishermen-Anchovy-Environment). Furthermore, shift occurred in 2005/11-12 towards a weaker variability but a positive trend in Environment and Anchovy PCA scores (Fig. 6.5). Such shift was also observed in Fishermen that presented more favorable activity budgets (higher scores) from 2005/11-12. For each compartment, variability – computed as the sum of squared residuals from a fitted linear regression – was higher before the shift (2005/11-12) than after (Table 6.6). Moreover, variability measures before the shift decreased

when ecological levels rose.

Table 6.6: Sum of squared residuals

Compartment	From 2000/06 – 07 to 2004/11 – 12	From 2005/11 – 12 to 2009/02 – 04
Environment	10.88	1.24
Anchovy	6.77	2.26
Fishermen	2.58	1.31

6.4.2 Fishermen response to ecosystem scenarios

Based on the statistical analyses, we constructed four ecosystem scenarios (Fig. 6.9). Each scenario belonged to a distinct quadrant in the coinertia space (Fig. 6.8, right panel). The first two scenarios correspond to overall favorable ecosystem conditions and are associated with the time-periods 2002/02-03 and 2008/02-04 while the third and forth are less favorable and can be illustrated by the time-periods 2003/10-12 and 2001/10-11. Here, we analyze and discuss those scenarios taking into account that changes in biomass of organisms are not always simultaneous to changes in environmental conditions (if bad environmental conditions can trigger mortalities and translate into low biomasses almost instantaneously, good environmental conditions will affect recruitment, producing increased population biomass with delay); instead, patterns of geographic distribution or organism movements may adjust rapidly to physical forcing (Bertrand *et al.*, 2008b).

The favorable scenarios

Scenarios I and II (Fig. 6.9, upper panels), illustrated by the time-periods 2002/02-03 and 2008/02-04, were characterized by high SST (relatively to the range of temperatures encompassed in our dataset), high primary production and a shallow oxycline. Under both scenarios, anchovy was abundant and close to the coast. High SST and CHL values characterized productive summertime conditions (Echevin *et al.*, 2008; Gutiérrez *et al.*, 2011) with cold coastal waters close to the coast that favored locally-concentrated anchovy biomass (Gutiérrez *et al.*, 2007; Swartzman *et al.*, 2008). Evidence of positive relation between anchovy acoustic biomass and oxycline depth was also shown by Bertrand *et al.* (2011). They suggested that anchovy is capable of surviving in areas where the oxygen minimum zone is very shallow, where they escape from predation by larger fish that need more oxygen. The two

scenarios described above represent favorable conditions for fishermen because fish is easily catchable. Overall, fishermen had an efficient activity budget under both conditions: more time was spent fishing rather than cruising and vessels did not need to go far from the coast. Nevertheless, in 2008/02-04 (Fig. 6.9, right-upper panel), local biomass was higher, and highly concentrated near the coast. Compared to 2002/02-03, this scenario was even more favorable for fishermen not only in terms of activity budget but also in movement geometry. Strong patchiness of anchovy led to more diffusive spatial behavior of fishermen (Bertrand *et al.*, 2007, 2005). Since anchovy patches were very close to the coast, fishermen did not need to go far to fish, so the average length of the moves was very small. Because fish was very close and highly patchy, only a small proportion of time was spent cruising; most of the time they were searching and fishing (Fig. 6.3). That translated into movements composed of a large proportion of small moves (searching behavior within fish aggregation) and a small proportion of large moves (straight transit between port and anchovy aggregation), creating a super diffusive, Lévy-like movement.

Conversely, in 2002/02-03 (Fig. 6.9, left-upper panel) anchovy was spread over more space and exhibited a lower local biomass than in 2008/02-04. That meant that, in spite of the overall availability of anchovy, it was not very patchy so it pushed fishermen to visit more patches to fill the vessel hold. Thus, the relationship rare-large-moves vs. numerous-small-moves was not observed; instead moves were about the same length, producing a normally diffusive movement. Nonetheless, fish was close to the coast, so the average length of the moves was small.

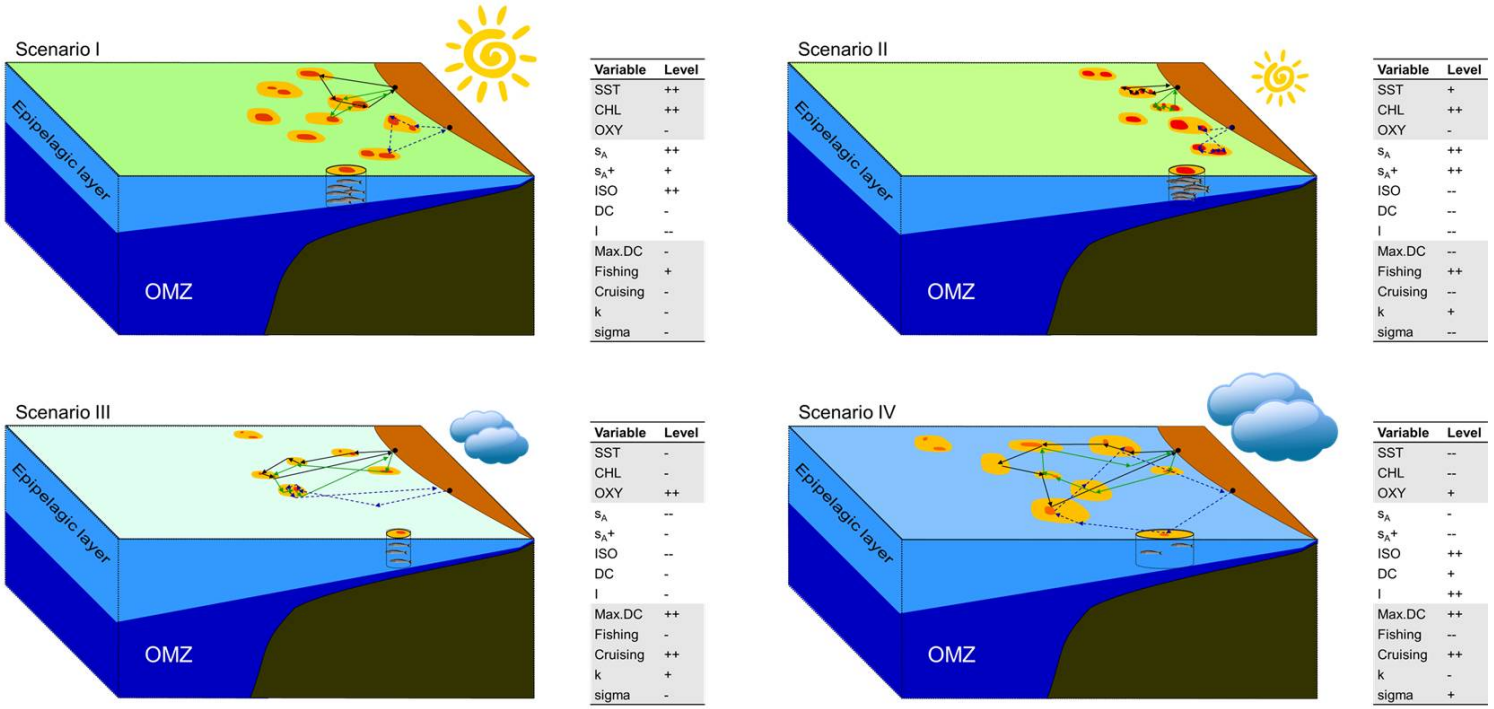
The unfavorable scenarios

Scenarios III and IV (Fig. 6.9, lower panels), illustrated by the time-periods 2003/10-12 and 2001/10-11, were characterized by low SST (relatively to the range of temperatures encompassed by our dataset), low primary production and deep oxycline. In both scenarios, anchovy was scarce. Low SST and CHL characterized winter conditions (Echevin *et al.*, 2008; Gutiérrez *et al.*, 2011), associated to low local biomass (Gutiérrez *et al.*, 2007).

In 2003/10-12, the extent of cold coastal upwelling waters was very limited, and so was anchovy habitat (Bertrand *et al.*, 2004a; Swartzman *et al.*, 2008), explaining the lowest ISO of the series. Although anchovy was not too far from the coast, it was very scarce (the lowest s_A value of the series), concentrated in very few spots

and most likely deep (since oxycline was remarkably deep). This corresponded to an adverse scenario for fishermen. For starters, fishermen went far from the coast and made fishing trips that lasted longer (Fig. 6.3) due to the scarceness of anchovy and the difficulty to find them. They spent most of the time cruising, and low time searching and fishing. Since the scarce anchovy was concentrated (low ISO), this led to super diffusive movements (less diffusive than in Scenario II). Fishermen made few large moves – 'large' compared to the 'small' ones – looking for a zone they presumed to be of high prey density, and many small moves within these zones trying to find attractive densities of prey to fish. Since fishermen went far from the coast looking for fish, the average length of moves (large and small all together) was higher than in Scenario II.

2001/10-11 did not stand out from the group of time-periods of quadrant IV in the coinertia space (Fig. 6.8, right panel). However, it did stand out in the univariate analyses, so we chose it for characterizing a forth scenario. In 2001/10-11, the extent of cold coastal upwelling waters was high, and so was the area occupied by anchovy (Bertrand *et al.*, 2004a; Swartzman *et al.*, 2008). Anchovy distribution was even but extended very far from the coast in the form of spare aggregations. This was a difficult scenario for fishermen too, mostly expressed through the high variability in their behavior (the highest variability in most Fishermen variables; Fig. 6.3 and 6.10). In general, fishermen went farther from the coast looking for prey. Because scarce anchovy was distributed in many ESDUs, close and far from the coast, fishermen seemed to use normally diffusive movements, moving constantly between aggregations of dispersed anchovy. That spatial behavior produces inefficient activity budgets, since most of the time at sea was spent cruising between aggregations and going far from the coast, rather than fishing/searching. Moreover and as a consequence, the average length of the moves was larger.



Variable	Level
SST	++
CHL	++
OXY	-
s_A	++
s_A+	+
ISO	++
DC	-
I	--
Max.DC	-
Fishing	+
Cruising	-
k	-
sigma	-

Variable	Level
SST	+
CHL	++
OXY	-
s_A	++
s_A+	++
ISO	--
DC	--
I	--
Max.DC	--
Fishing	++
Cruising	--
k	+
sigma	--

Variable	Level
SST	-
CHL	-
OXY	++
s_A	--
s_A+	-
ISO	--
DC	-
I	-
Max.DC	++
Fishing	-
Cruising	++
k	+
sigma	-

Variable	Level
SST	--
CHL	--
OXY	+
s_A	-
s_A+	--
ISO	++
DC	+
I	++
Max.DC	++
Fishing	--
Cruising	++
k	-
sigma	+

Figure 6.9: Idealized 3D representation of the ecological conditions in scenarios I (2002/02-03), II (2008/02-04), III (2003/10-12) and IV (2001/10-11). The sun or clouds and their size represent the amount of SST which for this study. Tones of blue-green color represent the amount of chlorophyll-a. Spots in dark yellow and red represent anchovy aggregations; the level of local aggregation is proportional to the darkness of the red color. Local aggregation is also represented by the amount of fish in the school. Arrows represent fishermen movements; each color is associated to a fishing trip. At the right, a table indicates the level of each variable for the given scenario. Levels can take values of: –, -, + and ++ and are associated to the values taken by each variable in Figures 6.2 and 6.3.

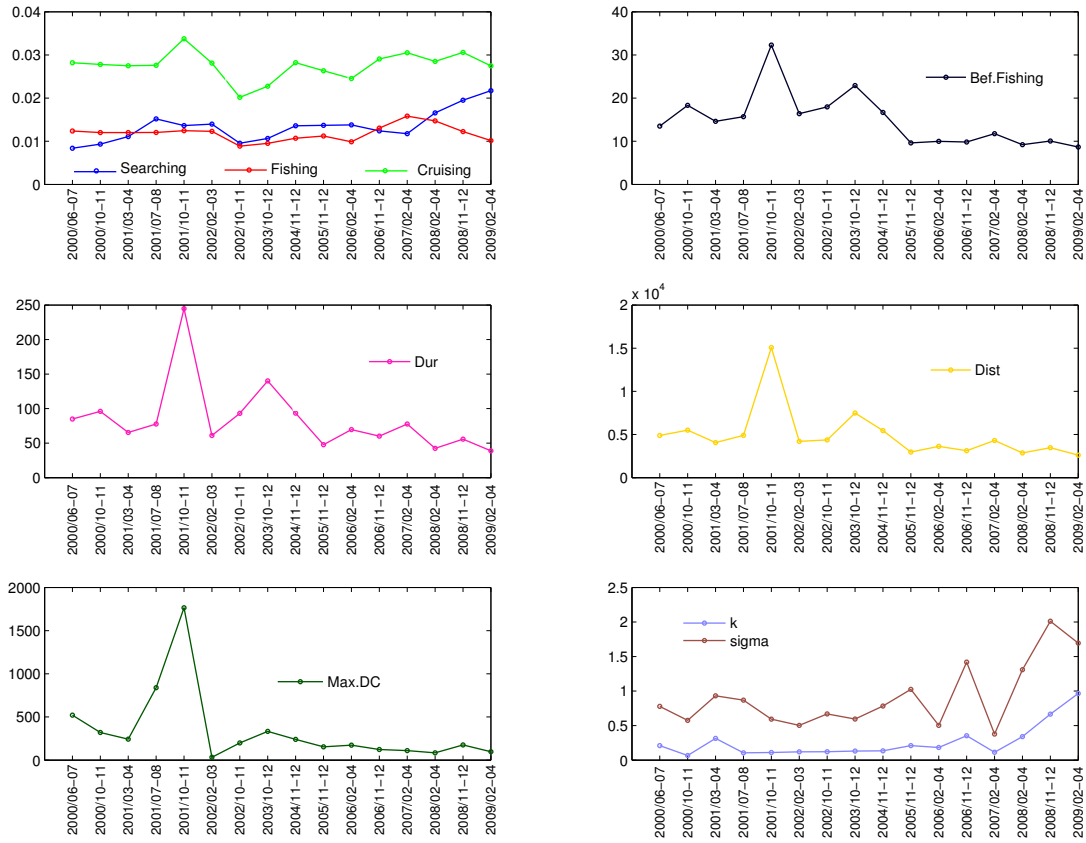


Figure 6.10: Series of variances of Fishermen variables.

It is important to notice that both unfavorable scenarios presented the greatest variability in fishermen spatial behavior among all the studied periods (greater in 2001/10-11 than in 2003/10-12; Fig. 6.3 and 6.10). It may evidence that in scenarios of high accessibility to prey, fishermen act more homogeneously, as it is easy for everyone to find fish and to catch it. But when conditions are tricky, each fisherman manages risk differently, and personal strategies count more. This type of behavior is likely to be enhanced in El Niño/La Niña conditions. It would be appealing to explore those scenarios.

6.4.3 Fishermen behavior and ecosystem approach to fisheries

This work provides a better understanding of how fishermen behavior is shaped by changes in environment and prey conditions. We call it spatio-temporal or spatial behavior because it is characterized by spatial and temporal descriptors of fishing trips (e.g. distance traveled, maximum distance to the coast, duration of the fishing

trip, time spent cruising, fishing, etc.).

As stated by Boyd *et al.* (2004), we intuitively think in terms of space and time dimensions partly because these are the easiest to measure but there are many other dimensions that characterize an animal's – or in this case human's – state. Some other variables may explain an important part of the behavior variance, such as the vessel holding capacity, the personality of the captain, oil and fish meal prices, technology improvements, government management measures for specific seasons, company strategies and tactics (i.e. how many vessels to deploy each day and in which zones, at which factories to land the fish), the degree of competition or co-operation between fishing vessels. Taking into account these other variables would improve our understanding of fishermen behavior, which is under the joint influence of economics, politics, oceanic and biological conditions (Wilen, 2004). A remaining challenge is the characterization of the collective behavior, since we always consider all fishing trips as independent from the others (the same approach is usually taken in animal behavior; Boyd *et al.*, 2004). Despite we did not considered all these potential factors, we have shown that prey availability and environmental conditions already play a great and significant role in shaping fishermen behavior, at least at the scale of average characteristics of fishing trips during two or three months periods.

In synthesis, in this study we showed how ecosystem scenarios shape fishermen spatial behavior. We also provided empirical evidence of bottom-up and smoothed out transfer of the high-frequency environmental fluctuations and magnification of low-frequency fluctuations by trophic transfer through the ecosystem compartments. Top predators are increasingly seen as integrators of the state of the ecosystem they inhabit (Boyd *et al.*, 2006). In the need for information to support the ecosystem approach to fisheries (EAF; Browman and Stergiou, 2004; Garcia and Cochrane, 2005; Jennings, 2005), top predators could be used as indicators of the ecosystem they inhabit. In the bottom-up dynamics we described, Fishermen PCA scores could be considered for use as ecosystem indicators of ecosystem scenarios. However, to test for the robustness of such indicators it is necessary to rely on larger time series encompassing strong El Niño/La Niña events, in order to cope with scales of variability in the NHCS that have not been included in the studied period. Implementation of an EAF management would also require understanding the interactions between fishermen and other top-predators competing for the same prey (e.g. seabirds and mammals; Bertrand *et al.*, 2012). Indeed, understanding the dynamics between

fishermen, other predators and ecosystem scenarios is crucial for incorporating conservation plans into the EAF management (Hooker *et al.*, 2011). Such aim can be pursued if data on seabirds and mammals are available via bio-logging devices, which is the case in Peru (e.g. Bertrand *et al.*, 2012). All these reasons highlight the importance of spatio-temporal data on top predators, including fishermen, for fisheries ecology and sustainable management of the fisheries (Hinz *et al.*, 2013).

Chapter 7

Fishermen spatial behavior and fish acoustic biomass: two sides of the same coin?

"The observer imposes a perceptual bias, a filter through which the system is viewed. (..) The challenge is the familiar one of trying to understand patterns observed at one level of detail in terms of mechanisms that are operating on other scales."

– Simon Levin (*The Problem of Pattern and Scale in Ecology*)

7.1 Introduction

Prey spatial structure and dynamics condition predator's spatial behavior (e.g., Fauchald *et al.*, 2000; Rose and Leggett, 1990). Indeed, predator's search for prey could be regarded as a search for spatial co-occurrence with its prey. Therefore, the study of foraging behavior contributes to understanding prey spatial distribution and dynamics. One of the most influential models in the literature for studying spatial allocation of foragers is known as the ideal free distribution (IFD; Fretwell and Lucas, 1969). When resources are distributed among a number of distinct areas, and when foragers must choose among those areas, the IFD in its simplest form predicts that the proportion of predators present in each aggregation is proportional to the available resource; this is based upon several assumptions (Bernstein *et al.*, 1991): (i) individual foragers are 'ideal', i.e., they seek to and are able to choose the area that maximizes fitness rewards; (ii) they are 'free', i.e., they experience no cost of moving between feeding areas; (iii) they are 'equal', genetically and otherwise; and (iv) predation does not affect the aggregations and so their is a permanent regeneration of the resources, i.e., the prey population.

The IFD theory could be a good starting point for characterizing fishermen spatial behavior in relation to fish; it has been applied in several works in fisheries science (Gillis, 2003; Gillis *et al.*, 1993; Rijnsdorp *et al.*, 2000; Swain and Wade, 2003). Nonetheless, IFD hypotheses may be far from reality, particularly in pelagic fisheries. In pelagic ecosystems, water masses and fish schools are constantly moving (Bertrand *et al.*, 2008a; Peraltilla and Bertrand, 2013; Swartzman *et al.*, 2008), so fishermen face great uncertainty on prey localization; i.e., they are not ideal foragers. Moreover, fuel limitations and, in some fisheries, the race for fish (e.g., Aranda, 2009), make the 'free' hypothesis unrealistic. Fishing vessels are neither 'equal'. Differences in spatial behavior and effort may be due to factors such as vessel sizes, technology used and captain's skills (e.g., Aranda (2009); Gaertner *et al.* (1999); Rijnsdorp *et al.* (2000); Vázquez-Rowe and Tyedmers (2013) and Chapter 5). Concerning the regeneration of prey aggregation, it has been shown that intensive fishing can cause local depletion (e.g., Bertrand *et al.*, 2012). For those reasons, some works have used modified versions of the IFD; for example, Poos *et al.* (2010b); Poos and Rijnsdorp (2007) incorporated interference competition and spatial segregation into IFD-based modeling. However, other works strongly reject the IFD theory for foraging movement (e.g., Pierce and Ollason, 1987). It thus seems imperative to assess to what extent fishermen spatial behavior can be used as indicator of fish spatial distribution.

Spatially explicit information on fishermen behavior can be obtained from Vessel Monitoring System (VMS) data. Efficient methods have been developed for obtaining high resolution spatio-temporal information on fishermen activities or behavioral modes (e.g., searching, fishing, cruising) from VMS positioning data (see Bertrand (2013) and Chapter 2 for a review). For example, in Walker and Bez (2010), hidden Markov models were used for inferring fishermen activities (i.e., cruising, tracking, stop and fishing) from VMS data on tuna purse-seiners in the Indian Ocean. Then, the probability of tune presence over a gridded map was inferred for each month from the time spent at each activity, using non-linear geostatistics Walker *et al.* (submitted). Due to the high migratory behavior of tuna, their habitat is too wide to prospect through scientific surveys. Thus, the maps of probability of presence elaborated could not be confronted and validated with other sources of data on tuna distribution.

Other pelagic fisheries do perform scientific acoustic surveys in a regular basis

for inferring spatial distribution and abundance of fish (Simmonds and MacLennan, 2005). An index of fish biomass is integrated for each georeferenced elementary distance sampling unit (EDSU) of 1 nm along the survey track. Then, those transect-based data are interpolated for the whole area of interest, typically based on geostatistical modeling.

Studying VMS and acoustic data could be like looking at fish through different glasses: fishermen and scientists perspectives. For inferring fish distribution from fishermen spatial behavior, the sampling method is rather preferential since trajectories are directed towards where fishermen expect to find large fish patches. For the scientific surveys, the sampling design is systematic (Fig. 7.1). Another important difference lies in their spatio-temporal scales. VMSs often provide positioning records at a frequency of $\sim$ 1 per hour (Bertrand, 2013). Because a fishing fleet spatially distributes throughout the whole fishing zone, analyzing the VMS records of an entire fleet means being able to cover, in general, the extent of the studied area every hour. This would provide snapshots of the vessels distribution – and thus behavior and effort – almost continuously (like photographs in burst mode). By contrast, acoustic data (echo responses) are collected every second, and acoustic biomass estimations are typically computed for each EDSU. Before obtaining a 'pseudo-snapshot' of the acoustic biomass and distribution, the survey track must be completed, which can take from weeks to months. Therefore, comparing the 'snapshots' obtained from both datasets is not straightforward. On the other hand, precisely because of the different nature of those data, the comparison is highly appealing.

We aim at evaluating the extent to which a VMS-based indicator of fish presence relates to acoustic biomass estimations. For that matter, we use VMS data from the Peruvian purse-seine anchovy (*Engraulis ringens*) fishery, the world's largest monospecific fishery. In this fishery, Bertrand *et al.* (2008c) applied an Artificial Neural Network (ANN) on a 2000-2002 VMS dataset covering a small zone off Peru, for inferring fishing set positions. They found significant monthly co-variations between spatial patterns (distance to the coast and clustering index) of fishing sets and fish acoustic biomass. At the scale of a fishing trip, management, fleet segmentation and skipper's personality were the main drivers found for trip patterns (Chapter 5). At the scale of a fishing season, a strong and significant – though not perfect – association was found between fishermen spatial behavior and fish biomass and distribution (Joo *et al.*, in review).

In this work, we analyze three anchovy fishing seasons in the NHCS. We use trajectory data from 11741 fishing trips, documented by 314214 VMS positioning records. Each record is associated with a given activity (i.e., fishing, searching and cruising) using a calibrated and validated hidden semi-Markov model (Joo *et al.*, 2013). We use the local density of these activities to build an indicator of anchovy presence and then interpolate this indicator for the whole fishing area off Peru using geostatistics. Next, we compare the inferred maps to those obtained based on fish acoustic biomass and we evaluate the strength of the correlation between spatial patterns of both fishermen and fish. In the light of our results, we discuss to what extent fishermen spatial effort could be used as proxy of fish spatial distribution.

7.2 Materials and methods

7.2.1 Anchovy data

Since 1983, IMARPE has been conducting on average two acoustic surveys per year for monitoring fish population distribution and biomass. These surveys consist of parallel cross-shore transects of $\sim$ 100 nm long, with a $\sim$ 15 nm spacing (Fig. 7.1). Simrad (kongsberg Maritime AS, Norway) scientific echosounders working at distinct frequencies are used to estimate biomasses (see Castillo *et al.*, 2009; Gutiérrez *et al.*, 2007; Simmonds *et al.*, 2009). An extensive midwater-trawl sampling completes the acoustic surveys for species identification. The nautical-area-backscattering coefficient (NASC or s_A, in $m^2.mn^{-2}$), an index of fish biomass (Simmonds and MacLennan, 2005), is recorded in each georeferenced elementary distance sampling unit (EDSU) of 1 nm. NASC is then log-transformed and, for purposes of this study, scaled from 0 to 1 for facilitating the comparison with a proxy of probability of anchovy presence. This scaled log-transformed NASC is henceforth called acoustic biomass. For this study, we selected data from three acoustic surveys corresponding to the following time-periods: (i) March-April, 2001; (ii) November-December, 2008; and (iii) December, 2009.

7.2.2 Fishermen data

Satellite tracking by VMS is mandatory for the whole Peruvian industrial fishing fleet since 2000. Vessel positions ($\pm$100m of accuracy; $\sim$ 1 record per hour) for

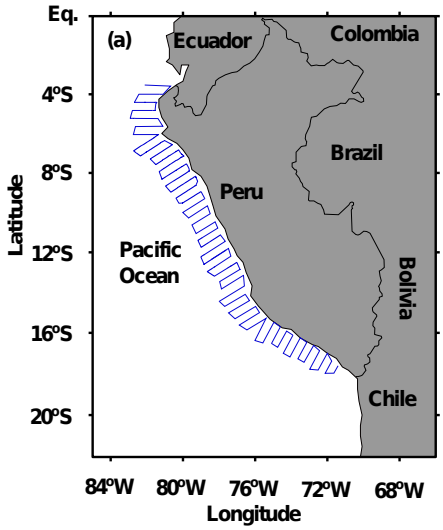


Figure 7.1: Typical survey design of an acoustic survey off Peru (blue line).

tens of thousands of fishing trips are thus available for scientific purposes since (e.g. Bertrand *et al.*, 2007, 2005, 2008b,c, 2012; Joo *et al.*, 2011, 2013). We report in Table 7.1 the number of fishing trips documented for each time-period. Pre-processing of VMS data are performed based on the criteria and algorithms described in Bertrand *et al.* (2007, 2005); Joo *et al.* (2011) and are detailed in appendix A.

Another source of information regarding fishing activities is IMARPE's program of on-board observers. For $\sim$ 1% of the fishing trips, on-board observers record the location and time of three main activities: fishing, searching and cruising (i.e. traveling following a predetermined course). In order to infer the activities for the remaining 99% of the fishing trips from VMS data only, a supervised hidden semi-Markov model was trained and validated using the on-board observer dataset (Joo *et al.*, 2013). This model reached a mean accuracy of 80% in the determination of the correct activities from the VMS data.

A synoptic sketch of the methodology from this chapter is shown in Figure 7.2. We will now describe each one of those analyses.

7.2.3 Composite maps

Because of the highly dynamic movement of fishing vessels and the mismatches in time-resolution and space-coverage between acoustic biomass and VMS data, instead of using a map aggregating all the behavioral modes corresponding to the period of time of the acoustic survey, a composite map is computed for each time-period.

Table 7.1: Number of fishing trips and VMS records corresponding to each time-period.

Time-period	Fishing trips	VMS records*	Fishing records	Searching records	Cruising records
2001 03/12 to 04/10	3795	82795	29416	22362	31017
2008 11/23 to 12/12	5063	101496	35166	37875	28455
2009 12/08 to 12/29	2883	63262	20185	23136	19941

*Number of VMS records after removing the ones from the last cruising segment from every fishing trip.

Note:

It consists in dividing each time-period (Table 7.1) into groups of $\sim$ 5 days; then, in separating the off Peru region according to the acoustic track: the latitudinal frontiers for 5-days periods are recovered for applying them in the behavioral mode maps. That way, each 5 days of acoustic biomass will be compared with concomitant 5 days of VMS-based data (Fig. 7.9).

7.2.4 Proxy of anchovy presence probability

Some basic assumptions are made a priori in order to construct this proxy. First, we assume that fishermen decide to fish due to a direct observation of a fish aggregation. Second, we assume that when a fishing vessel is in a searching mode, there is a non negligible probability of fish presence. Third, we assume that when a fishing vessel is cruising, no fish has been observed. These three assumptions were also adopted in Walker (2010). As a fourth assumption, we suppose that the last cruising segment in a fishing trip (after the last searching or fishing activity), fishermen have decided to go straight to the arrival port or factory, regardless of the amount of fish in the way. Thus, the last cruising segment of each trip is considered non informative and removed from the dataset. The number of records corresponding to each time-period is reported in Table 7.1.

The fishing area, from 3°S to 19°S and from the coastline (at $\sim$ 70°W or west) to 84°W, is divided into grid cells of 5 km $\times$ 5 km. The number of fishing, searching and cruising activities in each cell is computed. Then, a proxy of probability of prey presence P, henceforth called presence proxy, is computed for each cell:

$$P = \frac{(F \times 1) + (S \times W_S) + (C \times 0)}{F + S + (C \times W_C)} = \frac{F + (S \times W_S)}{F + S + (C \times W_C)} \tag{7.1}$$

where F, S and C are the number of VMS records corresponding to fishing, searching and cruising activities in the cell, respectively. $W_S \in [0, 1]$ is the weight of the searching activities for the presence proxy, and $W_C \in [0, 1]$ is the weight of the cruising activities.

The weights associated with fishing, searching and cruising activities in the numerator correspond to assumptions 1, 2 and 3. If $W_C = 1$, the denominator of P is the number of VMS records in the cell, and P is the weighted density of activities in the cell. However, we do not know to which extent cruising data is actually informative. In other words, it may be possible that sometimes fish is not present or in very small amounts when the vessel is cruising; it may also be possible that the fishermen are just not looking at the echosounders while they are going to a zone where they presume they will find fish. Because of this incertitude in the usefulness of the cruising information (i.e., in the third assumption), we attach a weight W_C to it. Here, we test three values for W_S (0.3, 0.5 and 0.7) and three values for W_C (0, 0.5 and 1). Maps resulting from each combination of weight values are compared.

7.2.5 Spatial interpolation

Geostatistical interpolations are used for both the acoustic biomass and presence proxy data. Geostatistics are recognized to be particularly suitable for describing spatial distributions of marine populations (Ciannelli *et al.*, 2008; Petitgas, 2001). In general, they provide tools for capturing and modeling the spatial variability of a given variable distributed in space or simultaneously in time and space (Chilès and Delfiner, 2012). Kriging, the geostatical interpolation technique, requires the use of a theoretical variogram model, which models the increase of variance between two points, apart from each other at a distance h, when h increases. Several variogram models exist in the literature (Chilès and Delfiner, 2012). Among them, the flexible Matérn model is recommended due to the inclusion of a smoothness parameter that allows describing the spatial process at small lags (Marchant and Lark, 2007; Pardo-Iguzquiza and Chica-Olmo, 2008). Exponential and Gaussian models are particular cases of the Matérn model. Here we use the Matérn model with geometrical anisotropy (directional dependence) to account for lower (vs. greater) correlation orthogonal (vs. parallel) to the coast. The parameters of the model are estimated

using ordinary least squares. For exploring the spatial structure of the data and for choosing the initial values of the parameters for the estimation, empirical variograms are computed for both acoustic biomass and presence proxy, for each time-period.

Given the estimated variogram model, ordinary kriging is used for interpolating both the acoustic biomass and the presence proxy for a region encompassing the anchovy habitat and the fishing zones (from 3°S to 19°S and from the coastline to 84°W) for grids of 2 km. The 2 km grids allow interpolation on a finer scale than the 5 km grids for which P is computed. The gstat package (Pebesma, 2004) in R (R Core Team, 2013) is used for fitting the Matérn model and for ordinary kriging.

7.2.6 VMS-derived and acoustic maps comparison

Because of the sampling methods and time-space resolution in both processes corresponding to acoustic biomass and presence proxy, one can hardly expect direct map-overlapping between them and classical linear correlation statistics appear poorly relevant. We thus established a simple criterion for comparison: to evaluate if zones with high fish biomass are associated with high presence proxy. Then, instead of comparing all grid cells from each process, we compare two levels: the highest half of both processes (grid cells taking values x, such that $x \geqslant \min + 0.5 \times [\max - \min]$) and the highest quarter of each process (grid cells taking values x, such that $x \geqslant \min + 0.25 \times [\max - \min]$). Two indicators are used for assessing to what extent fishermen spatial patterns reflect the spatial distribution of fish: distance to the coast (DC) and occupied area (A). DC is defined as the center of gravity of the distance to the coast (Joo *et al.*, in review) of the cells. It represents the longitudinal extent covered. A is defined as the number of cells associated with the level (highest half or quarter) times 4, since each cell measures $2 \times 2\ km^2$. Both indicators are measure at intervals of 1°of latitude. Then, Spearman correlation coefficients are computed for each level and indicator, for evaluating the latitudinal co-variation of the indicators for acoustic biomass and presence proxy.

We also assess at which scale the two processes, in general, co-variate. For that purpose, we fit an empirical cross-variogram to the kriged cells from both maps. A cross-variogram is an extension of the variogram for two variables, and measures, at a distance h, to which extent the increase in one of the variables corresponds on average to an increase or decrease in the other one, when h increases (Rivoirard *et al.*,

2000). For each time-period, empirical cross-variograms are computed over 500 subsampling replicas, each one of 10000 cell grids (subsampling replicas are made due to the high computational cost of computing an empirical cross-variogram accounting for all kriged grid cells). Then, the median, first and third quantiles are computed.

7.3 Results

7.3.1 Exploring fishermen activity data

The maps of fishermen activities for the five-day periods show the spatial dynamics of the anchovy fishery (Fig. 7.3, 7.4, and 7.5). For the three years, the dissimilarities in the spatial distributions of the fishermen activities through the five-day periods make that the composite maps result very different than the ones gathering data from the whole periods. These findings support the decision for making a map of presence proxy from the composite maps rather than from the whole-period maps (Fig. 7.3h, 7.4f, 7.5g).

7.3.2 Spatial interpolation

The variograms of acoustic biomass show a similar range of $\sim$ 30 km for 2008 and 2009 time-periods (Fig. 7.6). For 2001, the variogram shows an increasing trend (although it decays for a lag of $\sim$ 200 km). The $\sim$ 30 km structure is also found when computing a directional variogram shore-to-offshore (orthogonal to the coast). For the presence proxy, the spatial structure is strongly consistent over time. A first spatial range of $\sim$ 30 km is found. For 2001 and 2008, larger structures also appear.

For both datasets, kriging is done for a region encompassing the anchovy habitat and the fishing zones (from 3°S to 19°S and from the coastline to 84°W). From the whole kriged region, we select a smaller region based on the kriging variance. For acoustic biomass, the lowest variance is around the survey track (Fig. 7.7). Therefore, a polygon around the track is selected. For presence proxy, the lowest variances are around the cells associated with large amounts of VMS data, which in some cases, are split into several aggregations. Thus, they cannot be naturally contained in a single polygon. For that reason, instead of a polygon, a threshold on the variance is used for selection (Fig. 7.8).

7.3.3 Map comparison

For each year, we compare the acoustic biomass map with the ones issued from different presence proxies (with the different combinations of W_S and W_C values; Table 7.2). No significant positive correlations are found for 2008. For 2001, all W_C and W_S combinations give significant correlations in area and DC between highest halves of acoustic biomass and presence proxy values. For some cases with $W_C > 0$ and $W_S > 0.3$, correlation in A for the highest quarters is also significant. For 2009, when $W_C = 0$, significant correlations are found for DC, for both highest halves and highest quarters. Significant correlations in DC for highest quarters are also found for two other combinations of weights ($W_C = 0.5$ with $W_S = 0.5$, and $W_C = 1$ with $W_S = 0.7$).

Since we seek to fix a set of parameters for all years, we choose $W_C = 0$ (because significant correlations in DC are found with this value for 2001 and 2009) and $W_S = 0.5$ (because no difference is observed for $W_C = 0$ and all different values of W_S). A summary of the results with the chosen parameters is shown in Figure 7.9. Through a visual comparison, overall and at coarse scales, dense foraging spots correspond to medium and high densities of fish. However, some spots with high fish density were not even visited by the fishermen. Conversely, some spots with major fishing activity, were not detected by the acoustics as zones of high biomass.

Cross-covariograms (Fig. 7.10) evidence an overall positive co-variation between acoustic biomass and presence proxy maps. For 2001 and 2009 a maximum peak of covariation is achieved at $\sim$ 220 km, while for 2008 it was at $\sim$ 150 km (with a local minimum at 300 km). For this latter year, cross covariance decreases more than in the other two years after the peak is reached. The fact that the common structures found in the variograms are not detected in the co-variograms show that acoustic biomass and the presence proxy present structures of the same sizes but that do not co-occur.

Table 7.2: Spearman correlation coefficients for each indicator varying with latitude for all combinations of W_S and W_C. Values in bold correspond to significant correlations.

Year	W_C	W_S	Highest quarter		Highest half	
			DC	A	DC	A
2001	0	0.3	0.17	0.53	**0.70**	**0.82**
2001	0	0.5	0.17	0.53	**0.70**	**0.82**
2001	0	0.7	0.17	0.53	**0.70**	**0.82**
2001	0.5	0.3	0.23	0.45	**0.71**	**0.81**
2001	0.5	0.5	0.20	**0.62**	**0.72**	**0.77**
2001	0.5	0.7	0.22	**0.63**	**0.74**	**0.72**
2001	1	0.3	0.29	0.48	**0.71**	**0.80**
2001	1	0.5	0.30	0.57	**0.74**	**0.77**
2001	1	0.7	0.27	**0.60**	**0.74**	**0.71**
2008	0	0.3	−0.64	−0.09	−0.32	−0.50
2008	0	0.5	−0.64	−0.09	−0.32	−0.50
2008	0	0.7	−0.64	−0.09	−0.32	−0.50
2008	0.5	0.3	−0.11	**-0.77**	−0.32	−0.07
2008	0.5	0.5	−0.11	**-0.77**	−0.36	−0.14
2008	0.5	0.7	−0.11	−0.68	−0.29	−0.14
2008	1	0.3	−0.11	−0.65	−0.32	−0.07
2008	1	0.5	−0.11	−0.74	−0.36	0.00
2008	1	0.7	−0.11	−0.68	−0.36	0.00
2009	0	0.3	**0.75**	0.43	**0.72**	0.55
2009	0	0.5	**0.75**	0.44	**0.72**	0.55
2009	0	0.7	**0.75**	0.44	**0.72**	0.55
2009	0.5	0.3	0.65	0.26	0.55	0.23
2009	0.5	0.5	**0.69**	0.08	0.55	0.23
2009	0.5	0.7	0.60	−0.05	0.55	0.28
2009	1	0.3	0.65	0.26	0.55	0.30
2009	1	0.5	0.64	0.25	0.55	0.22
2009	1	0.7	**0.69**	0.10	0.55	0.27

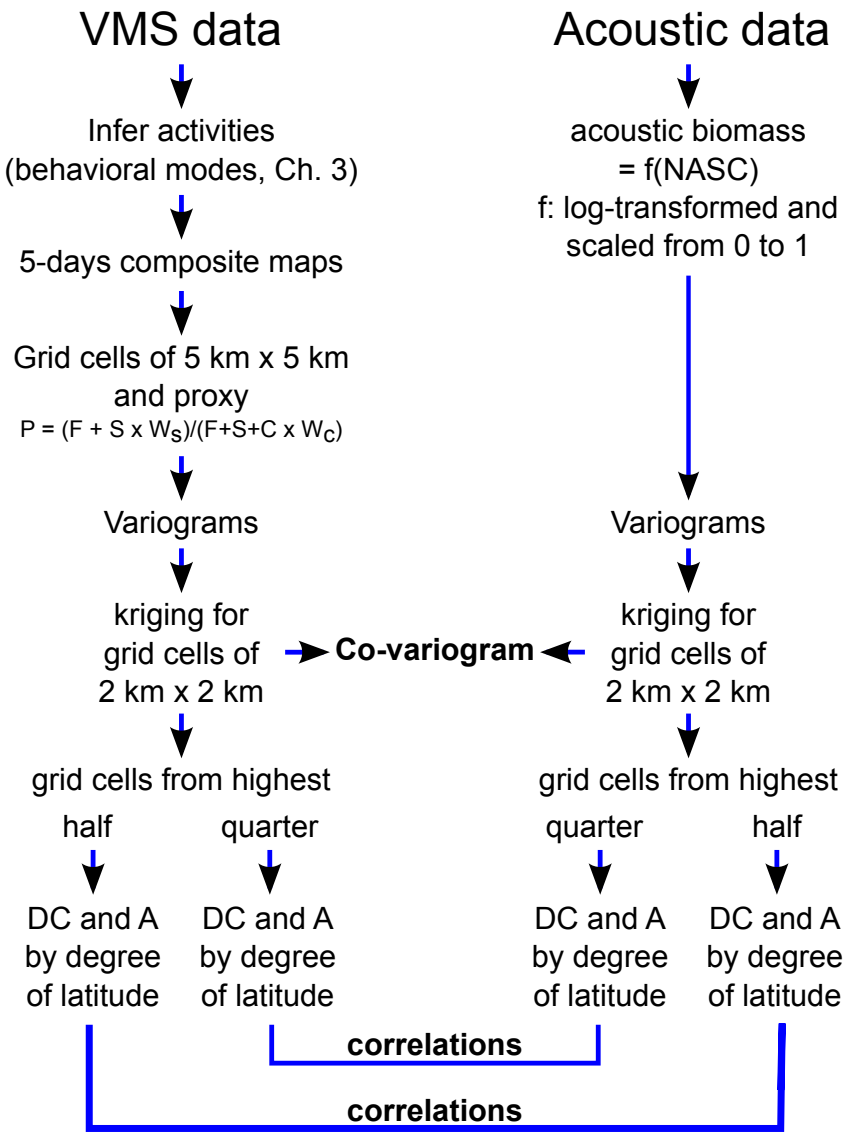
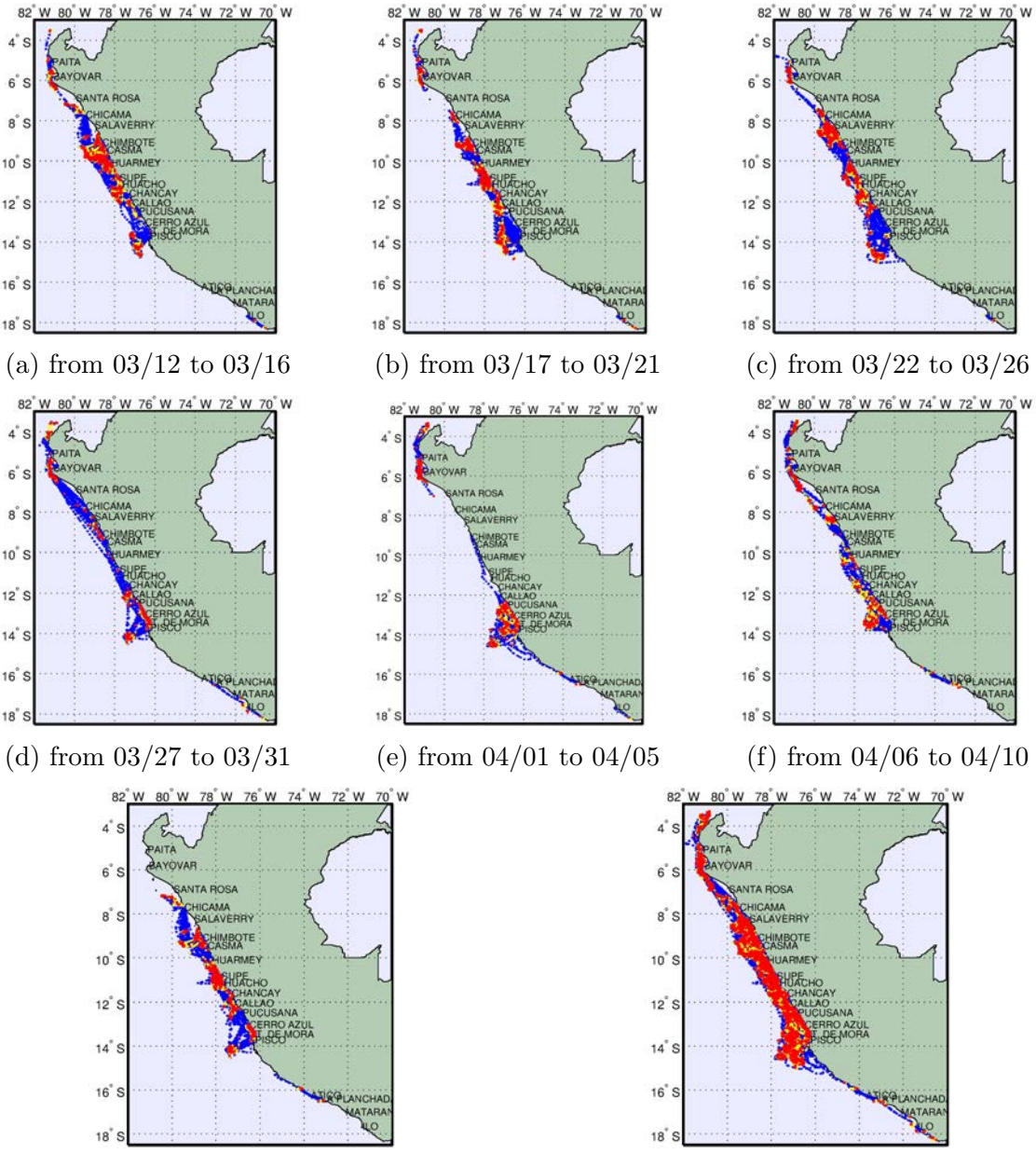


Figure 7.2: Synoptic sketch of the analyses. The elements for the final comparison are in bold.



(a) from 03/12 to 03/16

(b) from 03/17 to 03/21

(c) from 03/22 to 03/26

(d) from 03/27 to 03/31

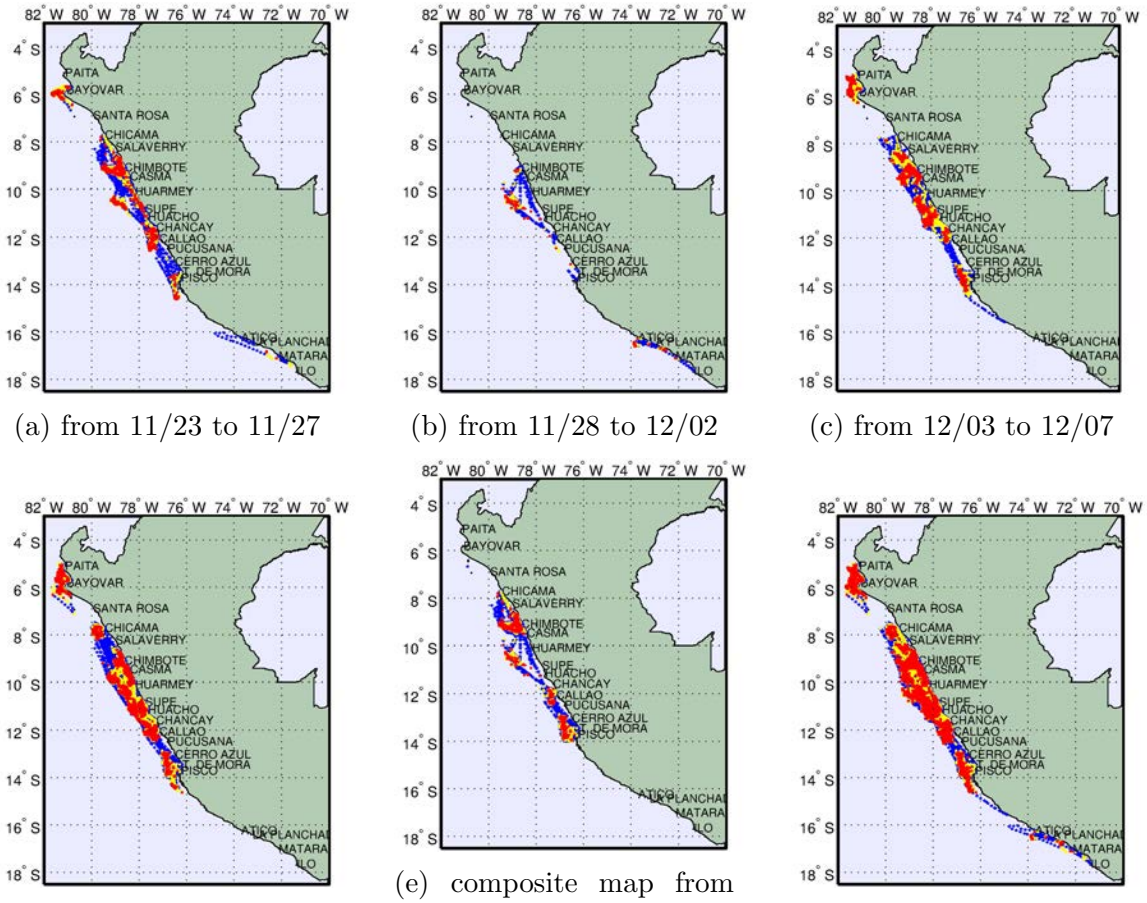
(e) from 04/01 to 04/05

(f) from 04/06 to 04/10

(g) composite map from 03/12 to 04/10

(h) from 03/12 to 04/10

Figure 7.3: Fishermen activities from 2001 associated to VMS records at different dates. For all maps, cruising records (blue dots) are plotted in the background, searching records (yellow dots) in the middle ground and fishing records in the foreground (red dots).



(a) from 11/23 to 11/27

(b) from 11/28 to 12/02

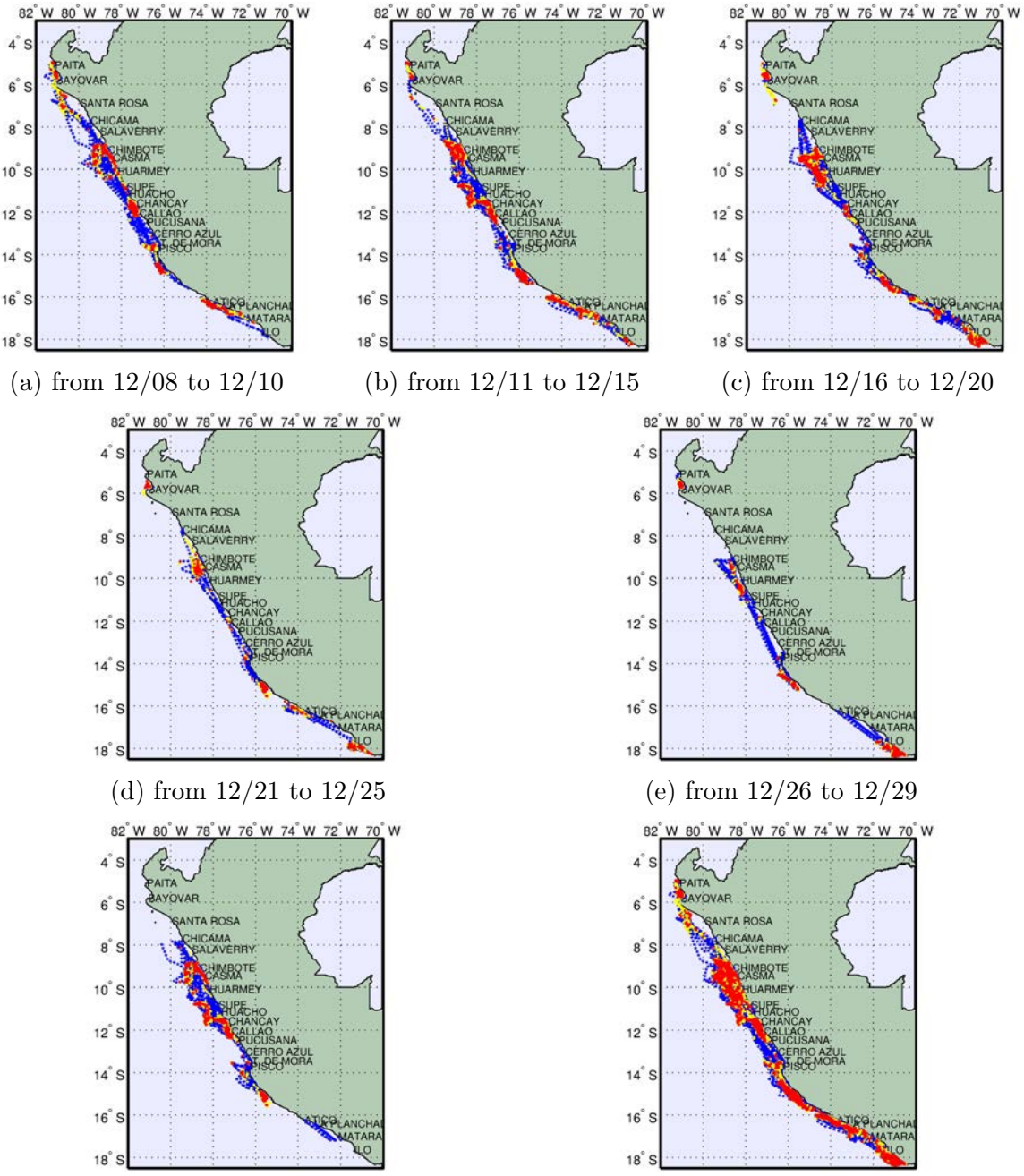
(c) from 12/03 to 12/07

(d) from 12/08 to 12/12

(e) composite map from 11/23 to 12/12

(f) from 11/23 to 12/12

Figure 7.4: Fishermen activities from 2008 associated to VMS records at different dates. For all maps, cruising records (blue dots) are plotted in the background, searching records (yellow dots) in the middle ground and fishing records in the foreground (red dots).

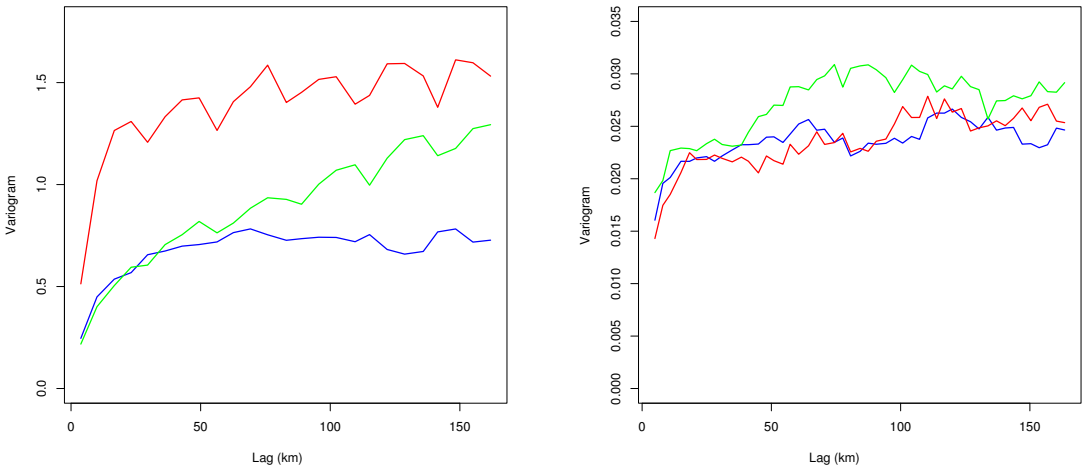


(a) from 12/08 to 12/10 (b) from 12/11 to 12/15 (c) from 12/16 to 12/20

(d) from 12/21 to 12/25 (e) from 12/26 to 12/29

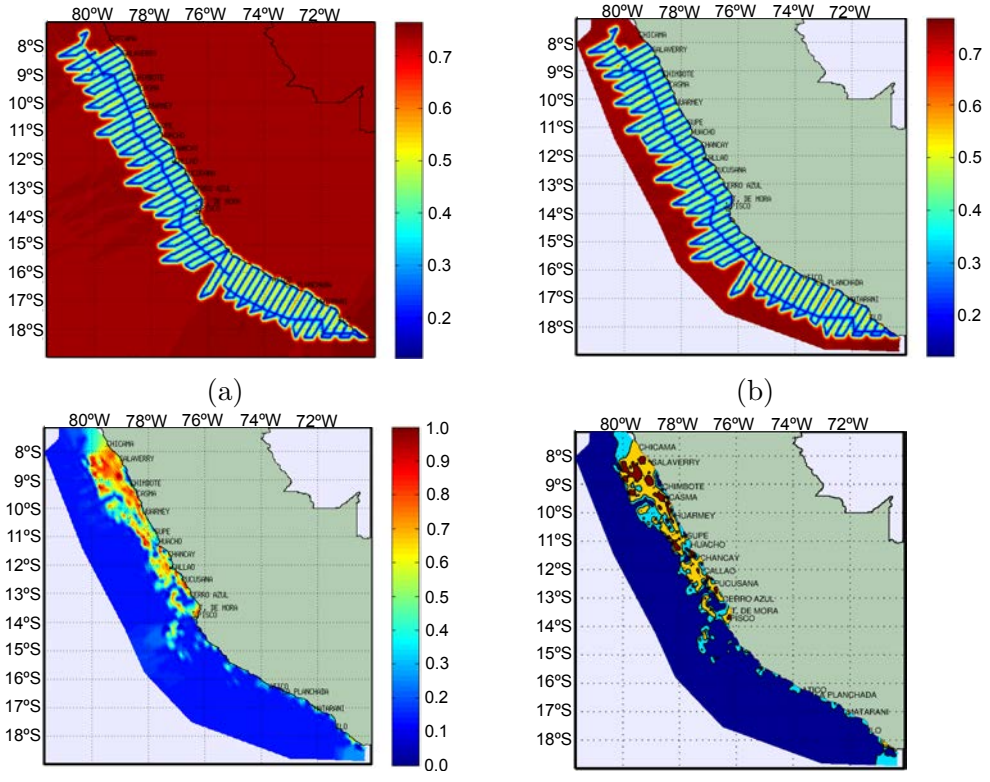
(f) composite map from 12/08 to 12/29 (g) from 12/08 to 12/29

Figure 7.5: Fishermen activities from 2009 associated to VMS records at different dates. For all maps, cruising records (blue dots) are plotted in the background, searching records (yellow dots) in the middle ground and fishing records in the foreground (red dots).



(a) Acoustic biomass (b) Presence proxy

Figure 7.6: Variograms of acoustic biomass and of presence proxy for each time-period. In each panel, the green line corresponds to the variogram of the 2001 period, the red line corresponds to the variogram of 2008, and the blue line, to variogram of 2009.



(a) (b)

(c) (d)

Figure 7.7: Krigged region for acoustic biomass (2001 example). (a) Kriged variance; (b) variance for area under the polygon selected around the lowest variance values; (c) kriged density for area under the polygon; and (d) kriged density divided in four quarters based on their values (the cells in vintage red compose the highest quarter and added to the cells in dark yellow, they compose the highest half).

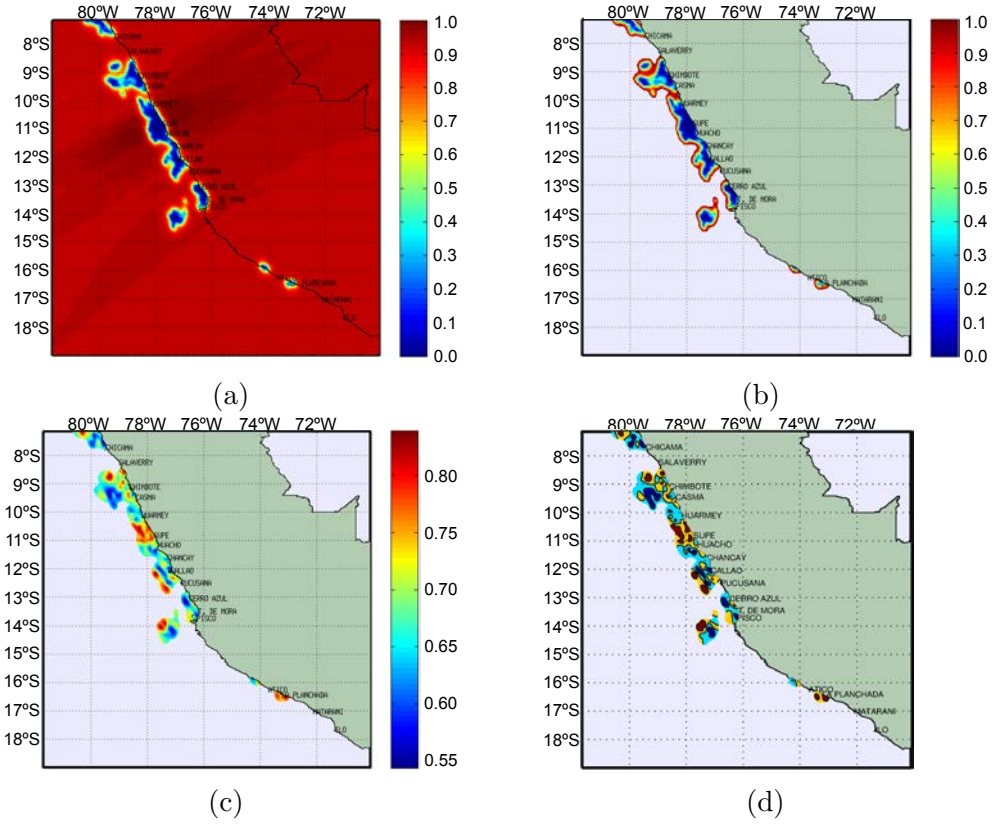


Figure 7.8: Krigged region for proxy of presence (2001 example). (a) Kriged relative variance; (b) variance for values below the threshold (0.95); (c) kriged proxy for area below the threshold; and (d) kriged proxy divided in four quarters based on their values (the cells in vintage red compose the highest quarter and added to the cells in dark yellow, they compose the highest half).

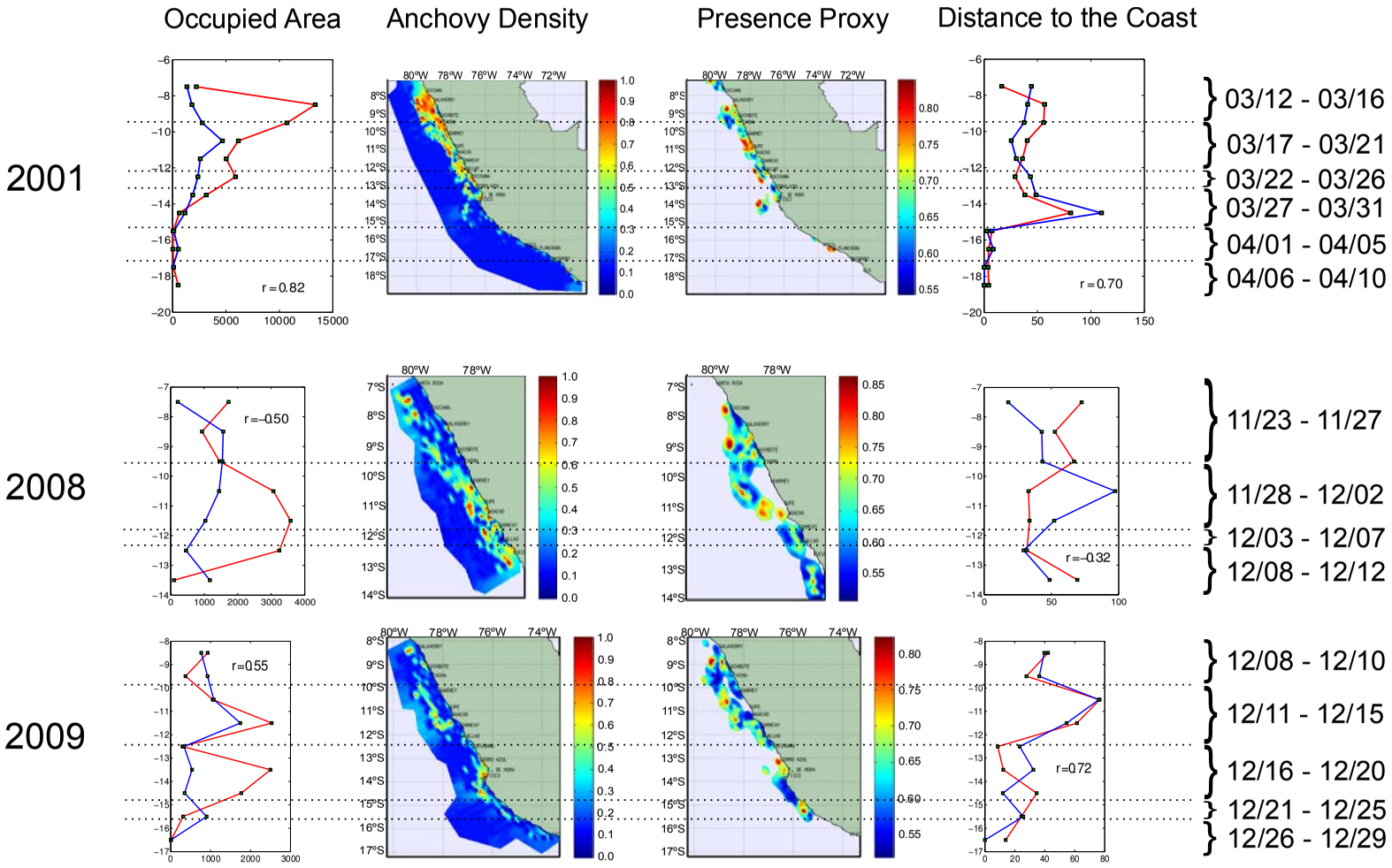


Figure 7.9: Summary of results by year for the chosen parameters. The first and last column compare the highest half of acoustic biomass and presence proxy by degree of latitude in occupied area and distance to the coast, respectively. The correlation between both, denoted by r, is also shown. In each of those plots, the blue solid line corresponds to presence proxy and the red solid line, to acoustic biomass. In the second and third columns, the kriged maps of acoustic biomass and presence proxy, respectively, are shown. Dotted lines indicate the latitude limits for each 5-day period.

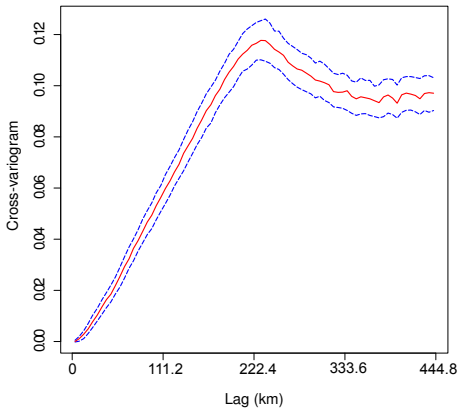
7.4 Discussion

In this study, we use VMS-based fishermen activity data to build maps of spatialized proxy of anchovy presence. We use different combinations of weights for searching and cruising activities and compare the maps with the ones obtained from acoustic surveys. Even though the underlying assumptions of the IFD theory are generally violated, significant correlations between some spatial patterns – of high density zones – are found. We show that, overall, (i) better associations are found when cruising activities are not considered in the computation of the proxy; (ii) co-variation in area and distance to the coast between acoustic biomass and presence proxy maps varies through the time-periods studied; (iii) co-variation of the interpolated presence proxy and acoustic biomass is always positive and reaches its highest value at a coarse scale for all time-periods; and (iv) acoustic biomass and presence proxy presented structures of the same sizes (of tens of km) that did not co-occur.

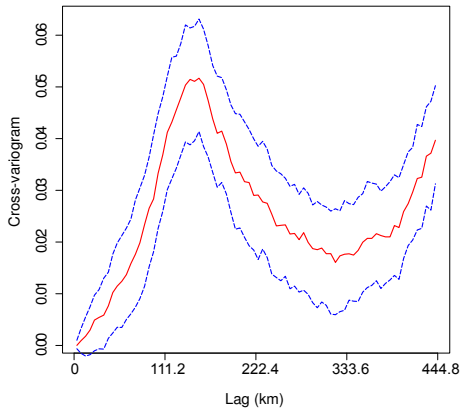
7.4.1 Implications of the results

The optimal value for the W_C parameter, 0, means that the cruising records, which represents about 30% of the data, are not reliable and add 'noise' to the analysis, so results are better when we discard the cruising records. A cruising activity means that the fishermen did not see enough fish to reduce speed and search more carefully, or start fishing. It does not always mean that there was no fish at that position, it may rather indicate that they did not see anything because they were not watching. In those cases, the cruising activity is not informative about the state of the fish and introduces noise to the model.

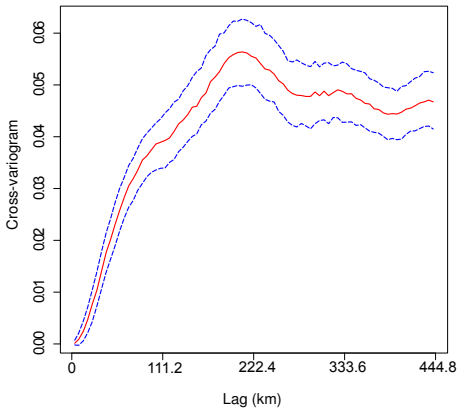
Regarding the DC and A features, results were quite variable among years. For 2001, concerning occupied area, although latitudes with high presence proxy generally correspond to high acoustic biomass, both the presence proxy and the acoustic biomass differed in their values; more specifically, the density is higher than the presence proxy (Fig. 7.9). By contrast, regarding DC, acoustic biomass and presence proxy are significantly correlated and have similar values at each degree of latitude. For 2008, no positive correlations are found for area neither for DC. However, neither of the maps (presence proxy vs. acoustic biomass) show systematically greater area or DC than the other. For 2009, DCs are significantly correlated. By contrast, areas have no significant correlation.



(a) 2001



(b) 2008



(c) 2009

Figure 7.10: Cross-variograms of acoustic biomass and presence proxy maps for each time-period. 500 subsampling replicas are done. The red solid line represents the median of the cross-variograms, while the blue dashed lines represent the first and third quartiles. The x axis represents the distance in kms, and the y axis represents the cross-variance function.

Bertrand *et al.* (2008c) found significant correlations in monthly-averaged distance to the coast between fishing set locations and acoustic biomass for a small region in Peruvian coastal waters (from 7°S to 10°S); and no significant correlations were found for occupied area. Joo *et al.* (in review) analyzed the associations between different features of anchovy (obtained from acoustic data) and fishermen spatial behavior (obtained from VMS) at a two-three months time resolution, and found a positive association between the center of gravity of the distance to the coast of fish and the maximum distance to the coast in fishing trips. It was then evidenced by both works that fishermen were highly sensitive to the longitudinal extent of anchovy, at least at coarse scales. Here, the center of gravity of the distance to the coast of fishermen and fish are computed at each degree of latitude for a one-month period. We thus show that, for two of the analyzed periods, fishermen are sensitive to the longitudinal extent of anchovy at a fine spatial scale and could be potentially considered as indicators of the longitudinal extent of anchovy. Bertrand *et al.* (2008c) postulated that the reason why there were no significant correlations in occupied area was that an indicator of occupied area is more sensitive to mismatches in spatial occupation between fishermen and acoustic biomass than a mean (or center of gravity, in our case) of distance to the coast.

Cross-variograms evidenced spatial co-variation of presence proxy and acoustic biomass at different scales, with peaks at: $\sim$ 220km for 2001, $\sim$ 150km for 2008, and two peaks, $\sim$ 100km and $\sim$ 220km for 2009. A more detailed analysis on the scales of co-variation could be done using cross-wavelets (Torrence and Compo, 1998), a data analysis technique that would allow detecting the spatial zones of co-variation at different lags (comparable to the 'h' of cross-variograms).

Although spatial covariation was found at coarse scales, both processes did not seem to co-occur. It may suggest that the dynamics of fishermen and acoustics, and more importantly, the dynamics of fishermen and anchovy may be influenced by factors operating at different scales (Hengeveld, 1987, we will describe acoustics and fishermen dynamics in the following section). Besides, the lack of perfect spatial synchrony may actually explain the sustainability of anchovy. If fishermen knew exactly where to fish and they distributed in space exactly like the fish, the spatial co-occurrence would indicate an eventual – or proximate – collapse of the fish stock. The fact that fishermen exploit only some of the fish patches, and that they do not necessarily target the biggest anchovy patches, but settle with patches large enough for fulfilling their holding capacity, contributes to sustainability.

7.4.2 Two datasets of different nature

Acoustic data dynamics

Anchovy density maps are based on acoustic survey data. Acoustic biomass information is usually assumed to provide a snapshot of fish distribution valid for the time of the survey . Nevertheless, the systematic survey design makes that, in practice, the northern transects may be about a month away from the southern transects. Since fish schools are known to move up to $\sim$ 26 km/day (Peraltilla and Bertrand, 2013), depending on the scale of the analyzed patterns, what happened at the north may not be valid anymore for the days when the survey vessel is at the south.

VMS data dynamics

The maps of fishermen activities show strong variabilities in the spatial distribution of activities for consecutive five-day periods (Fig. 7.3, 7.4, 7.5). These differences are due to dynamics at fine (e.g., hours) and coarse temporal and spatial scales (e.g., days and weeks within a fishing season; fishing zones). In a previous work on fishing dynamics, Passuni (unpublished) analyzed fishing set positions extracted from VMS data on the same anchovy fishery (inferred using artificial neural networks; Joo *et al.*, 2011). She estimated bivariant kernel densities of the fishing set positions (Botev *et al.*, 2010), detected regions with high densities and grouped them into clusters (Haralick and Shapiro, 1992; Klusch *et al.*, 2003). She used this method for the identification of clusters of fishing sets at several time scales (6, 12 and 24 hours) and found that only at a six-hour scale it was possible to track the movement of those clusters (Fig. 7.11). By contrast, at 12 and 24 hours, identified clusters were very large ($\sim$ 100km) and their movement was not easy to track visually in consecutive images. The high daily dynamics of the fishery in a small zone off Peru are also shown in Bertrand *et al.* (2012).

At coarser temporal and spatial scales, several factors contribute to variations in fishermen spatial behavior throughout a fishing season. Fishermen could have more uncertainty on the localization of the anchovy aggregations at the beginning of a season, than at the middle or the end. It could be appealing to analyze the differences in their spatial behavior throughout a whole fishing season. Fishing activities dynamism does not only varies at different time scales. It may also differ from one region to another; for instance, fishing activities may be more concentrated near the

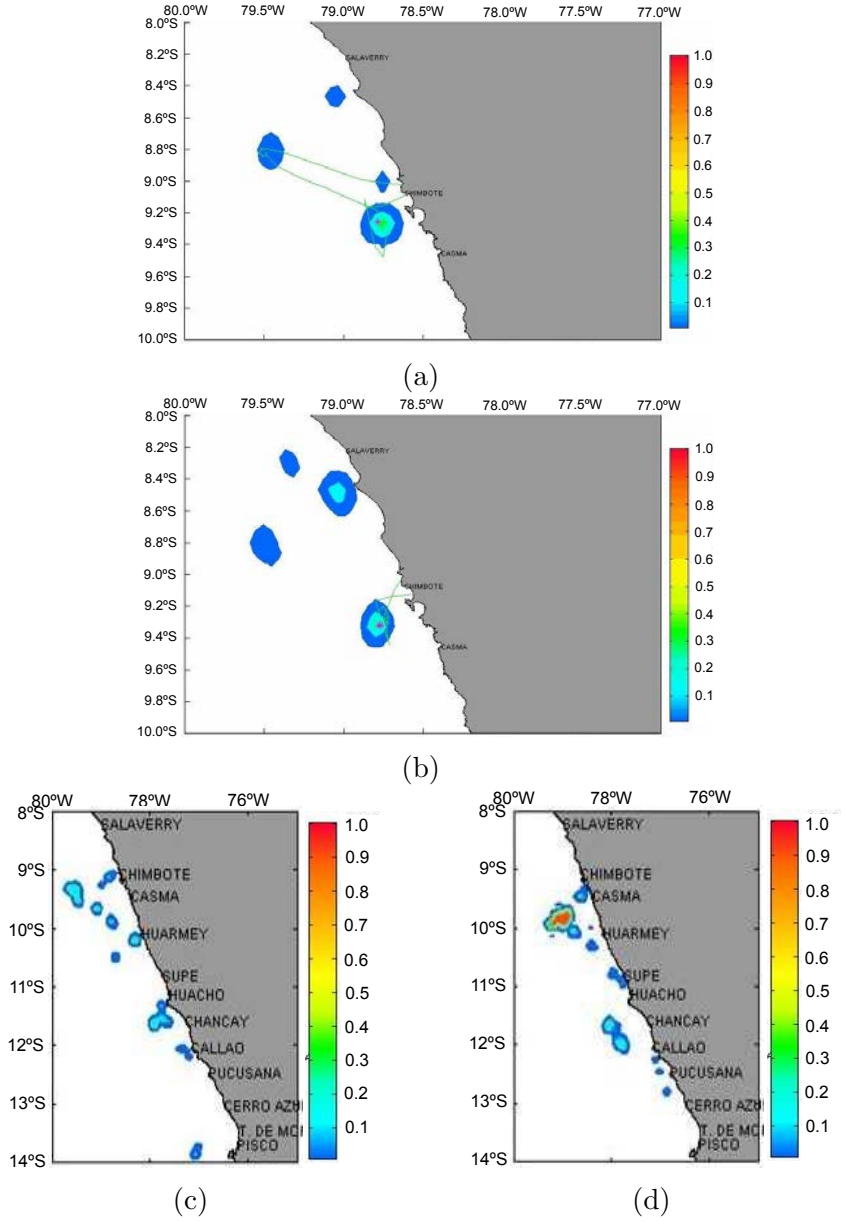


Figure 7.11: Clusters of fishing sets identified by kernels of density. Evolution of clusters are shown when separated by six hours (a) and (b), and when separated by a day (c) and (d).

most important ports. Other factors, like local closures of the fishery (e.g. after exceeding a catch limit of juveniles), or companies strategies (e.g. moving their fleet to a certain region where there are a priori good fishing spots or proximity to an available factory), also affect the activities distribution both in time and space.

From presence proxy to biomass proxy: not so straightforward

We have pointed out the intrinsic and extrinsic conditions than may introduce 'noise' (variability) in each dataset, which hinders the comparison between both maps. Another difficulty lies on the fact that the presence proxy is not compared to an 'actual' probability of anchovy presence, but to a scaled acoustic biomass. Actually, the interest behind the construction of a spatialized presence proxy is to use it for computing an index of biomass or abundance (Walker *et al.*, submitted). One could question why did we not build a proxy of biomass instead of a proxy of probability of presence. Building a proxy of biomass from fishermen activities may not be as straightforward as the probability of presence. There are no self-evident assumptions to make for the former. Based on the IFD theory, it was assumed that foraging effort, would distribute in space proportionally to the available resource. Foraging effort in space was represented here by weighted frequencies of activities by unit of space, defining an index called presence proxy. In turn, the available resource was represented by scaled log-transformed acoustic biomass. If indeed the available resource was well-represented by the acoustic biomass and the IFD-based assumption was met, it would be conceivable to use the presence proxy for extrapolating a biomass proxy. However, the assumption does not seem to hold. A key issue could lie in the fact that the availability of prey is not always proportional to its attractiveness for fishermen. This is important since the decision for fishing is based on how attractive an aggregation is considered. How concentrated, superficial and closed to port should a patch be and how high should its biomass be for it to be considered attractive, is a subjective question. Its answer could differ from one fishermen to another and depend on several other factors (e.g. the number of competitors interested in the patch, the local and global environmental conditions, their previous fishing success since the fishing season opened, and their previous success since the fishing trip started or how full is the vessel's hold). These relationships should be carefully explored before a proxy of biomass could be built.

7.4.3 Future work (for improving map comparison and matching)

Several issues could be worth exploring for improving map matching.

Developing a methodology for comparing maps. In this work, we built some indicators for comparing overall consistency between maps. Although this may be a

bit too 'hand-made' (for comparing high densities zones, highest-half and highest-quarter cell-grids from both maps were compared in terms of occupied area and center of gravity of distance to the coast), there are no standard methods for comparing map consistency (not grid cell by grid cell), like what was needed here. Methods for comparing maps by several criteria should be developed.

Preferential sampling. Pennino (2013) used preferential sampling geostatistical modeling (Diggle *et al.*, 2010), for estimating the distribution of European hake (*Merluccius merluccius*) abundance, using data from on-board observers deployed on board of fishing vessels in the Gulf of Alicante. It would be appealing to apply the preferential geostatistical modeling approach into the presence proxy model, and compare those results to both the classic geostatistical modeling of fishermen behavior and acoustic biomass results obtained here.

No searching. Concerning the weight parameters, the performance of the proxy interpolation show only small variations when changing the weights of searching activities. Since the inference of the searching activity was the least reliable (Joo *et al.*, 2013), using only the proportion of fishing activities at each cell grid for computing the presence proxy could be tested; i.e. considering searching and fishing weights equal to 0 and 1, respectively. However, we hypothesize that it will not cause fundamental improvements, since a 0.3 searching weight did not either.

Cruising at night. Here we assumed that a cruising activity indicate absence of anchovy, for all cruising activities excepting the last cruising segment of each trip, considering that in those cases, the fishermen have a priori decided to head back to port regardless of how much fish there is on the way. Since foraging (represented by searching and fishing activities) is mostly done during daytime (Fig. 7.12), we could also consider discarding the cruising activities occurring at night, as they are not based on fish presence/absence. The cruising records at night may be the ones causing the noise that we found in the data.

Bayesian non-informative priors. Fishermen behavior is shaped by the ecological, economical and management conditions (Joo *et al.*, in review). The different results obtained here for each year may be partly explained by that. Then, instead of fixing the values of the searching and cruising weights, we could model them using priors (Carlin and Louis, 2000; Congdon, 2006) that could cope with that variability.

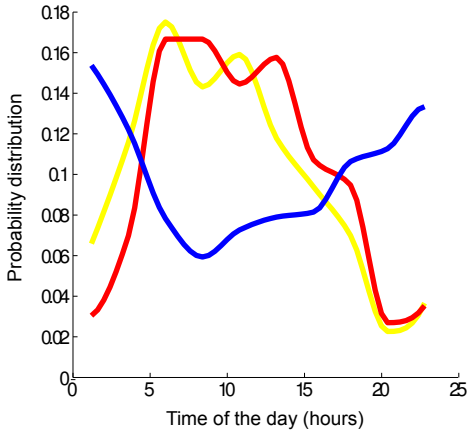


Figure 7.12: Distribution of activities during a day. The blue, red and yellow solid lines correspond to cruising, fishing and searching, respectively.

Fish schools depth. We tacitly assumed that fish depth does not constrain fishermen decisions. In practice, purse-seiners take into consideration the school depth in relation to the net size before deciding to fish. A threshold on fish school depth ($\sim$ 40m is the averaged maximum depth of fished schools registered by on-board observers) could be applied to the acoustic biomass dataset before spatial interpolation. In that case, we would assess how the spatial distribution of fishermen activities could be used as a proxy of the spatial distribution of shallow anchovy. If no information on fish school depth is available, oxycline depth (extracted from acoustic surveys; Bertrand *et al.*, 2010) could be used as a proxy.

Physical covariates. Previous works have shown a strong relationship between environmental variables (such as water masses distribution or Chlorophyll-a) and anchovy spatial distribution and biomass (Gutiérrez *et al.*, 2007; Joo *et al.*, in review; Swartzman *et al.*, 2008). Those variables could be used as spatial covariates for improving biomass interpolation, instead on relying only in the acoustic data collected through the transects (Georgakarakos and Kitsiou, 2008, obtained better results when using covariates for mapping abundance distribution of small pelagic species).

Indicators at fish aggregation level. It would be attractive to evaluate if clusters of high presence proxy correspond to fish patches (i.e., comparing 'fish clusters' identified by both methods). Using kernel densities (Botev *et al.*, 2010), for instance, as in Passuni (unpublished), clusters could be identified in both maps and the proportion of identified clusters could be used as an indicator of cluster recognition.

7.4.4 Synthesis

The use of spatialized data on fishermen behavior for producing maps of levels of fish presence or, even better, maps of a proxy of biomass, is highly appealing. However, the analyses from this work have shown that the maps of acoustic biomass and presence proxy produced did not strongly match. Nonetheless, an indicator of fishermen spatial distribution, distance to the coast, has proven to be a robust indicator of the longitudinal extent of anchovy. Besides, we found positive covariations between both maps at larger scales than the ones used for interpolation: between ~ 150 km and ~ 220 km. The potential of cross-wavelets analysis for simultaneously studying the space and scale of co-variations seems highly appealing. Several other issues have been discussed for improving the quality of the results and will be addressed in future works.

Chapter 8

General Conclusions and Perspectives

"It is not the facts that are of chief importance, but the light thrown upon them, the meaning in which they are dressed, the conclusions which are drawn from them, and the judgments delivered upon them."

– Mark Twain (variant of quote from *Amended Obituaries*, by Boyd *et al.* (2004))

In this work, we aimed at providing a first approach to fishermen behavioral ecology at multiple time and spatial scales (Fig. 8.1). We characterized the behavior of Peruvian anchovy fishermen by means of their trajectories and, for that reason, we began this manuscript by presenting a movement ecology framework (Chapter 2) and a broad outline on the Northern Humboldt Current System (Chapter 3).

As stated in Chapter 2, a movement path results from the succession of distinct behavioral modes. In Chapter 4, we analyzed the behavioral modes within the fishing trips. We compared several discriminative and Markovian models for inferring behavioral modes associated with fishermen trajectories, for a subset of $\sim$ 300 fishing tracks for which behavioral modes were known. They represented $\sim$ 1% of the total number of tracks from the fishing fleet in 2008. The better performance of Markovian models over the discriminative models highlighted the importance of modeling state dynamics for accurately inferring the behavioral mode sequences. Semi-Markov processes represent better the behavioral mode sequences than first-order Markov processes, since they explicitly model state duration and consider transitions at a segment scale (i.e. a sequence of consecutive steps associated with a same behavioral mode). Through a simulation experiment, it was shown that increasing time resolution (to at least 1 record each 30 minutes) significantly

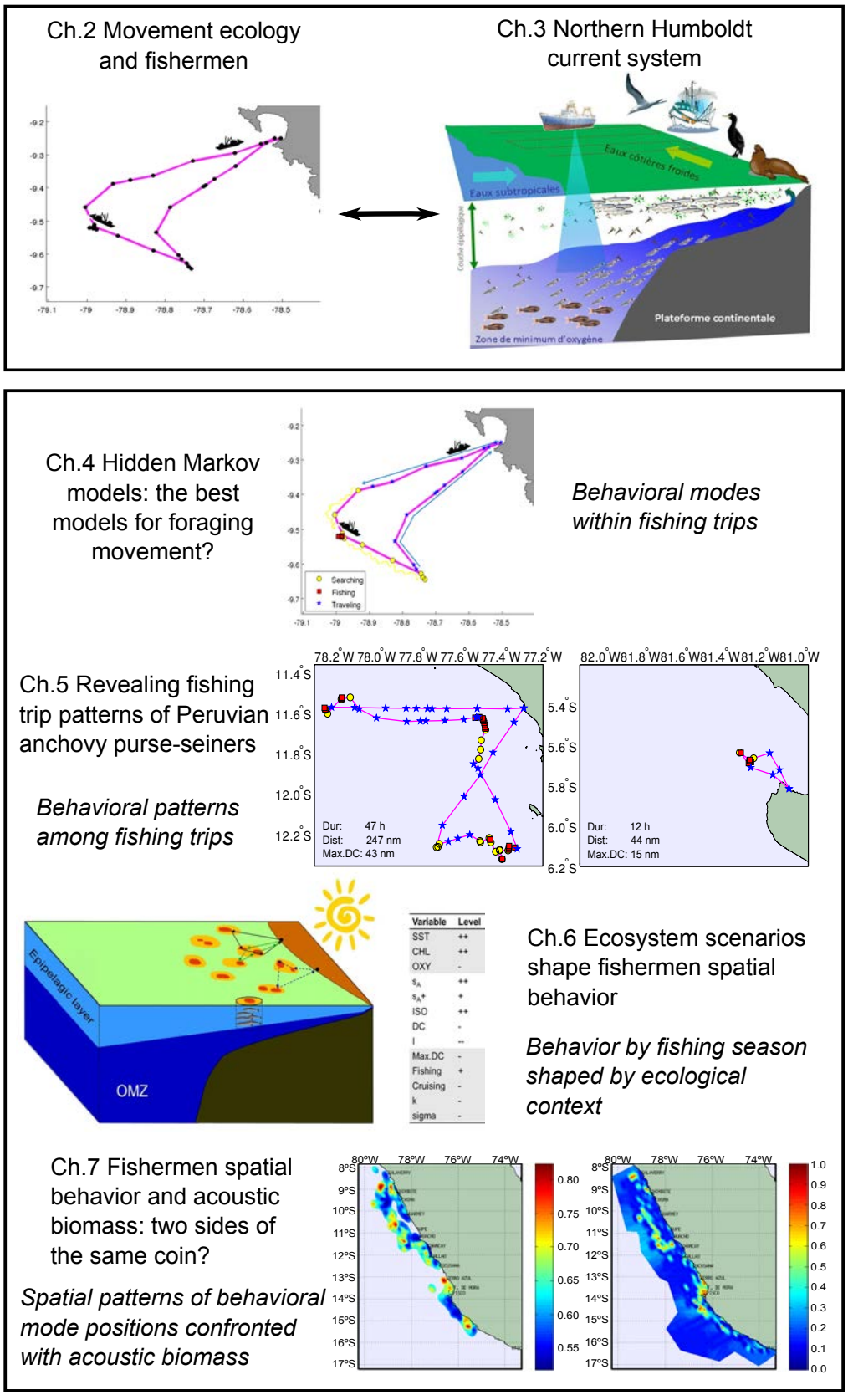


Figure 8.1: Synthesis of thesis work by chapter (Ch.). In italics, scales at which movement is analyzed.

increased the accuracy of Hidden Semi-Markov Model (HSMM) inference. We thus used HSMMs for inferring the behavioral modes on the tracks where the modes remained unknown, not only for 2008, but for the whole 2000-2009 period.

Then, we analyzed the behavioral patterns among fishing trips (Chapter 5). We characterized each fishing trip by a set of features; some of them based on the inference made in the previous chapter, such as time spent fishing or cruising. We used hierarchical clustering for analyzing how fishing trip patterns grouped. Four groups were found, associated with management zones (north-center and south), fleet segments (steel and wooden fleet) and skippers' personality (risk takers vs. risk adverse).

Next, we analyzed the adaptive behavior of fishermen to its ecological context at the scale of a fishing season (Chapter 6). We found significant associations between fishermen, anchovy and environment dynamics, and characterized the fishermen response to the observed ecosystem scenarios. Moreover, we found evidence that environmental fluctuations smooth out along trophic levels up to fishermen.

Finally, we analyzed the spatial patterns of behavioral mode positions of fishermen and compared them with acoustic biomass of anchovy (Chapter 7). Investigating if fishermen spatial behavior reflected anchovy spatial distribution, we built a proxy of anchovy presence using the geo-referenced behavioral modes and compared its spatial patterns to those of the acoustic biomass. Positive spatial co-variations between the presence proxy and the acoustic biomass were found at coarse scales. In addition, we found significant correlations in distance to the coast for two out of three fishing seasons studied. However, the maps of presence proxy did not accurately reflect the maps of acoustic biomass at a small scale. Both sources of information seem to complement each other; the maps of presence proxy could be rather used as maps of spatialized effort and/or maps of the anchovy seen by the fishermen. Potential improvements in the applied methods (proposed in section 7.4.3) could provide stronger coherence between the maps.

As stated in Chapter 2, the most common use of fishermen trajectory data is the identification of fishing set positions; the classification method employed often considers each positioning record as independent of all the others (e.g. a threshold on speed). Under such approach, one chooses to forget that those positions come from a trajectory, and as such, have a sequential nature. Neglecting the temporal

dimension of a trajectory is – to put it in other words – like taking the verses in a poem and re-writing them in one plain sentence: it is not poetry anymore. In this work, we considered the spatio-temporal nature of the trajectories when using hidden semi-Markov models for modeling the behavioral modes within the fishing trips. Despite the global accuracy is not overwhelmingly superior when compared to discriminative models, we proved that this kind of model represents better the sequences of behavior and their duration. As rhythm gives sound to a poem, behavioral mode sequences enrich our understanding of fishermen behavior, greatly contributing to the characterization of the fishing trips in Chapter 5, the response of fishermen to ecosystem scenarios in Chapter 6 and the elaboration of a map of anchovy presence according to what fishermen 'saw' in Chapter 7.

Each analysis was done at a different scale: the fishing trip scale, a fishing season scale, and a broader scale of spatialized behavioral patterns off Peru. The challenge was to understand the patters observed at each scale and their drivers. When studying behavior at the fishing trip scale, fleet segment and management factors showed to be key for discriminating the trips into different groups. No information on the biotic or abiotic conditions were available at very fine scale (e.g. fish biomass at each VMS record) for testing for environmental and biological factors conditioning fishermen behavior at that scale. Those data were available and analyzed at the fishing-season-like scale ($\sim$ 2-3 months), and the response of fishermen to anchovy and environment conditions was significant. This result contrasts with the comparison between the acoustic biomass and the anchovy presence proxy (computed from behavioral mode positions) at a $\sim$ 30 days scale. Although spatial covariation was found (at coarse scales), they did not seem to co-occur.

These apparently odd results may be the confirmation that, just like other patterns in ecology (Halley *et al.*, 2004), fishermen behavioral patterns are not scale invariant. It may suggest, thus, that in average, fishermen behavior does strongly respond to global changes in anchovy biomass, but, when it comes to spatial co-occurrence (with acoustic biomass), their dynamics may be influenced by factors operating at different scales (Hengeveld, 1987). The predating power of fishermen and their average trip duration ($\sim$ 24 hours, wherein they aim at filling their holding capacity) might make them more dynamic than anchovy; and certainly more dynamic than acoustic surveys. Besides, the fact that the zones with the highest fishing effort do not always correspond with the ones with the highest acoustic biomass may not be as strange as it seems. As stated by Hastings (2010), lack of

spatial synchrony may be the ultimate explanation for persistence of strongly interacting exploiter and 'victim' systems. Indeed, if fishermen knew exactly where to fish and they distributed in space exactly like the fish, the spatial co-occurrence would indicate an eventual – or proximate – collapse of the fish stock. The fact that fishermen exploit only some of the fish patches, and that they do not necessarily choose the biggest anchovy patches, but settle with patches large enough for fulfilling their holding capacity, contributes to sustainability. This could represent another piece in the puzzle of the anchovy paradox and the maintenance of the high productivity in the NHCS.

From a methodological point of view, the low co-occurrence between fishermen behavioral patterns and acoustic biomass may be partly conditioned by the differences in sampling method and interests: fishermen vs. scientists. Therefore, we could regard fishermen effort as another – preferential – sampling method. Pennino (2013) used preferential sampling geostatistical modeling (Diggle *et al.*, 2010), for estimating the distribution of European hake (*Merluccius merluccius*) abundance, using data from observers deployed on board fishing vessels in the Gulf of Alicante. It would be appealing to apply the preferential geostatistical modeling approach in future works, and compare those results to both the classic geostatistical modeling of fishermen behavior and acoustic biomass results obtained here.

The analyses made in this work provide a better understanding of fishermen behavior and their drivers at different scales. In Figure 8.3, a Stommel diagram of fishermen behavior is presented. We show the behavioral units studied (behavioral mode, fishing trip, fishing season), the factors proven to condition behavior at each scale (black arrow), and the fish and environment structures corresponding to the studied scales. Through this work, we showed that, at the scale of a behavioral mode, it is the internal states that accounts mostly for inferring behavioral modes. This internal state manifested itself through the observed track (i.e. speeds and turning angles) and the behavioral sequences (e.g. the behavior in the previous segment in the sequence conditioned the behavior in the next segment). Those components allowed obtaining 80% of accuracy when inferring behavioral modes through hidden semi-Markov models. At the fishing trip scale, the main drivers were management rules (north-center and south regions), fleet segments (steel vs. wooden) and skippers personality (risk takers vs. followers). The explained variance by the clusters denoting these drivers was 61%. At the fishing season, anchovy biomass and distribution, and environmental factors such as sea surface temperature, Chlorophyll-a

and oxycline depth significantly conditioned fishermen spatial behavior (0.63 and 0.55 of association between fishermen and fish, and fishermen and environment, respectively). Evidently these results are opportunistic, because they dependent on the available data at each scale. In spite of this, the levels of explained variance are high enough to ensure that the enlightened processes are the most important ones.

Concerning behavioral modes, for example, fishing schools could play a role in the decision on whether to fish or continue searching. As shown in this work, fishermen preferred to forage (search and fish) at daytime. This is probably because at day, fish are aggregated in schools, most likely as a response to predation (Gerlotto *et al.*, 2006); while at night, the vertical upward migration of zooplankton and reduced predation on anchovy create a more suitable habitat, where anchovy is distributed as loose shoals and scattered fish (Bertrand *et al.*, 2008a, Fig. 8.2). Bertrand *et al.* (2008a) showed that internal waves play an important role in the distribution of fish at very small scales (from meters to kilometers, Fig. 8.3), basically in two ways. First, internal waves create convergences that concentrate prey above the oxycline, which is particularly important during the day when most zooplankton is otherwise distributed below the oxycline. Second, they increase the available habitat by deepening the oxycline, allowing anchovy to form larger and more elongated schools. Fishermen could find these large schools attractive enough for triggering a fishing operation. If not, they may expect a set of consecutive large schools, which respond to submesoscale structures: cluster size depends on submesoscale physical features (e.g., upwelling plumes, eddies) that shape the distribution of zooplankton patches (Bertrand *et al.*, 2008a). We hypothesize that eventual data on internal waves and submesoscale processes (e.g., using echosounder records of fishing vessels, or acoustic Doppler current profilers and vertical temperature observations as in Nakada *et al.*, 2013) could improve the inference of behavioral modes and thus our representation of the behavioral dynamics at that scale, particularly for discriminating between fishing and searching behavior. Still, it is up to the fishermen to decide whether the school is 'attractive enough' for fishing it; criteria for 'attractiveness' may not be the same for every skipper; as stated by Gaertner *et al.* (1999), the cause of any particular skipper decision can be multiple and conjectural.

Regarding a fishing trip (as a behavioral unit), in a context where fishing trips last $\sim$ 24 hours, the trip descriptors describing where to go and how much time to stay in each zone and how much time to spend in each behavioral mode, could be conditioned to knowledge or discovery of zones of large anchovy clusters. In a

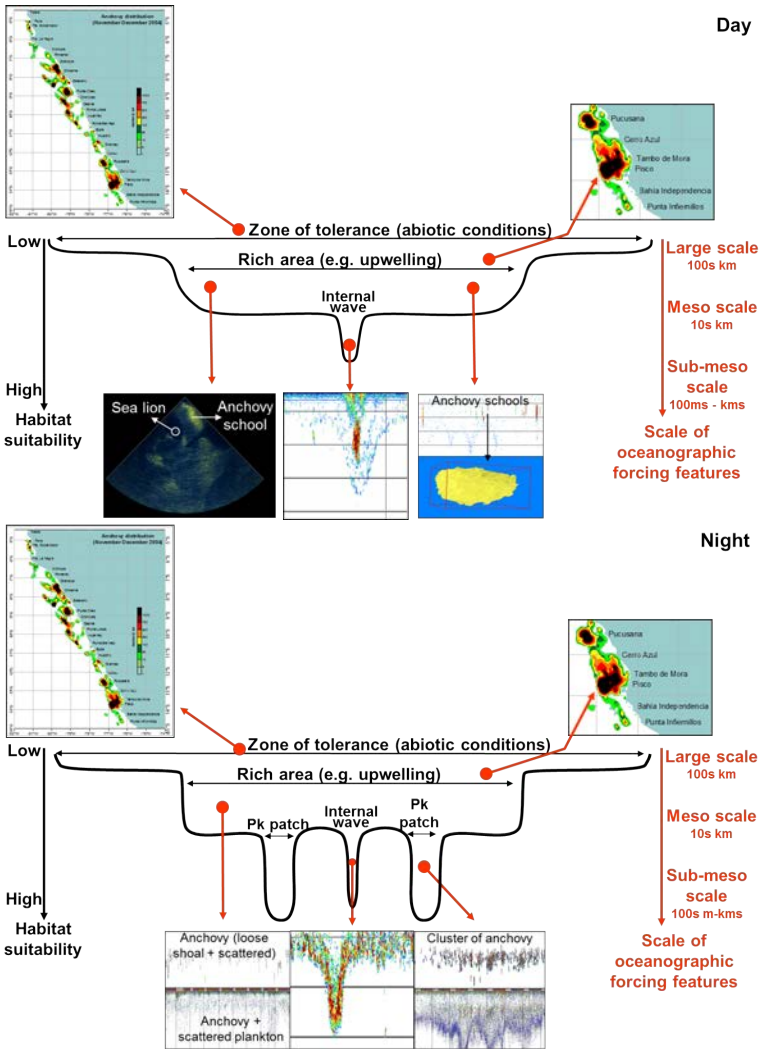


Figure 8.2: Synthetic representation of the 3D spatial distribution of anchovy, adapted from the 'habitat-based basin model framework' of MacCall (1990). The depth of the basin increases with habitat quality in terms of oceanographic and biotic features. The larger basin corresponds to the zone of anchovy distribution. Habitat quality increases in areas rich in prey surrounding an upwelling zone for example. Inside the rich areas, submesoscale structures and internal waves concentrate prey, further increasing habitat quality. During the day, the depth of the basin is shallower than during the night since prey are less accessible (some of the plankton have migrated below the oxycline) and predation by visual top predators is higher. Anchovy form dense schools. During the night (lower figure) the depth of the basin increases as prey become more available and predation is reduced. Fish are no longer able to construct schools but are concentrated in prey patches or internal waves, when present. Encapsulated figures above the basin figures show anchovy distribution as evaluated during an acoustic survey performed just after the experiment illustrating the range of distribution of anchovy off Peru (left) and an upwelling area (right). Encapsulated figures below the basin figures show typical examples of fish collective structure in each case as observed with the multibeam sonar or with the echosounder. Source: Bertrand *et al.* (2008a)

fishing trip, the goal basically consists in filling the holding capacity of the vessel in $\sim$ 24 hours and spending as less fuel as possible. Fishermen need to develop tactics for achieving that goal. In the southern region, the platform is very close to the coast, which makes anchovy distribute near the coast, conditioning the geometry and proximity to coast of fishermen trips. In general, in north-center and south regions, fishing trips are not too far from the coast, since anchovy can be found not too far away in its cold coastal water habitat (Swartzman *et al.*, 2008). Among the vessels, the wooden fleet is more sensitive to getting further away, because they need to catch less and have typically worse refrigeration system than industrial vessels. Among both fleet segments, most fishermen showed a Cartesian behavior, i.e. a way of decision making where fishermen only go to locations where present information tells them that the 'highest returns' can be found (Allen and McGlade, 1986). And the few risk takers, or stochasts, are willing to explore new zones for identifying large aggregations of prey. Risk takers are mostly from the steel fleet; those are bigger vessels with typically better refrigeration systems. Fish biomass and distribution could have an impact on fishing trips, but when considering what triggers each fishing trip, it seems that the vessel and fishermen characteristics, as well as the division between management zones, are the most important factors.

At the fishing season, anchovy and environmental conditions were computed. At this scale, fishing season showed to be conditioned by anchovy and environmental factors. In turn, fish stocks are determined by mesoscale physical features increasing and concentrating productivity in the upwelling cells. Because the analyzed fishing seasons did not encompass any El Niño event, the seasonal dynamics were the strongest environmental dynamics shaping fishermen spatial behavior. Summer seasons, associated with higher productivity, more superficial oxycline and cold coastal waters close to the coast, favor abundant, locally-concentrated anchovy biomass (Gutiérrez *et al.*, 2007; Swartzman *et al.*, 2008) and closer to the coast. These conditions make anchovy more easily available and average effort for fishing is reduced (trips are shorter, closer to the coast and more time of the trip is spent fishing). Conversely, winter and spring seasons, characterized by lower sea surface temperature and productivity, and a deeper oxycline, are associated to lower local and global biomass. As a result, fishermen have to go farther to fish, fishing trips last longer and more time is spent cruising.

For a larger time-series, – for which on-board observers data were used for characterizing fishermen behavior at the season scale since 1996 – Bertrand *et al.* (2008b)

analyzed how large scale oceanic forcing, via Kelvin waves, affected the coastal ecosystem. Downwelling Kelvin waves, associated with warm (El Niño) scenarios, were characterized by an increase in sea surface temperature and a smaller extent of cold coastal water masses, making anchovy to be closer to the coast and deeper, following the deepening of the oxycline. Because of the high patchiness of anchovy, fishermen movement is more diffusive (i.e., there are a lot more short moves than long moves, as explained in section 2.4), trips are shorter and closer to the coast; and because anchovy is deeper, it is less accessible and less catches are made. Conversely, upwelling Kelvin waves, associated with normal and La Niña conditions, were characterized by lower temperatures and a larger extent of cold coastal waters, providing anchovy with a larger habitat. Anchovy is thus distributed in larger areas and less patchy than in El Niño conditions, but also more superficial. Fishermen movement is then less diffusive, as they need to explore larger areas, fishing trips last longer and more catches are made.

Hence, the analyses made in this work and previous works can give a rich picture of the multiscale fishermen behavior.

Still, many other components could enrich this analysis. First, behavioral pattern dynamics throughout a fishing season seem appealing to study. Two contrasting hypothesis could be tested: (1) in the first days of the season, vessels will start close to the coast, with similar durations and distances traveled, but because they do not know the fishing grounds, they will present more sinuous paths and high variance in the time spent at each behavioral mode; (2) due to fishing ground prospection and fixed tactics by fishing companies at the beginning of the season, fishermen will have decided where to fish in advance, so their paths will be more ballistic, although, as in the first hypothesis, the variance in time spent in behavioral modes will be high because a priori fixed strategies will work for only some of them; also they will remain near the coast in the first days and move away progressively.

Other aspects could be studied as well. Future works should include the social (collective behavior) and economical conditions (e.g. oil price, fishmeal price, and environmental repercussions through life cycle assessment) that also shape fishermen behavior. Another factor is the quota system. In 2009, an individual vessel quota (IVQ) system was introduced in the Peruvian anchovy fishery. Interviews with fishermen and fishing company managers suggest that fishermen collective behavior changed drastically after the IVQ system started: the race for fish ended

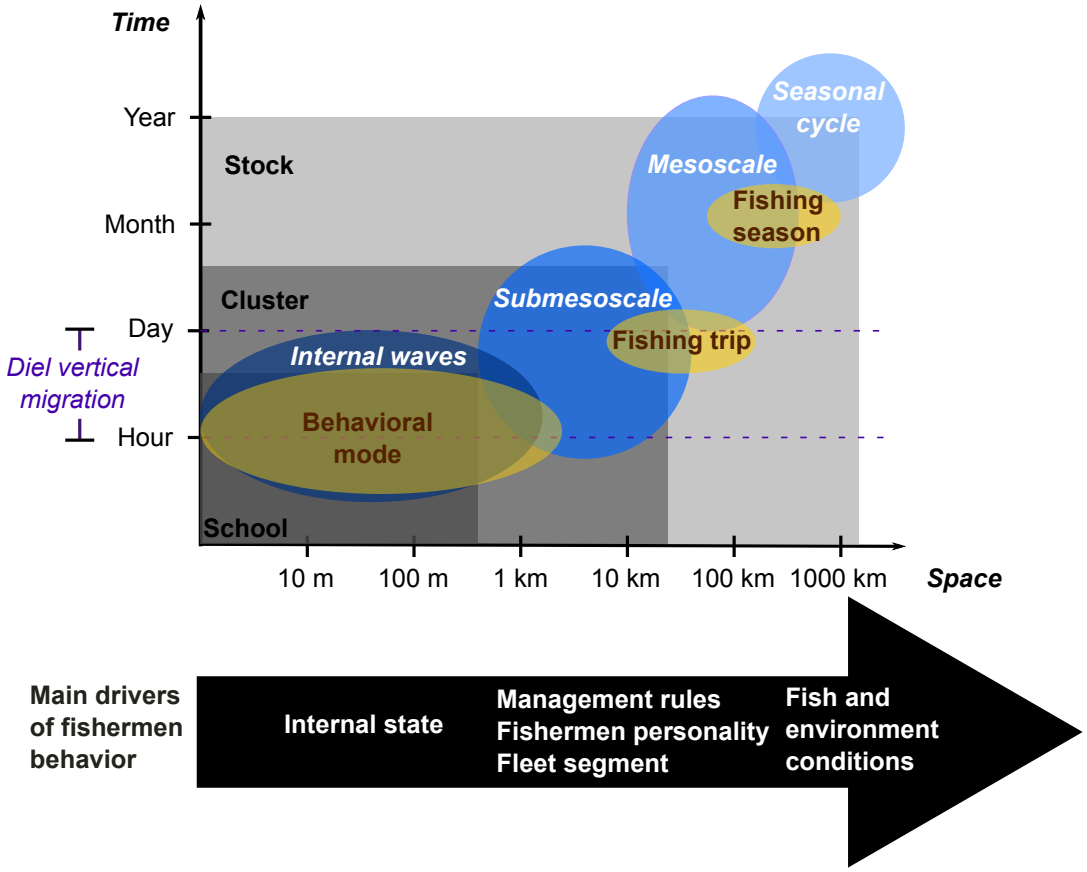


Figure 8.3: Stommel diagram of fishermen behavioral units and their drivers. The spatio-temporal scales for fish and environmental structuring where based on Chelton (2001); Dickey and Lewis (2006) and Bertrand *et al.* (2008a).

and so the competition turned into association. Due to the observed dynamism of the fishery, we propose that the competition/association be analyzed at the scale of the trajectory rather than at each VMS record, in order to take into account the spatio-temporal interactions between tracks. Individual-based models (Reynolds, 1987) or other models used in animal collective behavior (Sumpter *et al.*, 2012) could be explored; interactions between individuals could differ based on the main drivers found in this work for several scales (Fig. 8.3). With the increasing availability of data from the monitoring of ecosystems, the framework presented here can be used for analyzing fishermen behavioral ecology in other fisheries and ecosystems.

In order to improve the accuracy of the analyses and thus the fisheries management, we stress some important recommendations. First, there is a need for increasing the frequency of the VMS records. We have shown via simulation that higher frequencies (than 1 record per hour) can increase the accuracy of the behavioral mode inference, especially for modes that can last shorter than the time between consecutive emissions (the searching mode in our case). Time steps of 30 minutes or even better, 15 minutes, could greatly increase the inference performance and so the accuracy of the subsequent analyses. It would also open the opportunity of using many other models, like spatio-temporal models, or wavelet analyses, which would provide new insights into fishermen behavior. Another important issue is to take advantage of complementary sources of fishermen monitoring. In this work, the on-board observers data has been essential for calibrating and then validating the models for inferring the modes. Although recognizing the great effort of IMARPE (Instituto del Mar del Perú) for maintaining their on-board observers program, which allowed obtaining > 200 fishing trips with concomitant VMS and observers data per year, an increase in the number of observed fishing trips would increase the generalization power of the models and thus, their validation performance. Because each database had a distinct vessel ID system, the name of the vessel was used for relating the databases; names were sometimes misspelled and in some cases, vessels were renamed, causing problems for relating the databases and probable loss of information during the process. For that reason, we strongly advise that a unique vessel ID system be used in all databases, so that relating the databases be more efficient. In regions where on-board observers programs are not implemented, logbook data can be used for training the models, although it should be considered that they are not independent samples (from the vessel crew).

The indicators of fishermen behavior computed here (e.g. time spent in each behavioral mode, maximum distance to the coast, sigma, k) showed to be sensitive to ecosystem and management conditions. Therefore, we recommend their use as effort indicators, additionally to the CPUE indicators commonly used. In addition, the availability of the spatial distribution of behavioral modes, and particularly, the maps of presence proxy can be used as maps of fishing effort. Another recommendation for ecosystem-based management, is the incorporation of seabirds in fisheries management, as mentioned in section 2.7.5. Despite not accounting for the interaction between vessels, seabirds and marine mammals in the present work, high-resolution tracking data is being collected on samples of seabirds and fur seals in reproduction zones during breeding seasons. A first study on competition for prey between seabirds and vessels was published (Bertrand *et al.*, 2012), and more work is in progress. This first work has shown that the competition of fishing vessels affects seabirds foraging behavior, probably causing local depletion and making them go farther in the search for prey. The interactions of fishermen and competitor predators should be monitored and managed in order to guarantee the sustainability of those predators populations.

From a methodological perspective, we would like to emphasize the importance of model validation. Indeed, the increasing access to Vessel Monitoring System data is opening wide opportunities for monitoring and modeling. As we have shown here, and have also shown Bertrand *et al.* (2008c); Joo *et al.* (2011); Palmer (2008); Walker and Bez (2010) is that the evaluation of model performances (e.g. for identifying fishing sets or behavioral modes in general) should not be disregarded. Especially for management purposes, it should be critical to measure the accuracy and precision of our estimates, instead of blindly trusting our results. Referring to the lack of validation, Jerome Sacks said that in this era of methodological advances, 'uncertainty is as uncertain as ever' (Sacks and Ylvisaker, 2012). Objective model validation implies (1) having an independent groundtruthed dataset to evaluated the model, and (2) a priori defining criteria for evaluating a model, i.e. indicators of performance that depend on how we define that results are good. Performance indicators under those criteria will help respecting the parsimony principle or Occam's razor, by choosing a simple model until its simplicity can be traded for greater explanatory power (i.e. better performance). Indeed, model validation can be useful in two ways: On the one hand, it saves us from the risk of using a very simple and practical model but that may give very inaccurate results; in this case, validation would tell us to look for a more complex model that fits better the data. On the

other hand, it saves us from the growing tendency to look for a more complex (i.e. more parametrized) model that can better represent reality as we think it is or as our data used for calibration tell us it is. In this case, validation would tell us that the computational prize we are paying for a complex model is not worth it and we should stop at a more parsimonious model.

Regarding complex models, end-to-end models combine physicochemical oceanographic descriptors and organisms ranging from microbes to top predators (Rose *et al.*, 2010). The demand for such approaches arises from the need for quantitative tools for ecosystem-based management, for accounting for multiple changes in the ecosystem, including climate change and fisheries management policies. However, in those models, fishermen are not modeled as a component of the ecosystem that is conditioned by it; catches only vary due to management restrictions (when it does); we have proven that it is not accurate. Both the inclusion of humans and of organisms' behavioral movement had been recognized as important issues for further development of end-to-end models (Rose *et al.*, 2010). The complementary indicators of fishing effort and behavior presented here, could be incorporated to those models to take into account the variability in fishermen behavior; and the explored relations with the environment and anchovy could give a starting point for determining the relationships between fishermen and the ecosystem in those models.

Bibliography

Aguilar Ibarra, A., C. Reid, and A. Thorpe, 2000. The political economy of marine fisheries development in Peru, Chile and Mexico. Journal of Latin American Studies 32:503–527. 61

Akaike, H., 1981. Likelihood of a model and information criteria. Journal of Econometrics 16:3–14. 79

Alheit, J. and M. Ñiquen, 2004. Regime shifts in the Humboldt Current ecosystem. Progress in Oceanography 60:201–222. 55, 60, 133

Allauca, S., 1990. Presencia de la corriente costanera ecuatoriana. Acta Oceanográfica del Pacífico 6:10–17. 47

Allen, P. M. and J. M. McGlade, 1986. Dynamics of discovery and exploitation: the case of the Scotian Shelf Groundfish fisheries. Canadian Journal of Fisheries and Aquatic Sciences 43:1187–1200. xlv, 98, 108, 178

Aranda, M., 2009. Developments on fisheries management in Peru: The new individual vessel quota system for the anchoveta fishery. Fisheries Research 96:308–312. xxix, 60, 61, 63, 65, 67, 98, 144

Arias Schreiber, M., M. Ñiquen, and M. Bouchon, 2011. Coping strategies to deal with environmental variability and extreme climatic events in the Peruvian anchovy fishery. Sustainability 3:823–846. 65, 66, 67

Arntz, W. E., E. Fahrbach, C. Wosnitza-Mendo, and J. Mendo, 1996. El Niño: Experimento climático de la naturaleza. México. 53, 58

Aswani, S. and M. Lauer, 2006. Incorporating fishermen's local knowledge and behavior into Geographic Information Systems (GIS) for designing marine protected areas in Oceania. Human Organization 65:81–102. 35

AVISO, 2013. What is El Niño/La Niña? http://www.aviso.oceanobs.com/en/applications/climate/el-nino/what-is-el-ninola-nina.html. Accessed: 2013-10-08. xxviii, 50

Ayón, P., M. I. Criales-Hernandez, R. Schwamborn, and H. J. Hirche, 2008a. Zooplankton research off Peru: A review. Progress in Oceanography 79:238–255. xxviii, 48, 53

Ayón, P., G. Swartzman, A. Bertrand, M. Gutiérrez, and S. Bertrand, 2008b. Zooplankton and forage fish species off Peru: Large-scale bottom-up forcing and local-scale depletion. Progress in Oceanography 79:208–214. 116

Bai, J., J. Goldsmith, B. Caffo, T. A. Glass, and C. M. Crainiceanu, 2012. Movelets: A dictionary of movement. Electronic Journal of Statistics 6:559–578. 21

Bakun, A. and R. Mendelssohn, 1989. Along shore wind strees 1953-82: correction, reconciliation and update through 1986. In D. Pauly, P. Muck, J. Mendo, and I. Tsukayama, editors, The Peruvian upwelling ecosystem: dynamics and interactions, pages 77–81. International Center for Living Aquatic Resources Management (ICLARM), Manila, Philippines. 59

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 79:290–299. 53, 58

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 91:360–381. 53

Barange, M., M. Bernal, M. Cercole, L. Cubillos, G. Daskalov, C. Cunningham, J. de Oliveira, M. Dickey-Collas, D. Gaughan, K. Hill, L. Jacobson, F. Köster, J. Massé, M. Ñiquen, H. Nishida, Y. Oozeki, I. Palomera, S. Saccardo, A. Santojanni, R. Serra, S. Somarakis, Y. Stratoudakis, A. Uriarte, C. van der Lingen, and A. Yatsu, 2009. Current trends in the assessment and management of stocks. In D. Checkley, J. Alheit, Y. Oozeki, and C. Roy, editors, Climate change and small pelagic fish. vol. 9, pages 191–255. Cambridge University Press, Cambridge. 69

Barber, R. T. and F. P. Chavez, 1983. Biological consequences of El Niño. Science 222:1203–1210. 58, 133

Bartumeus, F., 2007. Lévy processes in animal movement: an evolutionary hypothesis. Fractals 15:151–162. 13

Bartumeus, F., 2009. Behavioral intermittence, Lévy patterns, and randomness in animal movement. Oikos 118:488–494. 14

Bendezú, S., 2008. Concentración de la propiedad de la flota y de plantas de harina de pescado en el Perú. Pesca 95:4–9. 65

Bengio, Y., 1999. Markovian models for sequential data. Neural Computing Surveys 2:129–162. 78

Benhamou, S., 2004. How to reliably estimate the tortuosity of an animal's path: straightness, sinuosity, or fractal dimension? Journal of theoretical biology 229:209–20. 10

Benhamou, S., 2007. How many animals really do the Lévy walk? Ecology 88:1962–1969. 13, 14, 23

Bergman, C. M., J. A. Schaefer, and S. N. Luttich, 2000. Caribou movement as a correlated random walk. Oecologia 123:364–374. 74

Bernstein, C., A. Kacelnik, and J. R. Krebs, 1991. Individual decisions and the distribution of predators in a patchy environment. II. The influence of travel costs and structure of the environment. Journal of Animal Ecology 60:205–225. 143

Bertrand, A., M. Ballón, and A. Chaigneau, 2010. Acoustic observation of living organisms reveals the upper limit of the oxygen minimum zone. PLoS one 5:e10330. 49, 168

Bertrand, A., A. Chaigneau, S. Peraltilla, J. Ledesma, M. Graco, F. Monetti, and F. P. Chavez, 2011. Oxygen: a fundamental property regulating pelagic ecosystem structure in the coastal southeastern tropical Pacific. PloS one 6:e29558. xlii, 49, 55, 60, 116, 119, 135

Bertrand, A., F. Gerlotto, S. Bertrand, M. Gutiérrez, L. Alza, A. Chipollini, E. Díaz, P. Espinoza, J. Ledesma, R. Quesquén, S. Peraltilla, and F. Chavez, 2008a. Schooling behaviour and environmental forcing in relation to anchoveta distribution: An analysis across multiple spatial scales. Progress in Oceanography 79:264–277. xxvii, xxxv, xlviii, 144, 176, 177, 180

Bertrand, A., M. Segura, M. Gutierrez, and L. Vasquez, 2004a. From small-scale habitat loopholes to decadal cycles: a habitat-based hypothesis explaining fluctuation in pelagic fish populations off Peru. Fish and Fisheries 5:296–316. 51, 54, 55, 57, 58, 60, 61, 116, 133, 136, 137

Bertrand, S., 2005. Analyse comparée des dynamiques spatiales des poissons et des pêcheurs: mouvements et distributions dans la pêcherie d'anchois *(Engraulis ringens)* du Pérou. Ph.D. thesis, Ecole Nationale Supérieure Agronomique de Rennes. 25

Bertrand, S., 2013. Comportement spatial des prédateurs supérieurs et approche écosystémique des pêche: Pêcheurs, oiseaux et otaries dans lecosystème dupwelling du courant de humboldt. Mémoire d'Habilitation à Diriger des Recherches. Université de Montpellier II. 27, 111, 144, 145

Bertrand, S., A. Bertrand, R. Guevara-Carrasco, and F. Gerlotto, 2007. Scale-invariant movements of fishermen: The same foraging strategy as natural predators. Ecological Applications 17:331–337. xli, 2, 26, 27, 39, 70, 77, 99, 116, 134, 136, 147, 221

Bertrand, S., J. M. Burgos, F. Gerlotto, and J. Atiquipa, 2005. Lévy trajectories of Peruvian purse-seiners as an indicator of the spatial distribution of anchovy *(Engraulis ringens)*. ICES Journal of Marine Science 62:477–482. xli, 15, 27, 29, 70, 77, 99, 136, 147, 221

Bertrand, S., B. Dewitte, J. Tam, E. Díaz, and A. Bertrand, 2008b. Impacts of Kelvin wave forcing in the Peru Humboldt Current system: Scenarios of spatial reorganizations from physics to fishers. Progress in Oceanography 79:278–289. 29, 38, 58, 70, 116, 117, 133, 135, 147, 178

Bertrand, S., E. Diaz, and M. Lengaigne, 2008c. Patterns in the spatial distribution of Peruvian anchovy *(Engraulis ringens)* revealed by spatially explicit fishing data. Progress In Oceanography 79:379–389. 20, 27, 28, 29, 39, 70, 76, 117, 145, 147, 161, 163, 182

Bertrand, S., E. Diaz, and M. Niquen, 2004b. Interactions between fish and fisher's spatial distribution and behaviour: an empirical study of the anchovy fishery of Peru. ICES Journal of Marine Science 61:1127–1136. 130

Bertrand, S., R. Joo, and R. Fablet, in review. Generalized Pareto for a pattern-oriented random walk modelling of organisms' move . xxvii, xxviii, 14, 15, 16, 26, 27, 28, 39, 121

Bertrand, S., R. Joo, C. A. Smet, Y. Tremblay, and H. Weimerskirch, 2012. Local depletion by a fishery can affect seabird foraging. Journal of Applied Ecology . 38, 70, 140, 141, 144, 147, 164, 182

Bestley, S., I. D. Jonsen, M. A. Hindell, C. Guinet, and J.-B. Charrassin, 2012. Integrative modelling of animal movement: incorporating in situ habitat and behavioural information for a migratory marine predator. Proceedings of the Royal Society B: Biological Sciences . 22

Bestley, S., T. A. Patterson, M. A. Hindell, and J. S. Gunn, 2010. Predicting feeding success in a migratory predator: integrating telemetry, environment, and modeling techniques. Ecology 91:2373–2384. 22, 74

Beyreuther, M. and J. Wassermann, 2011. Hidden semi-Markov model based earthquake classification system using weighted finite-state transducers. Nonlinear Processes in Geophysics 18:81–89. 80

Bez, N. and S. Bertrand, 2010. The duality of fractals: roughness and self-similarity. Theoretical Ecology . 10

Bez, N., E. Walker, D. Gaertner, J. Rivoirard, and P. Gaspar, 2011. Fishing activity of tuna purse seiners estimated from vessel monitoring system (VMS) data. Canadian Journal of Fisheries and Aquatic Sciences 68:1998–2010. 33, 40

Bhattacharya, K. and T. Vicsek, 2010. Collective decision making in cohesive flocks. New Journal of Physics 12:093019. 24

Bird, D. W. and J. F. O'Connell, 2006. Behavioral ecology and archaeology. Journal of Archaeological Research 14:143–188. xxxix, 2

Black, A. J. and A. J. McKane, 2012. Stochastic formulation of ecological models and their applications. Trends in ecology and evolution 27:337–45. 9

Block, B. A., I. D. Jonsen, S. J. Jorgensen, A. J. Winship, S. A. Shaffer, S. J. Bograd, E. L. Hazen, D. G. Foley, G. A. Breed, A.-L. Harrison, J. E. Ganong, A. Swithenbank, M. Castleton, H. Dewar, B. R. Mate, G. L. Shillinger, K. M. Schaefer, S. R. Benson, M. J. Weise, R. W. Henry, and D. P. Costa, 2011. Tracking apex marine predator movements in a dynamic ocean. Nature 475:86–90. 10

Bograd, S. J., B. A. Block, D. P. Costa, and B. J. Godley, 2010. Biologging technologies: new tools for conservation. Introduction. Endangered Species Research 10:1–7. 115

Botev, Z. I., J. F. Grotowski, and D. P. Kroese, 2010. Kernel density estimation via diffusion. The Annals of Statistics 38:2916–2957. 164, 168

Bouchon, M., P. Ayón, J. Mori, C. Peña, P. Espinoza, L. Hutchings, B. Buitron, A. Perea, C. Goicochea, and M. Messié, 2010a. Biología de la anchoveta peruana, *Engraulis ringens* Jenyns. Boletín Instituto del Mar del Perú 25:23–30. 54

Bouchon, M., E. Diaz, D. Espinoza, and J. Limache, 2010b. Effects of the individual vessel quota allocations on the anchoveta fishery: Peruvian experience. Presented at the International Conference: Environment and Resources of the South Pacific, Viña del Mar, Chile. 111

Bourlard, H. and N. Morgan, 1994. Connectionist speech recognition: A hybrid approach. Kluwer Academic Publishers. 94

Bourlard, H. and N. Morgan, 1998. Hybrid HMM/ANN systems for speech recognition: Overview and new research directions. In Adaptive Processing of Sequences and Data Structures, ser. Lecture Notes in Artificial Intelligence. 94

Boyd, I. L., A. Kato, and Y. Ropert-Coudert, 2004. Bio-logging science: sensing beyond the boundaries. Memoirs of National Institute of Polar Research 58:1–14. 115, 140, 171

Boyd, I. L., S. Wanless, and C. Camphuysen, editors, 2006. Top predators in marine ecosystems: their role in monitoring and management. Cambridge University Press. 140

Breed, G. A., D. P. Costa, I. D. Jonsen, P. W. Robinson, and J. Mills-Flemming, 2012. State-space methods for more completely capturing behavioral dynamics from animal tracks. Ecological Modelling 235-236:49–58. 19

Breiman, L., 2001. Random forests. Machine Learning 45:5–32. xliii, 81

Browman, H. I. and K. I. Stergiou, 2004. Marine Protected Areas as a central element of ecosystem-based management: defining their location, size and number. Marine Ecology Progress Series 274:269–303. 36, 140

Brown, R., 1828. XXVII. A brief account of microscopical observations made in the months of June, July and August 1827, on the particles contained in the pollen of plants; and on the general existence of active molecules in organic and inorganic bodies. The Philosophical Magazine 4:161–173. 13

Buhl, J., D. J. T. Sumpter, I. D. Couzin, J. J. Hale, E. Despland, E. R. Miller, and S. J. Simpson, 2006. From disorder to order in marching locusts. Science 312:1402–1406. 24

Burges, C. J. C., 1998. A tutorial on support vector machines for pattern recognition. Data Mining and Knowledge Discovery 2:121–167. xliii, 81

Byun, H. and S.-W. Lee, 2002. Applications of support vector machines for pattern recognition: A survey. In Proceedings of the First International Workshop on Pattern Recognition with Support Vector Machines, SVM '02, pages 213–236, London, UK. Springer-Verlag. 76

Carlin, B. P. and T. A. Louis, 2000. Bayes and empirical Bayes methods for data analysis. Chapman & Hall/CRC, second edi edition. 167

Carr, M.-E. and E. J. Kearns, 2003. Production regimes in four Eastern Boundary Current systems. Deep Sea Research Part II: Topical Studies in Oceanography 50:3199–3221. 43

Castillo, R., S. Peraltilla, A. Aliaga, M. Flores, M. Ballón, J. Calderón, and M. Gutiérrez, 2009. Protocolo técnico para la evaluación acústica de las áreas de distribución y abundancia de recursos pelágicos en el mar peruano. Instituto del Mar del Perú . xli, 68, 119, 146

Catry, P., R. T. Lemos, P. Brickle, R. A. Phillips, R. Matias, and J. P. Granadeiro, 2013. Predicting the distribution of a threatened albatross: The importance of competition, fisheries and annual variability. Progress in Oceanography 110:1–10. 36

Chaigneau, A., 2013. Dynamique de grande échelle et activité de mésoéchelle dans le Système du Courant de Humboldt. Mémoire d'Habilitation à Diriger des Recherches. Université Paul Sabatier de Toulouse III. xxviii, 46, 47

Chaigneau, A., N. Dominguez, G. Eldin, and L. Vasquez, 2013. Near-coastal circulation in the Northern Humboldt Current System from shipboard ADCP data. Journal of Geophysical Research: Oceans 118. xxviii, 47, 48

Chang, C.-C. and C.-J. Lin, 2011. LIBSVM: A library for support vector machines. ACM Transactions on Intelligent Systems and Technology 2:27:1–27:27. 82

Chapelle, O., B. Schölkopf, and A. Zien, 2006. Semi-supervised learning. The MIT Press, Cambridge, Massachusetts. 93

Chavez, F., 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 79:95–105. xxxix, 3, 47, 51, 53, 60, 67, 116, 117

Chavez, F. P. and M. Messié, 2009. A comparison of Eastern Boundary Upwelling Ecosystems. Progress in Oceanography 83:80–96. xxviii, 43, 44, 45, 51

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 299:217–221. 57, 60

Chelton, D., editor, 2001. Report of the high-resolution ocean topography science working group meeting. Oregon State University, College of Oceanic and Atmospheric Sciences. xxvii, xxxv, xlviii, 180

Chessel, D. and M. Hanafi, 1996. Analyses de la co-inertie de K nuages de points. Revue de Statistique Appliquee XLIV:35–60. xliv, 123

Chilès, J.-P. and P. Delfiner, 2012. Geostatistics: modeling spatial uncertainty. John Wiley & Sons, Inc. xliv, 149

Ciannelli, L., P. Fauchald, K. S. Chan, V. N. Agostini, and G. E. Dingsør, 2008. Spatial fisheries ecology: Recent progress and future prospects. Journal of Marine Systems 71:223–236. 149

CIDEF, 2002. Proceso de privatización de Pesca Perú 1992-2001. Technical report, Congreso de la República del Perú. 61

Codling, E. A., M. J. Plank, and S. Benhamou, 2008. Random walk models in biology. Journal of The Royal Society Interface 5:813–834. 13, 15

Collins, C., A. Mascarenhas, and R. Martinez, 2013. Structure of ocean circulation between the Galápagos Islands and Ecuador. Advances in Geosciences 33:3–12. 47

Congdon, P., 2006. Bayesian statistical modelling. John Wiley & Sons, Ltd, second edi edition. 167

Costa, D. P., 1993. The secret life of marine mammals. Oceanography 6:120–128. 115

Costello, C., A. Rassweiler, D. Siegel, G. De Leo, F. Micheli, and A. Rosenberg, 2010. The value of spatial information in MPA network design. Proceedings of the National Academy of Sciences of the United States of America 107:18294–19299. 34

Crisci, C., B. Ghattas, and G. Perera, 2012. A review of supervised machine learning algorithms and their applications to ecological data. Ecological Modelling 240:113–122. 76

Croll, D., B. Marinovic, S. Benson, F. Chavez, N. Black, R. Ternullo, and B. Tershy, 2005. From wind to whales: trophic links in a coastal upwelling system. Marine Ecology Progress Series 289:117–130. 115

Cronk, L., 1991. Human behavioral ecology. Annual Review of Anthropology 20:25–53. xxxix, 2

Cury, P. and Y. Miserey, 2008. Une mer sans poissons. Calmann-Lévy. xxxix, 2

Cutler, D. R., T. C. Edwards, K. H. Beard, A. Cutler, K. T. Hess, J. Gibson, and J. J. Lawler, 2007. Random forests for classification in ecology. Ecology 88:2783–2792. 76

Czeschel, R., L. Stramma, F. U. Schwarzkopf, B. S. Giese, A. Funk, and J. Karstensen, 2011. Mid-depth circulation of the eastern tropical South Pacific and its link to the oxygen minimum zone. Journal of Geophysical Research 116:C01015. 47

Davies, N. B., J. R. Krebs, and S. A. West, 2012. An introduction to behavioural ecology. John Wiley & Sons, Ltd. 1

Dean, B., R. Freeman, H. Kirk, K. Leonard, R. A. Phillips, C. M. Perrins, and T. Guilford, 2012. Behavioural mapping of a pelagic seabird: combining multiple sensors and a hidden Markov model reveals the distribution of at-sea behaviour. Journal of the Royal Society, Interface / the Royal Society 20120570. 19, 74

Decker, C. J. and R. O'Dor, 2003. A census of marine life: unknowable or just unknown? Un recensement de la vie marine : impossible à connaître ou simplement inconnu ? Oceanologica Acta 25:179–186. 115

Delgado, E., S. Sanchez, F. Chang, P. Villanueva, and C. Fernandez, 2001. El fitoplancton frente a la costa Peruana durante El Niño 1997-98. In J. Tarazona, W. E. Arntz, and E. Castillo de Maruenda, editors, El Niño en América Latina: Impactos Biológicos y Sociales, pages 29–38. Consejo Nacional de Ciencia y Tecnología, Lima, Perú. 53

Dempster, A. P., N. M. Laird, and D. B. Rubin, 1977. Maximum likelihood from incomplete data via the EM algorithm. Journal of the Royal Statistical Society: Series B (Methodological) 39:1–38. 93

Deng, R., C. Dichmont, D. Milton, M. Haywood, D. Vance, N. Hall, and D. Die, 2005. Can vessel monitoring system data also be used to study trawling intensity and population depletion? The example of Australia's northern prawn fishery 62:611–622. 32

Deporte, N., C. Ulrich, S. Mahévas, S. Demaneche, and F. Bastardie, 2012. Regional métier definition: a comparative investigation of statistical methods using a workflow applied to international otter trawl fisheries in the North Sea. ICES Journal of Marine Science 69:331–342. 27, 101

Dewitte, B., J. Vazquez-Cuervo, K. Goubanova, S. Illig, K. Takahashi, G. Cambon, S. Purca, D. Correa, D. Gutierrez, A. Sifeddine, and L. Ortlieb, 2012. Change in El Niño flavours over 1958-2008: Implications for the long-term trend of the upwelling off Peru. Deep Sea Research Part II: Topical Studies in Oceanography 77-80:143–156. 57, 133

Di Lorenzo, E. and M. D. Ohman, 2013. A double-integration hypothesis to explain ocean ecosystem response to climate forcing. Proceedings of the National Academy of Sciences of the United States of America 110:2496–9. 134

Dickey, T. and M. Lewis, 2006. Optical oceanography: Recent advances and future directions using global remote sensing and in situ observations. Review of Geophysics 44:RG1001. xxvii, xxxv, xlviii, 180

Dietterich, T. G., 2002. Machine learning for sequential data: A review. In Structural, Syntactic, and Statistical Pattern Recognition, pages 15–30. Springer-Verlag. 79

Diggle, P. J., R. Menezes, and T.-L. Su, 2010. Geostatistical inference under preferential sampling. Applied Statistics 59:191–232. 167, 175

Dinmore, T. A., D. E. Duplisea, B. D. Rackham, D. L. Maxwell, and S. Jennings, 2003. Impact of a large-scale area closure on patterns of fishing disturbance and the consequences for benthic communities. ICES Journal of Marine Science 60:371–380. 34

Dolédec, S. and D. Chessel, 1994. Co-inertia analysis: an alternative method for studying species-environment relationships. Freshwater Biology 31:277–294. xliv, 123

Dong, M. and D. He, 2007. A segmental hidden semi-Markov model (HSMM)-based diagnostics and prognostics framework and methodology. Mechanical Systems and Signal Processing 21:2248–2266. 80

Dragon, A., A. Bar-Hen, P. Monestiez, and C. Guinet, 2012. Comparative analysis of methods for inferring successful foraging areas from Argos and GPS tracking data. Marine Ecology Progress Series 452:253–267. 20

Dray, S., D. Chessel, and J. Thioulouse, 2003. Co-inertia analysis and the linking of ecological data tables. Ecology 84:3078–3089. xliv, 123

Dray, S., A. Dufour, and D. Chessel, 2007. The ade4 package-II: Two-table and K-table methods. R News 7:47–52. 123

Eastwood, P. D., C. M. Mills, J. N. Aldridge, C. A. Houghton, and S. I. Rogers, 2007. Human activities in UK offshore waters: an assessment of direct, physical pressure on the seabed. ICES Journal of Marine Science 64:453–463. 33

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 79:167–176. 53, 59, 129, 135, 136

Echevin, V., O. Aumont, J. Tam, and J. Pasapera, 2004a. The seasonal cycle of surface Chlorophyll along the Peruvian coast: comparison between SeaWIFS satellite observations and dynamical/biogeochemical coupled model simulations. Gayana 68:325–326. 49

Echevin, V., I. Puillat, C. Grados, and B. Dewitte, 2004b. Seasonal and mesoscale variability in the Peru upwelling system from in situ data during the years 2000 to 2004. Gayana 68:167–173. 49

Edwards, A. M., R. A. Phillips, N. W. Watkins, M. P. Freeman, E. J. Murphy, V. Afanasyev, S. V. Buldyrev, M. G. E. da Luz, E. P. Raposo, H. E. Stanley, and G. M. Viswanathan, 2007. Revisiting Lévy flight search patterns of wandering albatrosses, bumblebees and deer. Nature 449:1044–8. 13, 14

Escoufier, Y., 1973. Le traitement des variables vectorielles. Biometrics 29:751–760. xliv, 122

Espinoza, P. and A. Bertrand, 2008. Revisiting Peruvian anchovy (Engraulis ringens) trophodynamics provides a new vision of the Humboldt Current system. Progress in Oceanography 79:215–227. 53, 54

Espinoza, P., A. Bertrand, C. D. van der Lingen, S. Garrido, and B. Rojas 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. 53

Estrella Arellano, C. and G. Swartzman, 2010. The Peruvian artisanal fishery: Changes in patterns and distribution over time. Fisheries Research 101:133–145. 63

Fablet, R., A. Chaigneau, and S. Bertrand, 2013. Multiscale analysis of geometric planar deformations: application to wild animals electronic tracking and satellite ocean observation data. IEEE Transactions on Geoscience and Remote Sensing . 20

FAO, 2013a. Fishing vessel monitoring systems (VMS). http://www.fao.org/fishery/vms/en. Accessed: 2013-10-03. 25

FAO, 2013b. Global capture production 1950-2011. http://www.fao.org/figis/servlet/TabSelector?tb_ds=Capture&tb_mode=TABLE&tb_act=SELECT&tb_grp=SPECIES. Accessed: 2013-10-03. 44, 54

Fauchald, P., K. E. Erikstad, and H. Skarsfjord, 2000. Scale-dependent predator-prey interactions : The hierarchical spatial distribution of seabirds and prey. Ecology 81:773–783. 143

Fauchald, P. and T. Tveraa, 2003. Using first-passage time in the analysis of area-restricted search and habitat selection. Ecology 84:282–288. 20

Fock, H. O., 2008. Fisheries in the context of marine spatial planning: Defining principal areas for fisheries in the German EEZ. Marine Policy 32:728–739. 36

Franke, A., T. Caelli, and R. J. Hudson, 2004. Analysis of movements and behavior of caribou *(Rangifer tarandus)* using hidden Markov models. Ecological Modelling 173:259–270. 21, 74

Franke, A., T. Caelli, G. Kuzyk, and R. J. Hudson, 2006. Prediction of wolf *(Canis lupus)* kill-sites using hidden Markov models. Ecological Modelling 197:237–246. 21, 74

Freeman, R., R. Mann, T. Guilford, and D. Biro, 2011. Group decisions and individual differences: route fidelity predicts flight leadership in homing pigeons *(Columba livia)*. Biology letters pages 63–66. 23

Fréon, P., M. Barange, and J. Arístegui, 2009. Eastern boundary upwelling ecosystems: integrative and comparative approaches. Progress in Oceanography 83:1–14. xxviii, 43, 44

Freón, P., M. Bouchon, C. Mullon, C. García, and M. Ñiquen, 2008. Interdecadal variability of anchoveta abundance and overcapacity of the fishery in Peru. Progress in Oceanography 79:401–412. xxix, 60, 61, 62, 63

Freón, P. and O. A. Misund, 1999. Dynamics of pelagic fish distribution and behaviour: effects on fishery and stock assessment. Fishing New Books, Oxford. 115

Fretwell, S. D. and H. L. Lucas, 1969. On territorial behavior and other factors influencing habitat distribution in birds. Acta Biotheoretica 19:16–36. 30, 143

Fritz, H., S. Said, and H. Weimerskirch, 2003. Scale-dependent hierarchical adjustments of movement patterns in a long-range foraging seabird. Proceeding of the Royal Society of London B 270:1143–1148. 20, 22

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 Research Part II: Topical Studies in Oceanography 56:1027–1038. xxix, 49, 51, 52

Gaertner, D., M. Pagavino, and J. Marcano, 1999. Influence of fishers' behaviour on the catchability of surface tuna schools in the Venezuelan purse-seiner fishery in the Caribbean Sea. Canadian Journal of Fisheries and Aquatic Sciences 56:394–406. 144, 176

Ganapathiraju, A., J. Hamaker, and J. Picone, 2000. Hybrid SVM/HMM architectures for speech recognition. ICSLP 4:504–507. 94, 249

Garcia, H. E., R. A. Locarnini, T. P. Boyer, and J. I. Antonov, 2006a. World Ocean Atlas 2005. In S. Levitus, editor, Nutrients (phosphate, nitrate, silicate), NOAA Atlas NESDIS 63, vol. 4, page 396. US Government Printing Office, Washington, DC. xxviii, 45

Garcia, H. E., R. A. Locarnini, T. P. Boyer, and J. I. Antonov, 2006b. World Ocean Atlas 2005. In S. Levitus, editor, Dissolved Oxygen, Apparent Oxygen Utilization, and Oxygen Saturation, NOAA Atlas NESDIS 63, vol. 3, page 342. US Government Printing Office, Washington, DC. xxviii, 45

Garcia, S. and K. Cochrane, 2005. Ecosystem approach to fisheries: a review of implementation guidelines. ICES Journal of Marine Science 62:311–318. 36, 97, 140

Gaucherel, C., 2011. Wavelet analysis to detect regime shifts in animal movement. Computational Ecology and Software 1:69–85. 20

Georgakarakos, S. and D. Kitsiou, 2008. Mapping abundance distribution of small pelagic species applying hydroacoustics and co-Kriging techniques. Hydrobiologia 612:155–169. 168

Gerlotto, F., S. Bertrand, N. Bez, and M. Gutierrez, 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 63:1405–1417. 54, 176

Gerritsen, H. D., C. Lordan, C. Minto, and S. B. M. Kraak, 2012. Spatial patterns in the retained catch composition of Irish demersal otter trawlers: High-resolution fisheries data as a management tool. Fisheries Research 129-130:127–136. 34

Gerritsen, H. D., C. Minto, and C. Lordan, 2013. How much of the seabed is impacted by mobile fishing gear? Absolute estimates from Vessel Monitoring System (VMS) point data. ICES Journal of Marine Science 70:523–531. 32

Gillis, D. M., 2003. Ideal free distributions in fleet dynamics: a behavioral perspective on vessel movement in fisheries analysis. Canadian Journal of Zoology 81:177–187. 144

Gillis, D. M., R. M. Peterman, and A. V. Tyler, 1993. Movement dynamics in a fishery: Application of the ideal free distribution to spatial allocation of effort. Canadian Journal of Fisheries and Aquatic Sciences 50:323–333. 144

Gimpel, K. and D. Rudoy, 2008. Statistical inference in graphical models. Technical report, Lincoln Laboratory, Massachusetts Institute of Technology, Lexington, Massachusetts. 75

Granadeiro, J. P., P. Brickle, and P. Catry, 2013. Do individual seabirds specialize in fisheries' waste? The case of black-browed albatrosses foraging over the Patagonian Shelf. Animal Conservation pages n/a–n/a. 36

Granadeiro, J. P., R. A. Phillips, P. Brickle, and P. Catry, 2011. Albatrosses following fishing vessels: how badly hooked are they on an easy meal? PloS one 6:e17467. 36

Greenman, J. V. and T. G. Benton, 2003. The amplification of environmental noise in population models: causes and consequences. The American Naturalist 161:225–239. 134

Gripenberg, S. and T. Roslin, 2007. Up or down in space? Uniting the bottom-up versus top-down paradigm and spatial ecology. Oikos 116:181–188. 115

Guédon, Y., 2003. Estimating hidden semi-Markov chains from discrete sequences. Journal of Computational and Graphical Statistics 12:604–639. xliii, 79, 231

Guédon, Y., 2007. Exploring the state sequence space for hidden Markov and semi-Markov chains. Computational Statistics & Data Analysis 51:2379–2409. 81, 235

Guilford, T., S. Roberts, D. Biro, and I. Rezek, 2004. Positional entropy during pigeon homing II: navigational interpretation of Bayesian latent state models. Journal of theoretical biology 227:25–38. 74

Gurarie, E., R. D. Andrews, and K. L. Laidre, 2009. A novel method for identifying behavioural changes in animal movement data. Ecology letters 12:395–408. 17

Gutiérrez, D., I. Bouloubassi, A. Sifeddine, S. Purca, K. Goubanova, M. Graco, D. Field, L. Méjanelle, F. Velazco, A. Lorre, R. Salvatteci, D. Quispe, G. Vargas, B. Dewitte, and L. Ortlieb, 2011. Coastal cooling and increased productivity in the main upwelling zone off Peru since the mid-twentieth century. Geophysical Research Letters 38:n/a–n/a. 129, 135, 136

Gutiérrez, D., E. Enríquez, S. Purca, L. Quipúzcoa, R. Marquina, G. Flores, and M. Graco, 2008. Oxygenation episodes on the continental shelf of central Peru: Remote forcing and benthic ecosystem response. Progress in Oceanography 79:177–189. 54, 129

Gutierrez, D., A. Sifeddine, D. B. Field, L. Ortlieb, G. Vargas, F. Chavez, F. Velazco, V. Ferreira, P. Tapia, R. Salvatecci, H. Boucher, M. Morales, J. Valdes, J.-L. Reyss, A. Campusano, M. Boussafir, M. Mandeng-Yogo, M. Garcia, and T. Baumgartner, 2009. Rapid reorganization in ocean biogeochemistry off Peru towards the end of the Little Ice Age. Biogeosciences 6:835–848. 57

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 16:155–168. xli, 55, 58, 68, 119, 135, 136, 146, 168, 178

Halley, J. M., S. Hartley, A. S. Kallimanis, W. E. Kunin, J. J. Lennon, and S. P. Sgardelis, 2004. Uses and abuses of fractal methodology in ecology. Ecology Letters 2004:254–271. 174

Hänggi, P. and F. Marchesoni, 2005. Introduction: 100 years of Brownian motion. Chaos 15:26101. 13

Hanks, E. M., M. B. Hooten, D. S. Johnson, and J. T. Sterling, 2011. Velocity-based movement modeling for individual and population level inference. PLoS ONE 6:e22795. 22

Haralick, R. M. and L. G. Shapiro, 1992. Computer and robot vision. Volume I. 164

Hart, T., R. Mann, T. Coulson, N. Pettorelli, and P. Trathan, 2010. Behavioural switching in a central place forager: patterns of diving behaviour in the macaroni penguin *(Eudyptes chrysolophus)*. Marine Biology 157:1543–1553. 20, 74

Hastie, T., R. Tibshirani, and J. Friedman, 2009. The elements of statistical learning: data mining, inference and prediction. Springer, second edition. xliii, 76, 81, 101

Hastings, A., 2010. Timescales, dynamics, and ecological understanding. Ecology 91:3471–3480. 174

Hazen, E. L., D. P. Nowacek, L. St Laurent, P. N. Halpin, and D. J. Moretti, 2011. The relationship among oceanography, prey fields, and beaked whale foraging habitat in the Tongue of the Ocean. PloS one 6:e19269. 116

He, X., L. Deng, and W. Chou, 2008. Discriminative learning in sequential pattern recognition. IEEE Signal Processing Magazine 14:14–36. 77

Helbing, D., I. Farkas, and T. Vicsek, 2000. Simulating dynamical features of escape panic. Nature 407:487–490. 24

Helbing, D., J. Keltsch, and P. Molnár, 1997. Modelling the evolution of human trail systems. Nature 388:1–11. 24

Helly, J. J. and L. A. Levin, 2004. Global distribution of naturally occurring marine hypoxia on continental margins. Deep Sea Research Part I: Oceanographic Research Papers 51:1159–1168. 51

Hengeveld, R., 1987. Scales of variation: their distinction and ecological importance. Ann. Zool. Fennici 24:195–202. 163, 174

Heo, M. and K. R. Gabriel, 1998. A permutation test of association between configurations by means of the RV coefficient. Communications in Statistics - Simulation and Computation 27:843–856. 122

Herbert-Read, J. E., A. Perna, R. P. Mann, T. M. Schaerf, D. J. T. Sumpter, and A. J. W. Ward, 2011. Inferring the rules of interaction of shoaling fish. Proceedings of the National Academy of Sciences 108:18726–18731. 24

Hernández, O., 2012. Modélisation eulérienne de l'habitat de ponte et de la dynamique des larves des anchois et sardines dans le système d'upwelling du Pérou. Ph.D. thesis, Université de Toulouse. 9

Heskes, T. and W. Wigerinck, 1996. A theoretical comparison of batch-mode, online, cyclic, and almost-cyclic learning. IEEE transactions on neural networks 7:919–25. 94

Hiddink, J. G., T. Hutton, S. Jennings, and M. Kaiser, 2006a. Predicting the effects of area closures and fishing effort restrictions on the production, biomass, and species richness of benthic invertebrate communities. ICES Journal of Marine Science 63:822–830. 32, 34

Hiddink, J. G., S. Jennings, and M. J. Kaiser, 2006b. Indicators of the ecological impact of bottom-trawl disturbance on seabed communities. Ecosystems 9:1190–1199. 32

Hiddink, J. G., S. Jennings, and M. J. Kaiser, 2007. Assessing and predicting the relative ecological impacts of disturbance on habitats with different sensitivities. Journal of Applied Ecology 44:405–413. 33

Hijmans, R. J., 2012. Cross-validation of species distribution models: removing spatial sorting bias and calibration with a null model. Ecology 93:679–688. 75

Hilborn, R. and C. J. Walters, 1992. Quantitative fisheries stock assessment choice, dynamics & uncertainty. Chapman & Hall. 60

Hintzen, N. T., G. J. Piet, and T. Brunel, 2010. Improved estimation of trawling tracks using cubic Hermite spline interpolation of position registration data. Fisheries Research 101:108–115. 10

Hinz, H., L. G. Murray, G. I. Lambert, J. G. Hiddink, and M. J. Kaiser, 2013. Confidentiality over fishing effort data threatens science and management progress. Fish and Fisheries 14:110–117. 141

Holmes, S. J., N. Bailey, N. Campbell, R. Catarino, K. Barratt, A. Gibb, and P. G. Fernandes, 2011. Using fishery-dependent data to inform the development and operation of a co-management initiative to reduce cod mortality and cut discards. ICES Journal of Marine Science 68:1679–1688. 34

Hooker, S., A. Cañadas, K. Hyrenbach, C. Corrigan, J. Polovina, and R. Reeves, 2011. Making protected area networks effective for marine top predators. Endangered Species Research 13:203–218. 141

Horta, S. and O. Defeo, 2012. The spatial dynamics of the whitemouth croaker artisanal fishery in Uruguay and interdependencies with the industrial fleet. Fisheries Research 125-126:121–128. 30

Humphries, N. E., N. Queiroz, J. R. M. Dyer, N. G. Pade, M. K. Musyl, K. M. Schaefer, D. W. Fuller, J. M. Brunnschweiler, T. K. Doyle, J. D. R. Houghton, G. C. Hays, C. S. Jones, L. R. Noble, V. J. Wearmouth, E. J. Southall, and D. W. Sims, 2010. Environmental context explains Lévy and Brownian movement patterns of marine predators. Nature 465:1066–1069. 13, 14, 23

Husson, F., J. Josse, S. Le, and J. Mazet, 2013. FactoMineR: multivariate exploratory data analysis and data mining with R. R package version 1.25. 102, 123

Husson, F., S. Lê, and J. Pagès, 2010. Exploratory multivariate analysis by example using R, volume 20105550 of Chapman & Hall/CRC Computer Science & Data Analysis. CRC Press. xxx, 101, 102, 105

IFFO, 2008. Fishmeal and fish oil statistical yearbook. 65

IMARPE, 2013. Resultados de los límites máximos por embarcación - anchoveta. región norte centro. http://www.imarpe.pe/imarpe/index.php?id_seccion=I013104020000000000000. 111

Jaiantilal, A., 2009. Classification and regression by randomforest-matlab. http://code.google.com/p/randomforest-matlab. 82

Jennings, S., 2005. Indicators to support an ecosystem approach to fisheries. Fish and Fisheries 6:212–232. 140

Jennings, S. and J. Lee, 2012. Defining fishing grounds with vessel monitoring system data. ICES Journal of Marine Science 69:51–63. 32

Jennings, S., J. Lee, and J. G. Hiddink, 2012. Assessing fishery footprints and the trade-offs between landings value, habitat sensitivity, and fishing impacts to inform marine spatial planning and an ecosystem approach. ICES Journal of Marine Science 69:1053–1063. 32, 35

Johnson, A. R., J. A. Wiens, B. T. Milne, and T. O. Crist, 1992. Animal movements and population dynamics in heterogeneous landscapes. Landscape Ecology 7:63–75. xliv, 74

Johnson, R. A. and D. W. Wichern, 2007. Applied multivariate statistical analysis. Pearson Prentice Hall. 101

Jones, J. B., 1992. Environmental impact of trawling on the seabed: A review. New Zealand Journal of Marine and Freshwater Research 26:59–67. 32

Jonsen, I., M. Basson, S. Bestley, M. Bravington, T. Patterson, M. Pedersen, R. Thomson, U. Thygesen, and S. Wotherspoon, 2013. State-space models for bio-loggers: A methodological road map. Deep-Sea Research Part II 88–89:34–46. 10, 91

Jonsen, I. D., R. A. Myers, and M. C. James, 2007. Identifying leatherback turtle foraging behaviour from satellite telemetry using a switching state-space model. Marine Ecology Progress Series 337:255–264. 20, 23, 74

Jonsen, I. D., R. A. Myers, and J. Mills Fleming, 2003. Meta-analysis of animal movement using state-space models. Ecology 84:3055–3063. xxvii, 11

Joo, R., A. Bertrand, M. Bouchon, M. Segura, A. Chaigneau, H. Demarcq, J. Tam, M. Simier, M. Gutierrez, D. Gutierrez, R. Fablet, and S. Bertrand, in review. Ecosystem scenarios shape fishing spatial behavior. The case of the anchovy fishery in the northern Humboldt Current system . 98, 145, 150, 163, 167, 168

Joo, R., S. Bertrand, A. Chaigneau, and M. Ñiquen, 2011. Optimization of an artificial neural network for identifying fishing set positions from VMS data: An example from the Peruvian anchovy purse seine fishery. Ecological Modelling 222:1048–1059. xli, 20, 28, 70, 76, 77, 82, 99, 117, 147, 164, 182, 221

Joo, R., S. Bertrand, J. Tam, and R. Fablet, 2013. Hidden Markov models: the best models for forager movements? PLOS ONE 8:e71246. 74, 117, 121, 146, 147, 167

Kaiser, M. J., 2005. Marine Ecology Processes, Systems and Impacts. Oxford University Press. xxix, 56

Klusch, M., S. Lodi, and G. Moro, 2003. Distributed clustering based on sampling local density estimates. In Proceedings of International Joint Conference on Artificial Intelligence (IJCAI 2003), pages 485–490, Mexico. 164

Kohavi, R. and F. Provost, 1998. Glossary and terms. Maching Learning 30:271–274. 83

Langrock, R., R. King, J. Matthiopoulos, L. Thomas, D. Fortin, and J. M. Morales, 2012. Flexible and practical modeling of animal telemetry data: hidden Markov models and extensions. Ecology 93:2336–2342. xl, 19, 74, 75, 76, 91

Lau, K.-K., S. Roberts, D. Biro, R. Freeman, J. Meade, and T. Guilford, 2006. An edge-detection approach to investigating pigeon navigation. Journal of theoretical biology 239:71–8. 74

Laws, E., 1997. El Niño and the Peruvian anchovy fishery. University Science Books, California. 60

Lee, J., A. B. South, and S. Jennings, 2010. Developing reliable, repeatable, and accessible methods to provide high-resolution estimates of fishing-effort distributions from vessel monitoring system (VMS) data. ICES Journal of Marine Science 67:1260–1271. 27, 28, 38

Legendre, P. and L. Legendre, 1998. Numerical ecology. Second edition. 10

Lester, J., T. Choudhury, N. Kern, G. Borriello, and B. Hannaford, 2005. A hybrid discriminative/generative approach for modeling human activities. In IJCAI, pages 766–772. 21

Lett, C. and V. Mirabet, 2008. Modelling the dynamics of animal groups in motion. South African Journal Of Science 104:192–198. 24

Levin, S. A., 1992. The problem of pattern and scale in ecology. Ecology 73(6):1943–1967. 8, 115

Levinson, S., 1986. Continuously variable duration hidden Markov models for automatic speech recognition. Computer Speech & Language 1:29 – 45. 80

Levitis, D. A., W. Z. Lidicker, and G. Freund, 2009. Behavioural biologists don't agree on what constitutes behaviour. Animal behaviour 78:103–110. 1

Linderman, M., 2012. Rclusterpp: Linkable C++ clustering. R package version 0.2.1. 102

Lukas, R., 1986. The termination of the Equatorial Undercurrent in the eastern Pacific. Progress in Oceanography 16. 47

MacCall, A. D., 1990. Dynamic geography of marine fish populations. Washington Sea Grant Program Seattle, Washington. xxxv, 177

Maiorano, L., V. Bartolino, F. Colloca, A. Abella, A. Belluscio, P. Carpentieri, A. Criscoli, G. J. Lasinio, A. Mannini, F. Pranovi, B. Reale, G. Relini, C. Viva, and G. D. Ardizzone, 2009. Systematic conservation planning in the Mediterranean: a flexible tool for the identification of no-take marine protected areas. ICES Journal of Marine Science 66:137–146. 35

Mandelbrot, B. B., 1977. The fractal geometry of nature. 10

Mandelbrot, B. B., 1979. Fractals: form, chance and dimension. 10

Mann, K. and J. Lazier, 2006. Dynamics of Marine Ecosystems. Biological-Physical Interactions in the Oceans. Backwell Publishing, third edition. 56, 115, 117, 134

Mann, R. P., J. Faria, D. J. T. Sumpter, and J. Krause, 2013a. The dynamics of audience applause. Journal of The Royal Society Interface 10:20130466. 24

Mann, R. P., A. Perna, D. Strömbom, R. Garnett, J. E. Herbert-Read, D. J. T. Sumpter, and A. J. W. Ward, 2013b. Multi-scale inference of interaction rules in animal groups using Bayesian model selection. PLoS Computational Biology 9:e1002961. 21, 24, 74

Marchant, B. P. and R. M. Lark, 2007. The Matérn variogram model : Implications for uncertainty propagation and sampling in geostatistical surveys. Geoderma 140:337–345. 149

Marrs, S. J., I. D. Tuck, R. J. A. Atkinson, T. D. I. Stevenson, and C. Hall, 2002. Position data loggers and logbooks as tools in fisheries research: results of a pilot study and some recommendations. Fisheries Research 58:109–117. 32

Martiskainen, P., M. Järvinen, J.-P. Skön, J. Tiirikainen, M. Kolehmainen, and J. Mononen, 2009. Cow behaviour pattern recognition using a three-dimensional accelerometer and support vector machines. Applied Animal Behaviour Science 119:32–38. 20

Matheron, G., 1978. Estimer et choisir: essai sur la pratique des probabilités. Ecole nationale supérieure des mines de Paris. 14

Matson, P. A. and M. D. Hunter, 1992. Special feature: The relative contributions to top-down and bottom-up forces in population and community ecology. Ecology 73:723. 115

McManus, M. A. and C. B. Woodson, 2012. Plankton distribution and ocean dispersal. The Journal of experimental biology 215:1008–1016. 134

McSherry, F. and M. Najork, 2008. Computing information retrieval performance measures efficiently in the presence of tied scores. In C. Macdonald, I. Ounis, V. Plachouras, I. Ruthven, and R. White, editors, Advances in Information Retrieval, volume 4956 of Lecture Notes in Computer Science, pages 414–421. Springer Berlin / Heidelberg. 83

Meyer, D., F. Leisch, and K. Hornik, 2003. The support vector machine under test. Neurocomputing 55:169–186. xliii, 81

Miramontes, O. and P. Rohani, 1998. Intrinsically generated coloured noise in laboratory insect populations. Proceedings of the Royal Society of London, Series B 265:785–792. 134

Mjolsness, E. and D. DeCoste, 2001. Machine learning for science: state of the art and future prospects. Science (New York, N.Y.) 293:2051–5. 77

Montecino, V. and C. B. Lange, 2009. The Humboldt Current System: Ecosystem components and processes, fisheries, and sediment studies. Progress In Oceanography 83:65–79. 44, 112

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:C09003. 47

Morales, J. M., D. Fortin, J. L. Frair, and E. H. Merrill, 2005. Adaptive models for large herbivore movements in heterogeneous landscapes. Landscape Ecology 20:301–316. 20, 76

Morales, J. M., D. T. Haydon, J. Frair, K. E. Holsinger, and J. M. Fryxell, 2004. Extracting more out of relocation data: building movement models as mixures of random walks. Ecology 85:2436–2445. 22

Morales, J. M., P. R. Moorcroft, J. Matthiopoulos, J. L. Frair, J. G. Kie, R. A. Powell, E. H. Merrill, and D. T. Haydon, 2010. Building the bridge between animal movement and population dynamics. Philosophical transactions of the Royal Society of London. Series B, Biological sciences 365:2289–301. 9

Mountrakis, G., J. Im, and C. Ogole, 2011. Support vector machines in remote sensing: A review. ISPRS Journal of Photogrammetry and Remote Sensing 66:247–259. 76

Mullowney, D. R. and E. Dawe, 2009. Development of performance indices for the Newfoundland and Labrador snow crab *(Chionoecetes opilio)* fishery using data from a vessel monitoring system. Fisheries Research 100:248–254. 33

Murawski, S., S. Wigley, M. Fogarty, P. Rago, and D. Mountain, 2005. Effort distribution and catch patterns adjacent to temperate MPAs. ICES Journal of Marine Science 62:1150–1167. 35

Murphy, K., 1998. Hidden Markov model toolbox for Matlab. `http://www.ai.mit.edu/~murphyk/Software/hmm.html/`. 79

Nakada, S., N. Hirose, T. Senjyu, K.-I. Fukudome, T. Tsuji, and N. Okei, 2013. Operational ocean prediction experiments for smart coastal fishing. Progress In Oceanography . 176

Nallapati, R., 2004. Discriminative models for information retrieval. In Proceedings of the 27th annual international ACM SIGIR conference on Research and development in information retrieval, SIGIR '04, pages 64–71, New York, NY, USA. ACM. 77

Nams, V., 1996. The VFractal : a new estimator for fractal dimension of animal movement paths. Landscape Ecology 11:289–297. 74

Nathan, R., W. M. Getz, E. Revilla, M. Holyoak, R. Kadmon, D. Saltz, and P. E. Smouse, 2008. A movement ecology paradigm for unifying organismal movement research. PNAS 105:19052–19059. xxvii, xxviii, xxxix, 7, 8, 9, 19, 74

Needle, C. L. and R. Catarino, 2011. Evaluating the effect of real-time closures on cod targeting. ICES Journal of Marine Science 68:1647–1655. 34

Nettle, D., M. A. Gibson, D. W. Lawson, and R. Sear, 2013. Human behavioral ecology: current research and future prospects. Behavioral Ecology 24:1031–1040. 1

O'Connor, S., R. Ono, and C. Clarkson, 2011. Pelagic fishing at 42,000 years before the present and the maritime skills of modern humans. Science 334:1117–1121. 2

Olden, J. D., J. J. Lawler, and N. L. Poff, 2008. Machine learning methods without tears: a primer for ecologists. The Quarterly Review of Biology 83:171–193. 76

Palmer, M. C., 2008. Calculation of distance traveled by fishing vessels using GPS positional data: A theoretical evaluation of the sources of error. Fisheries Research 89:57–64. 10, 26, 39, 182

Palmer, M. C. and S. E. Wigley, 2009. Using positional data from vessel monitoring systems to validate the logbook-reported area fished and the stock allocation of commercial fisheries landings. North American Journal of Fisheries Management 29:928–942. 27

Pardo-Iguzquiza, E. and M. Chica-Olmo, 2008. Geostatistics with the Matern semivariogram model: A library of computer programs for inference, kriging and simulation. Computers and Geosciences 34:1073–1079. 149

Passuni, G., unpublished. Análisis del comportamiento de barcos de pesca. 164, 168

Patlak, C. S., 1953. Random walk with persistence and external bias. The Bulletin of Mathematical Biophysics 15:311–338. 13

Patterson, T. A., M. Basson, M. V. Bravington, and J. S. Gunn, 2009. Classifying movement behaviour in relation to environmental conditions using hidden Markov models. The Journal of animal ecology 78:1113–1123. 20, 74

Patterson, T. A., L. Thomas, C. Wilcox, O. Ovaskainen, and J. Matthiopoulos, 2008. State-space models of individual animal movement. Trends in Ecology and Evolution 23:87–94. 74

Paulmier, A., D. Ruiz-Pino, and V. Garcon, 2008. The oxygen minimum zone (OMZ) off Chile as intense source of CO2 and N2O. Continental Shelf Research 28:2746–2756. 49

Pearson, K., 1901. On lines and planes of closest fit to systems of points in space. Philosophical Magazine 2:559–572. xliii, 100, 122

Pebesma, E., 2004. Multivariable geostatistics in S: the gstat package. Computers and Geosciences 30:683–691. 150

Pedersen, M. W., T. A. Patterson, U. H. Thygesen, and H. Madsen, 2011. Estimating animal behavior and residency from movement data. Oikos 120:1281–1290. 20, 74

Pedersen, S. A., H. O. Fock, and A. F. Sell, 2009. Mapping fisheries in the German exclusive economic zone with special reference to offshore Natura 2000 sites. Marine Policy 33:571–590. 36

Peel, D. and N. M. Good, 2011. A hidden Markov model approach for determining vessel activity from vessel monitoring system data. Canadian Journal of Fisheries and Aquatic Sciences 68:1252–1264. 21, 28, 74

Pennino, M. G., 2013. Implementing Ecosystem Approach to Fisheries Management: Advances and New tools. Ph.D. thesis, Universitat de València. 167, 175

Peraltilla, S. and S. Bertrand, 2013. In situ measurements of the speed of Peruvian anchovy schools. Fisheries Research . 116, 144, 164

Perry, J. N., A. M. Liebhold, M. S. Rosenberg, J. Dungan, M. Miriti, A. Jakomulska, and S. Citron-Pousty, 2002. Illustrations and guidelines for selecting statistical methods for quantifying spatial pattern in ecological data. Ecography 25:578–600. 10

Petchey, O. L., 2000. Environmental colour affects aspects of single-species population dynamics. Proceedings of the Royal Society of London, Series B 267:747–754. 134

Petitgas, P., 2001. Geostatistics in fisheries survey design and stock assessment: models, variances and applications. Fish and Fisheries 2:231–249. 149

Pichegru, L., P. Ryan, C. van der Lingen, J. Coetzee, Y. Ropert-Coudert, and D. Grémillet, 2007. Foraging behaviour and energetics of Cape gannets Morus capensis feeding on live prey and fishery discards in the Benguela upwelling system. Marine Ecology Progress Series 350:127–136. 22

Pickands, J. I., 1975. Statistical inference using extreme order statistics. The Annals of Statistics 3:119–131. 14

Pierce, G. J. and J. G. Ollason, 1987. Eight reasons why optimal foraging theory is a complete waste of time. Oikos 49:111–118. 144

Piet, G., A. D. Rijnsdorp, J. N. Bergman, J. W. van Santbrink, J. Craeymeersch, and J. Buijs, 2000. A quantitative evaluation of the impact of beam trawling on benthic fauna in the southern North Sea. ICES Journal of Marine Science 57:1332–1339. 32

Piet, G. J. and N. T. Hintzen, 2012. Indicators of fishing pressure and seafloor integrity. ICES Journal of Marine Science 69:1850–1858. 32

Piet, G. J. and F. J. Quirijns, 2009. The importance of scale for fishing impact estimations. Canadian Journal of Fisheries and Aquatic Sciences 66:829–835. 32

Plank, M. J. and E. A. Codling, 2009. Sampling rate and misidentification of Lévy and non-Lévy movement paths. Ecology 90:3546–3553. 13

Plank, M. J. and A. James, 2008. Optimal foraging: Lévy pattern or process? Journal of the Royal Society, Interface / the Royal Society 5:1077–86. 14

Polansky, L. and G. Wittemyer, 2011. A framework for understanding the architecture of collective movements using pairwise analyses of animal movement data. Journal of the Royal Society, Interface / the Royal Society 8:322–33. 23

Polansky, L., G. Wittemyer, P. C. Cross, C. J. Tambling, and W. M. Getz, 2010. From moonlight to movement and synchronized randomness: Fourier and wavelet analyses of animal location time series data. Ecology 91:1506–18. 20

Polishchuk, L. V., J. Vijverberg, D. A. Voronov, and W. M. Mooij, 2013. How to measure top-down vs bottom-up effects: a new population metric and its calibration on Daphnia. Oikos 122:1177–1186. 115

Poos, J. J., J. a. Bogaards, F. J. Quirijns, D. M. Gillis, and a. D. Rijnsdorp, 2010a. Individual quotas, fishing effort allocation, and over-quota discarding in mixed fisheries. ICES Journal of Marine Science 67:323–333. 40

Poos, J.-J., F. J. Quirijns, and A. D. Rijnsdorp, 2010b. Spatial segregation among fishing vessels in a multispecies fishery. ICES Journal of Marine Science 67:155–164. 30, 144

Poos, J.-J. and A. D. Rijnsdorp, 2007. An experiment on effort allocation of fishing vessels : the role of interference competition and area specialization. Canadian Journal of Fisheries and Aquatic Sciences 64:304–313. 30, 40, 144

R Core Team, 2013. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 102, 123, 150

Rabiner, L. R., 1989. A tutorial on hidden Markov models and selected applications in speech recognition. Proceedings of the IEEE 77:257–286. xlii, 75, 79, 227, 229

Reynolds, A. M., 2012. Distinguishing between Lévy walks and strong alternative models. Ecology 93:1228–1233. 14

Reynolds, C. W., 1987. Flocks, herds and schools: a distributed behavioral model. In Proceedings of the 14th annual Conference on Computer Graphics and Interactive Techniques, pages 25–34, Anaheim, California. ACM Press, New York. 23, 181

Richard, M. D. and R. P. Lippmann, 1991. Neural network classifiers estimate Bayesian a posteriori probabilities. Neural Computation 3:461–483. 247

Ridley, M., 1995. Animal Behavior. Blackwell Sciencce Ltd, second edition. 2

Rijnsdorp, A. D., W. Dol, M. Hoyer, and M. A. Pastoors, 2000. Effects of fishing power and competitive interactions among vessels on the effort allocation on the trip level of the Dutch beam trawl fleet. ICES Journal of Marine Science 57:927–937. 144

Rijnsdorp, A. D., J. J. Poos, and F. J. Quirijns, 2011. Spatial dimension and exploitation dynamics of local fishing grounds by fishers targeting several flatfish species. Canadian Journal of Fisheries and Aquatic Sciences 68:1064–1076. 30

Rivoirard, J., J. Simmonds, K. G. Foote, P. Fernandes, and N. Bez, 2000. Geostatistics for Estimating Fish Abundance. Blackwell Science Ltd. xliv, 150

Roberts, S., T. Guilford, I. Rezek, and D. Biro, 2004. Positional entropy during pigeon homing I: application of Bayesian latent state modelling. Journal of theoretical biology 227:39–50. 19, 74

Rose, G. A. and W. C. Leggett, 1990. The importance of scale to predator-prey spatial correlations: An example of Atlantic fishes. Ecology 71:33–43. 143

Rose, K. A., J. I. Allen, Y. Artioli, M. Barange, J. Blackford, F. Carlotti, R. Cropp, U. Daewel, K. Edwards, K. Flynn, S. L. Hill, R. HilleRisLambers, G. Huse, S. Mackinson, B. Megrey, A. Moll, R. Rivkin, B. Salihoglu, C. Schrum, L. Shannon, Y.-J. Shin, S. L. Smith, C. Smith, C. Solidoro, M. St. John, and M. Zhou, 2010. End-to-end models for the analysis of marine ecosystems: Challenges, issues, and next steps. Marine and Coastal Fisheries 2:115–130. 183

Rousseeuw, P., C. Croux, V. Todorov, A. Ruckstuhl, M. Salibian-Barrera, T. Verbeke, M. Koller, and M. Maechler, 2012. robustbase: Basic robust statistics. http://CRAN.R-project.org/package=robustbase. R package version 0.9-7. 123

Rousseeuw, P. J., 1984. Least median of squares regression. Journal of the American Statistical Association 79:871–880. 122

Rouyer, T., A. Sadykov, J. Ohlberger, and N. C. Stenseth, 2012. Does increasing mortality change the response of fish populations to environmental fluctuations? Ecology letters 15:658–65. 134

Russo, T., A. Parisi, and S. Cataudella, 2011a. New insights in interpolating fishing tracks from VMS data for different métiers. Fisheries Research 108:184–194. 10

Russo, T., A. Parisi, M. Prorgi, F. Boccoli, I. Cignini, M. Tordoni, and S. Cataudella, 2011b. When behaviour reveals activity: Assigning fishing effort to métiers based on VMS data using artificial neural networks. Fisheries Research 111:53–64. 26, 39, 97

Rutz, C. and G. C. Hays, 2009. New frontiers in biologging science. Biology letters 5:289–292. 115

Ryan, P. G., S. L. Petersen, G. Peters, and D. Grémillet, 2004. GPS tracking a marine predator: the effects of precision , resolution and sampling rate on foraging tracks of African Penguins. Marine Biology 145:215–223. 10

Sacks, J. and D. Ylvisaker, 2012. After 50+ years in statistics, an exchange. Statistical Science 27:308–318. 91, 182

Salas, S. and D. Gaertner, 2004. The behavioural dynamics of fishers : management implications. Fish and Fisheries 5:153–167. 25

Salvatteci, R., 2012. Multi-decadal to millennial scale variability in OMZ intensity, export production and pelagic fish abundances from marine laminated sediments off Pisco, Peru during the last 25 000 years. Ph.D. thesis, Université Pierre et Marie Curie. 56

Salvatteci, R., D. B. Field, T. Baumgartner, V. Ferreira, and D. Gutierrez, 2012. Evaluating fish scale preservation in sediment records from the oxygen minimum zone off Peru. Paleobiology 38:52–78. 56

Sánchez Ramirez, S., 2000. Variación estacional e interanual de la biomasa fitoplantónica y concentración de clorofila A frente a la costa peruana durante 1976-2000. Boletín Instituto del Mar del Perú 19:29–43. 53

Santander, H., 1981. The zooplankton in an upwelling area off Peru. In F. A. Richards, editor, Coastal Upwelling Coastal and Estuarine Sciences, vol. 1. American Geophysical Union, Washington, DC. 53

Santora, J. A., J. C. Field, I. D. Schroeder, K. M. Sakuma, B. K. Wells, and W. J. Sydeman, 2012. Spatial ecology of krill, micronekton and top predators in the central California Current: Implications for defining ecologically important areas. Progress in Oceanography 106:154–174. 116

Schwarz, G., 1978. Estimating the dimension of a model. The Annals of Statistics 6:461–464. 79

Sharples, J., J. R. Ellis, G. Nolan, and B. E. Scott, 2013a. Fishing and the oceanography of a stratified shelf sea. Progress in Oceanography . 29, 38

Sharples, J., B. E. Scott, and M. E. Inall, 2013b. From physics to fishing over a shelf sea bank. Progress in Oceanography . 29, 38

Shephard, S., H. Gerritsen, M. J. Kaiser, and D. G. Reid, 2012. Spatial heterogeneity in fishing creates de facto refugia for endangered Celtic Sea elasmobranchs. PloS one 7:e49307. 35

Siegel, S., 1956. Nonparametric statistics for the behavioral sciences. McGraw-Hill, New York. 84

Silva, N. and S. Neshyba, 1979. On the southernmost extension of the Peru-Chile Undercurrent. Deep Sea Research Part A. Oceanographic Research Papers 26:1387–1393. 47

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 Journal of Marine Science 66:1341–1348. xli, 68, 119, 146

Simmonds, J. and D. MacLennan, 2005. Fisheries Acoustics: Theory and Practice. Blackwell Science Ltd, second edition. xli, 68, 119, 145, 146

Skaar, K. L., T. Jorgensen, B. K. H. Ulvestad, and A. Engas, 2011. Accuracy of VMS data from Norwegian demersal stern trawlers for estimating trawled areas in the Barents Sea. ICES Journal of Marine Science 68:1615–1620. 32

Smith, A. D., 2001. Vessel monitoring systems. In FAO, editor, Workshop on fisheries monitoring, control and surveillance in support of fisheries management., Goa, India. xxviii, 26

Smith, E. A., 1983. Anthropological applications of optimal foraging theory: A critical review. Current Anthropology 24:625–651. 2

Smith, M. D. and J. E. Wilen, 2003. Economic impacts of marine reserves: the importance of spatial behavior. Journal of Environmental Economics and Management 46:183–206. 25

Sonntag, N., H. Schwemmer, H. O. Fock, J. Bellebaum, and S. Garthe, 2012. Seabirds, set-nets, and conservation management: assessment of conflict potential and vulnerability of birds to bycatch in gillnets. ICES Journal of Marine Science 69:578–589. 36

Stadermann, J. and G. Rigoll, 2004. A hybrid SVM/HMM acoustic modeling approach to automatic speech recognition. In ICSLP - INTERSPEECH, Jeju Island, Korea. 94

Stelzenmuller, V., S. I. Rogers, and C. M. Mills, 2008. Spatio-temporal patterns of fishing pressure on UK marine landscapes, and their implications for spatial planning and management. ICES Journal of Marine Science 65:1081–1091. 32

Stouten, H., A. Heene, X. Gellynck, and H. Polet, 2011. Strategic groups in the Belgian fishing fleet. Fisheries Research 108:121–132. 99

Stramma, L., G. C. Johnson, E. Firing, and S. Schmidtko, 2010. Eastern Pacific oxygen minimum zones: Supply paths and multidecadal changes. Journal of Geophysical Research 115:C09011. 49

Sumpter, D. J. T., R. P. Mann, and A. Perna, 2012. The modelling cycle for collective animal behaviour. Interface focus 2:764–773. 24, 181

Swain, D. P. and E. J. Wade, 2003. Spatial distribution of catch and effort in a fishery for snow crab *(Chionoecetes opilio)*: tests of predictions of the ideal free distribution. Canadian Journal of Fisheries and Aquatic Sciences 60:897–909. 144

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 79:228–237. 57, 58, 107, 116, 135, 136, 137, 144, 168, 178

Takahashi, K., A. Montecinos, K. Goubanova, and B. Dewitte, 2011. ENSO regimes: Reinterpreting the canonical and Modoki El Niño. Geophysical Research Letters 38:n/a–n/a. 133

Tan, C. O., U. Özesmi, M. Beklioglu, E. Per, and B. Kurt, 2006. Predictive models in ecology: Comparison of performances and assessment of applicability. Ecological Informatics 1:195–211. 75

Tew Kai, E., S. Benhamou, C. D. van der Lingen, J. C. Coetzee, L. Pichegru, P. G. Ryan, and D. Grémillet, 2013. Are Cape gannets dependent upon fishery waste? A multi-scale analysis using seabird GPS-tracking, hydro-acoustic surveys of pelagic fish and vessel monitoring systems. Journal of Applied Ecology 50:659–670. 10, 37

Thiebault, A. and Y. Tremblay, 2013. Splitting animal trajectories into fine-scale behaviorally consistent movement units: breaking points relate to external stimuli in a foraging seabird. Behavioral Ecology and Sociobiology 67:1013–1026. 17

Thomas, A. C., M. E. Carr, and P. T. Strub, 2001. Chlorophyll variability in eastern boundary currents. Geophysical Research Letters 28:3421–3424. 59

Tillin, H., J. Hiddink, S. Jennings, and M. Kaiser, 2006. Chronic bottom trawling alters the functional composition of benthic invertebrate communities on a sea-basin scale. Marine Ecology Progress Series 318:31–45. 33

Tinbergen, N., 1951. The study of instinct. Clarendon Press/Oxford University Press. 2

Torrence, C. and G. P. Compo, 1998. A practical guide to wavelet analysis. Bulletin of the American Meteorological Society 79:61–78. 163

Torres, L. G., D. R. Thompson, S. Bearhop, S. Votier, G. A. Taylor, P. M. Sagar, and B. C. Robertson, 2011. White-capped albatrosses alter fine-scale foraging behavior patterns when associated with fishing vessels. Marine Ecology Progress Series 428:289–301. 36

Travis, J., 2007. Do wandering albatrosses care about math? Science 318. 13, 14

Tremblay, Y., S. A. Shaffer, S. L. Fowler, C. E. Kuhn, B. I. McDonald, M. J. Weise, C.-A. Bost, H. Weimerskirch, D. E. Crocker, M. E. Goebel, and D. P. Costa, 2006. Interpolation of animal tracking data in a fluid environment. The Journal of experimental biology 209:128–140. xxvii, 10, 12

Tsuchiya, M. and L. D. Talley, 1998. A Pacific hydrographic section at 88°W: Water-property distribution. Journal of Geophysical Research 103:12899–12918. 47

Turchin, P., 1998. Quantitative analysis of movement: measuring and modeling population redistribution in plants and animals. Sinauer Associates, Sunderland, MA. xxvii, 8, 10, 11, 17, 27

Tveteras, S., C. E. Paredes, and J. Peña Torres, 2011. Individual vessel quotas in Peru: Stopping the race for anchovies. Marine Resource Economics 26:225–232. 62, 63

Valdes, J., L. Ortlieb, D. Gutierrez, 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 79:198–207. 57

Vázquez-Rowe, I. and P. Tyedmers, 2013. Identifying the importance of the skipper effect within sources of measured inefficiency in fisheries through data envelopment analysis (DEA). Marine Policy 38:387–396. 144

Vermard, Y., E. Rivot, S. Mahévas, P. Marchal, and D. Gascuel, 2010. Identifying fishing trip behaviour and estimating fishing effort from VMS data using Bayesian hidden Markov models. Ecological Modelling 221:1757–1769. 21, 28, 74

Vicsek, T., A. Czirók, E. Ben-Jacob, I. Cohen, and O. Shochet, 1995. Novel type of phase transition in a system of self-driven particles. Physical Review Letters 75:1226–1229. 23

Villanueva, R., 1972. El programa peruano Eureka. Mimeo. 69

Vinther, M. and M. Eero, 2013. Quantifying relative fishing impact on fish populations based on spatio-temporal overlap of fishing effort and stock density. ICES Journal of Marine Science 70:618–627. 32, 33

Votier, S. C., S. Bearhop, M. J. Witt, R. Inger, D. Thompson, and J. Newton, 2010. Individual responses of seabirds to commercial fisheries revealed using GPS tracking, stable isotopes and vessel monitoring systems. Journal of Applied Ecology 47:487–497. 37

Walker, E., 2010. De la trajectoire des predateurs a la cartographie de leurs proies: Estimation spatiale de l'activité des senneurs et des thonidés dans l'Océan Indien. Ph.D. thesis, Ecole nationale superieure des mines de Paris. 148

Walker, E. and N. Bez, 2010. A pioneer validation of a state-space model of vessel trajectories (VMS) with observers' data. Ecological Modelling 221:2008–2017. 21, 28, 74, 75, 144, 182

Walker, E., J. Rivoirard, P. Gaspar, and N. Bez, submitted. From foragers' tracks to prey's distributions . 144, 166

Warner, B. and M. Misra, 1996. Understanding neural networks as statistical tools. The American Statistician 50:284–293. xliii, 81

Weimerskirch, H., S. Bertrand, J. Silva, C. Bost, and S. Peraltilla, 2012. Foraging in Guanay cormorant and Peruvian booby, the major guano-producing seabirds in the Humboldt Current System. Marine Ecology Progress Series 458:231–245. 55

Whitehead, H. and I. D. Jonsen, 2013. Inferring animal densities from tracking data using Markov chains. PloS one 8:e60901. 90

Wilcoxon, F., 1945. Individual comparisons by ranking methods. Biometrics Bulletin 1:80–83. 121

Wilen, J. E., 2004. Spatial management of fisheries. Marine Resource Economics 19:7–19. 25, 97, 140

Willems, N., H. van de Wetering, and J. J. van Wijk, 2009. Visualization of vessel movements. Computer Graphics Forum 28:959–966. 31

Wilson, R., E. Shepard, and L. N, 2008. Prying into the intimate details of animal lives: use of a daily diary on animals. Endang Species Res 4:123–137. 10

With, K. A., 1994. Using fractal analysis to assess how species perceive landscape structure. Landscape Ecology 9:25–36. 74

Witt, M. J. and B. J. Godley, 2007. A step towards seascape scale conservation: using vessel monitoring systems (VMS) to map fishing activity. PloS one 2:e1111. 28

Wittemyer, G., L. Polansky, I. Douglas-Hamilton, and W. M. Getz, 2008. Disentangling the effects of forage, social rank, and risk on movement autocorrelation of elephants using Fourier and wavelet analyses. Proceedings of the National Academy of Sciences 105:19108–19113. 23

Yu, S.-Z., 2010. Hidden semi-Markov models. Artificial Intelligence 174:215–243. 80

Zavalaga, C. B., G. Dell'Omo, P. Becciu, and K. Yoda, 2011. Patterns of GPS tracks suggest nocturnal foraging by incubating Peruvian pelicans *(Pelecanus thagus)*. PloS one 6:e19966. 55

Zhang, G., 2000. Neural networks for classification: A survey. IEEE Transactions on Systems, Man and Cybernetics - Part C: Applications and Reviews 30:451–462. 76

Appendices

Appendix A

Pre-processing of vessel monitoring system and on-board observers data

A.1 Vessel monitoring system data

For this work, we used the complete VMS data from the industrial Peruvian anchovy fishing fleet from 2000 to 2009. The entire steel fleet was covered by the VMS since 2000, while the coverage of the wooden fleet was done progressively over the years. VMS data is provided on a $\sim$ 1 hour basis, although some irregularities seldom occur (Fig. A.1). For each VMS record, the following information is provided: position, date-time (i.e. date and time), name of the vessel, ID (unique code) of the vessel, and an indicator of accuracy of the position. Over the years, several companies have provided the VMS service to the fishing companies. They all work with a global positioning system (GPS), associated with $\pm$100 m of accuracy. CLS, which at the beginning of the 2000s provided the VMS service to the whole fishing fleet and nowadays covers $\sim$ 30%, also uses an Argos system when the position cannot be estimated by GPS; they are associated with 5 accuracy levels (ranging from 150 m to 1000 m).

Pre-processing of VMS data are performed based on the criteria and algorithms described in Bertrand *et al.* (2007, 2005) and Joo *et al.* (2011). We identify positions at sea or at port based on their distance to the nearest port and their speed. We then consider each set of consecutive positions at sea as a trip. To exclude non-fishing trips, we only retain trips with a minimum speed lower than 3 knots (indicating possible fishing activity). To exclude the few trips targeting other pelagic

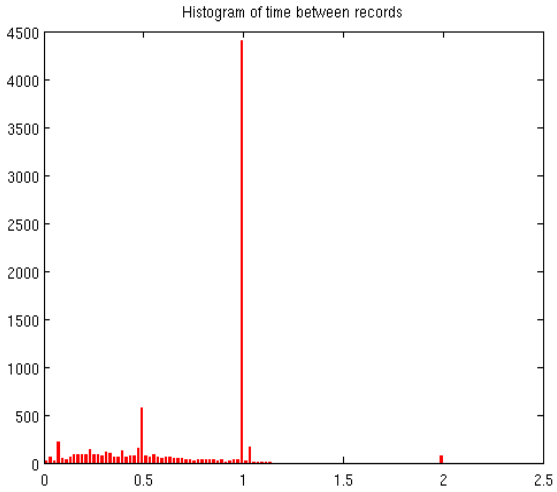


Figure A.1: Histogram of time between records (in hours) for 2008 VMS dataset (after pre-processing).

species (e.g. jack mackerel), trips from vessels with fishing authorization for multiple species that lasted more than 5 days are not considered (anchovy fishing trips typically last $\sim$ 24 hours). Finally, to exclude unreliable fishing trajectories, trips containing consecutive emissions separated by more than two hours are not kept in the dataset. Indeed, within an interval of more than two hours, we could be missing an important segment of the trajectory; even a fishing set could occur. The number of fishing trips per year kept in the VMS dataset is shown in Table A.1.

A.2 On-board observers data

IMARPE's program of on-board observers consists in $\sim$ 25 observers distributed along the coast of Peru (the number of observers varies each fishing season depending on the available funding). Each observer registers information related to the fishing trips, the fishing vessel and the crew of at least one vessel throughout the fishing season. Key information for short-term adaptive management, such as large proportions of juveniles in the catches, is reported immediately at the end of the fishing trip. However, other information, like the sequences of behavioral modes or activities associated with each fishing trip, are registered in the form of paper-written logbooks. Part of the information was already typed; for the remaining logbooks, three persons assisted us for completing that job. Annotation errors as well as typos could occur, so an algorithm was developed for recognizing possible mistakes, such

as inexistent or unlikely dates, or durations of behavioral modes suspiciously long or short, so that the paper-written sheet could be latter reviewed. If there was a typo, we corrected it. If the information on the sheet did not seem reliable, the trip was excluded from the dataset. The number of fishing trips per year kept in the logbook dataset is shown in Table A.1.

The behavioral modes registered by the on-board observers were: cruising (i.e. traveling following a predetermined course), searching for fish, fishing, drifting, helping other vessel, and receiving or offering catches to other vessel. Receiving/offering is very odd, and it can occur when a vessel has caught more fish than its holding capacity and calls a neighbor vessel to 'share the spoils'. The on-board observers registered the time at which each behavioral mode started and ended, the position at departure from port and arrival to port, and the position of each fishing set. This meant that we had a time-series of the behavioral modes within each fishing trip, but we could not make an accurate spatial representation of the trajectory followed (Fig. A.2).

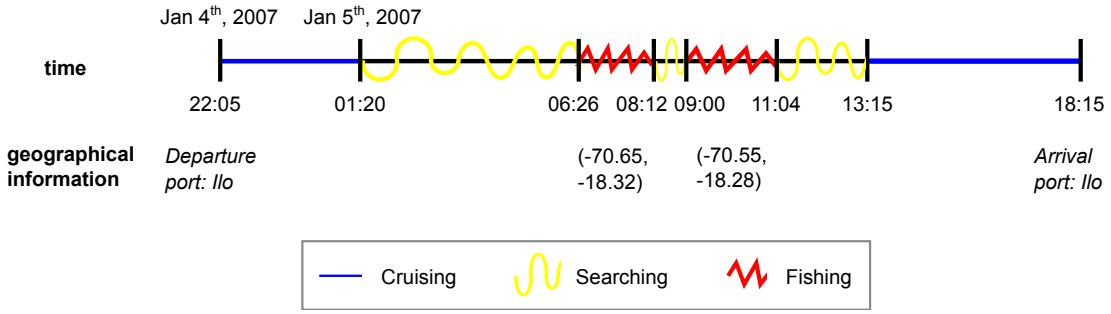


Figure A.2: Graphical example of fishing trip information recorded by on-board observers: time of start and end of each behavioral mode, ports of departure and arrival, and location (longitude, latitude) of each fishing set.

A.3 The groundtruthed dataset

In order to build a crossed dataset containing the activities information for each VMS record, two different algorithms were developed (if the reader is not interest in the description of both algorithms, he or she can go directly to the second algorithm, which we finally chose). The first one basically consists in (1) using the VMS records from reconstructing fishing trips (as in section A.1) and (2) crossing the VMS trips with its corresponding logbook trip. We now proceed to explain the second step. For crossing VMS and logbook trips, in both datasets, we search for

a trip done by the same vessel (i.e. with the same name, since VMS and logbook programs use different coding systems for the vessels) at about the same time (the time of the first and last VMS records are not expected to match exactly the time of departure and arrival to port, respectively). If the trip meets those conditions, we extract the date-time of the fishing sets registered in the logbooks and identify the VMS records corresponding to those date-time. We then compare the positions of those VMS records with the fishing set positions from the logbooks. If the distances between associated VMS record and fishing set position are less than 5 nm (arbitrary choice taking into account time mismatch and vessel speed), we keep the trip in the crossed dataset.

The second algorithm consists in (1) crossing the VMS records with logbook trips and (2) verifying if they are fishing trips (conditions from section A.1). We now proceed to explain the first step. For crossing VMS records and logbook trips, we search for records corresponding to the same vessel and between the date-time of departure and arrival to port of a logbook trip. If a sequence of more than 5 consecutive VMS records meets those conditions, we extract the date-time of the fishing sets registered in the logbooks and identify the VMS records corresponding to those date-time. We then compare the positions of those VMS records with the fishing set positions from the logbooks. If the distances between associated VMS record and fishing set position are less than 5 nm, and also distances between the first and last VMS records in respect to the departure and arrival positions of the logbook trip are less than 5 nm, we keep the trip in the crossed dataset. Then, if the trip meets the conditions from step 2, it remains in the dataset. This is a simplified version of the algorithm: other factors were taking into account when crossing both types of data, such as errors in the registration of time of depart and/or arrival by the on-board observer, or the possibility that the observers watch was 'out of sync' with the VMS satellite. Details on how we took into account this factors are not shown here since they did not produce any changes in the crossed dataset.

Both algorithms were applied on the VMS and logbook data from 2008. With the first method, a crossed dataset of $\sim$ 150 trips was obtained. However, with the second method, a dataset composed of $\sim$ 300 trips was obtained. We therefore chose the second algorithm for processing VMS and logbook data from all years. The number of crossed trips (obtained with the second algorithm) are shown in Table A.1. Compared to cruising, searching and fishing behavioral modes, a small number of records are associated with drifting, and almost no records are associated with

receiving/offering and helping. We consider that there is not enough information for correctly inferring those three behavioral modes, the final groundtruthed dataset will incorporated crossed fishing trips containing only fishing, cruising and searching modes. The number of fishing trips in the new groundtruthed dataset is also shown in Table A.1.

Table A.1: Fishing trips from VMS data, logbook data and the crossed data, and VMS records associated with each behavioral mode.

Dataset	Behavioral Mode	Year									
		2000	2001	2002	2003	2004	2005	2006	2007	2008	2009
VMS		28807	26176	44429	30764	43648	41518	31018	30606	36143	39602
Logbook		1056	764	1152	609	878	1163	790	899	934	540
Crossed		146	120	388	268	342	356	183	154	359	233
	C	*2403*	*1902*	*5783*	*5645*	*6740*	*6379*	*3275*	*2666*	*5207*	*4570*
	S	*780*	*809*	*3226*	*2578*	*3501*	*2527*	*1176*	*1165*	*2879*	*2249*
	F	*940*	*683*	*2717*	*1967*	*2839*	*2604*	*1294*	*941*	*2388*	*2018*
	D	*83*	*76*	*265*	*768*	*629*	*307*	*178*	*141*	*372*	*360*
	R/O	*0*	*0*	*0*	*0*	*0*	*3*	*6*	*0*	*22*	*27*
	H	*0*	*0*	*0*	*0*	*0*	*4*	*0*	*0*	*1*	*0*
Only S,F,C		132	109	356	193	265	309	155	127	269	182

Note: Numbers in regular style correspond to number of trips. In italic: number of records. H: Helping; S: Searching; F: Fishing; D: Drifting; R/O: Receiving/Offering; C: Cruising.

Appendix B

Notes on several Viterbi algorithms

In this appendix, we use the same notations as in section 4.2.

B.1 Model inversion

In many real applications, the observed values are known and the states corresponding to each one of them are not only hidden but also unknown. So the practical problem actually requires a model inversion: although observations are by definition state-dependent, we need to, given the observation sequence, find the corresponding hidden (and unknown) states. What we search for is an 'optimal' state sequence.

As Rabiner (1989) claims, there are several possible optimality criteria. One possible and very simple criterion is to choose the states S_t which are individually most likely. This optimality criterion maximizes the expected number of correct individual states. The chosen states at each time t in the sequence are the ones who maximize the probability

$$\gamma_t(j) = P(S_t = j \mid X_0^t = x_0^t) \tag{B.1}$$

of being in state j at time t given the observation sequence x_0^t. So $\hat{S}_t = \underset{0 \leqslant j \leqslant J}{\arg\max}\, \gamma_t(j)$, $\forall t$.

A potential problem with this criterion, is that, by determining the most likely state at every instant, we could end up with a sequence of states which is not likely

to happen. Therefore, instead of this criterion, we could choose the state sequence whose probability of occurrence (as a sequence) is maximal. The chosen states would thus be the ones to maximize the probability

$$P^* = P(S_0^T = s_0^T \mid X_0^T = x_0^T) \tag{B.2}$$

which is equivalent to maximizing

$$P(S_0^T = s_0^T, X_0^T = x_0^T) \tag{B.3}$$

A formal and efficient technique for finding this single best state sequence is the Viterbi algorithm. In the following section, we present the main features of the Viterbi algorithm for hidden Markov models (HMMs), then we will introduce the Viterbi algorithm for hidden semi-Markov models (HSMMs) and some other extensions (also for HSMMs). Some of the sections will have fonts of smaller sizes in order to be able to show the large equations.

B.2 Viterbi algorithm for hidden Markov models

Objective: To obtain the most likely sequence

$$\underset{S_0,S_1,\ldots,S_T}{\arg\max}\, P(S_0^T = s_0^T \mid X_0^T = x_0^T) \equiv \underset{S_0,S_1,\ldots,S_T}{\arg\max}\, P(S_0^T = s_0^T, X_0^T = x_0^T) \tag{B.4}$$

General ideas:

For each pair (j,t), $j \in \{0,\ldots,J\}$ and $t \in \{0,\ldots,T\}$, we need to define

$$\delta_t(j) = \max_{S_0,S_1,\ldots,S_{t-1}} P(S_0^{t-1} = s_0^{t-1}, S_t = j, X_0^t = x_0^t) \tag{B.5}$$

The algorithm is started at $t = 0$ for $j \in \{0,\ldots,J\}$

$$\delta_0(j) = P(S_0 = j, X_0 = x_0) = P(X_0 = x_0 \mid S_0 = j)P(S_0 = j) = b_j(x_0)\pi_j \tag{B.6}$$

The dynamic programming equation is written as

$$\delta_t(j) = \max_{S_0,S_1,\ldots,S_{t-1}} P(S_0^{t-1} = s_0^{t-1}, S_t = j, X_0^t = x_0^t) = b_j(x_t) \max_i \{p_{ij}\delta_{t-1}(i)\} \tag{B.7}$$

and we keep track of the argument that maximizes $\delta_t(j)$:

$$\psi_t(j) = \arg\max_i \{p_{ij}\delta_{t-1}(i)\} \tag{B.8}$$

for $j = 0, \ldots, J$.

Thus, at the end ($t = T$) we obtain a probability matrix $\delta_{(J+1)\times(T+1)}$ and a state sequence matrix $\psi_{(J+1)\times(T+1)}$

$$\delta = \begin{pmatrix} \delta_0(0) & \delta_1(0) & \cdots & \delta_t(0) & \cdots & \delta_T(0) \\ \vdots & \vdots & & \vdots & & \vdots \\ \delta_0(j) & \delta_1(j) & \cdots & \delta_t(j) & \cdots & \delta_T(j) \\ \vdots & \vdots & & \vdots & & \vdots \\ \delta_0(J) & \delta_1(J) & \cdots & \delta_t(J) & \cdots & \delta_T(J) \end{pmatrix};$$

$$\psi = \begin{pmatrix} \psi_0(0) & \psi_1(0) & \cdots & \psi_t(0) & \cdots & \psi_T(0) \\ \vdots & \vdots & & \vdots & & \vdots \\ \psi_0(j) & \psi_1(j) & \cdots & \psi_t(j) & \cdots & \psi_T(j) \\ \vdots & \vdots & & \vdots & & \vdots \\ \psi_0(J) & \psi_1(J) & \cdots & \psi_t(J) & \cdots & \psi_T(J) \end{pmatrix}$$

allowing us to reconstruct the complete sequence by going backwards

$$S_t^* = \psi_{t+1}(S_{t+1}^*) \tag{B.9}$$

Programming procedure:

The complete procedure can be stated as follows (Rabiner, 1989):

1. Initialization:

$$\delta_0(i) = \pi_i b_i(x_0) \tag{B.10}$$

$$\psi_0(i) = 0 \tag{B.11}$$

for $i = 0, \ldots, J$.

2. Recursion:

$$\delta_t(j) = \max_{0 \leqslant i \leqslant J} \{\delta_{t-1}(i)p_{ij}\} b_j(x_t) \tag{B.12}$$

$$\psi_t(j) = \arg\max_{0 \leqslant i \leqslant J} \{\delta_{t-1}(i)p_{ij}\} \tag{B.13}$$

for $t = 1, \ldots, T$ and $j = 0, \ldots, J$

3. Termination:

$$P^* = \max_{0 \leqslant i \leqslant J} \{\delta_T(i)\} \tag{B.14}$$

$$S_T^* = \arg\max_{0 \leqslant i \leqslant J} \{\delta_T(i)\} \tag{B.15}$$

4. Path (state sequence) backtracking:

$$S_t^* = \psi_{t+1}(S_{t+1}^*) \tag{B.16}$$

for $t = T-1, T-2, \ldots, 0$.

B.3 Viterbi algorithm for hidden semi-Markov models

The reconstruction of the most likely sequence is analogous to the one of the Viterbi algorithm for HMMs (section B.2), but in this case, it implies searching for the optimal duration for state occupancy. Thus,

Objective: To obtain the most likely sequence

$$\begin{aligned} &\arg\max_{S_0, S_1, \ldots, S_T, S_{t+1} \neq S_t} P(S_0^T = s_0^T \mid X_0^T = x_0^T) \\ &\equiv \arg\max_{S_0, S_1, \ldots, S_T, S_{t+1} \neq S_t} P(S_0^T = s_0^T, X_0^T = x_0^T) \end{aligned} \tag{B.17}$$

General ideas:

Since the state process is now a semi-Markov process, we define

$$\begin{aligned} \delta_t(j) &= \max_{S_0, S_1, \ldots, S_{t-1}} P(S_{t+1} \neq j, S_t = j, S_0^{t-1} = s_0^{t-1}, X_0^t = x_0^t) \\ &= \max \left\{ \max_{1 \leqslant u \leqslant t} \left[\left(\prod_{v=1}^{u-1} b_j(x_{t-v}) \right) d_j(u) \max_{i \neq j} (p_{ij} \delta_{t-u}(i)) \right], \left(\prod_{v=1}^{t} b_j(x_{t-v}) \right) d_j(t+1) \pi_j \right\} \\ &\quad \times b_j(x_t) \end{aligned} \tag{B.18}$$

for $t = 0, \ldots, T-1$; $j = 0, \ldots, J$; and due to the right-censoring of the occupancy time in the last visited state, for $t = T$; $j = 0, \ldots, J$:

$$\delta_T(j) = \max_{S_0,S_1,\ldots,S_{T-1}} P(S_T = j, S_0^{T-1} = s_0^{T-1}, X_0^T = x_0^T)$$
$$= \max\left\{ \max_{1\leqslant u\leqslant T}\left[\left(\prod_{v=1}^{u-1} b_j(x_{T-v})\right) D_j(u) \max_{i\neq j}\left(p_{ij}\delta_{T-u}(i)\right)\right], \left(\prod_{v=1}^{T} b_j(x_{T-v})\right) D_j(T+1)\pi_j \right\}$$
$$\times\, b_j(x_T) \quad \text{(B.19)}$$

where $D_j(u) = \sum_{v\geqslant u} d_j(v)$ is the survivor function for the occupancy time in state j.

We also keep track of the arguments maximizing $\delta_t(j)$:

$$\psi_t(j,1:2) = \begin{cases} (j,t), \text{ if } \left(\prod_{v=1}^{t} b_j(x_{t-v})\right) d_j(t+1)\pi_j > \max_{1\leqslant u\leqslant t}\left[\left(\prod_{v=1}^{u-1} b_j(x_{t-v})\right) d_j(u) \max_{i\neq j}\left(p_{ij}\delta_{t-u}(i)\right)\right] \\ \arg\max_{i,u}\left[\left(\prod_{v=1}^{u-1} b_j(x_{t-v})\right) d_j(u) p_{ij}\delta_{t-u}(i)\right], \text{ else} \end{cases} \quad \text{(B.20)}$$

Thus, at the end ($t = T$) we obtain a probability matrix $\delta_{(J+1)\times(T+1)}$ and a state sequence matrix $\psi_{(J+1)\times(T+1)\times 2}$ allowing us to reconstruct the complete sequence by going backwards

$$P^* = \max_{0\leqslant i\leqslant J}\{\delta_T(i)\} \quad \text{(B.21)}$$
$$\hat{s} = \arg\max_{0\leqslant i\leqslant J}\{\delta_T(i)\} \quad \text{(B.22)}$$
$$\hat{r} = \psi_T(\hat{s}, 2) \quad \text{(B.23)}$$
$$t = T \quad \text{(B.24)}$$
$$S^*_{t-\hat{r}+1:t} = \{s\}_{1\times\hat{r}} \quad \text{(B.25)}$$
$$\hat{s} = \psi_t(\hat{s}, 1) \quad \text{(B.26)}$$
$$t = t - \hat{r} \quad \text{(B.27)}$$

Programming procedure:

The complete procedure can be stated as follows (Guédon, 2003):

1. Initialization:

$$\begin{aligned}\delta_0(j) &= P(S_1 \neq j, S_0 = j, X_0 = x_0) \\ &= P(X_0 = x_0 \mid S_0 = j)P(S_1 \neq j \mid S_0 = j)P(S_0 = j) \\ &= b_j(x_0)d_j(1)\pi_j \end{aligned} \tag{B.28}$$

$$\psi_0(j, 1:2) = (0, 0) \tag{B.29}$$

for $j = 0, \ldots, J$.

2. Recursion:

$$\begin{aligned}&\delta_t(j) = b_j(x_t)\times \\ &\max\left\{\max_{1\leqslant u\leqslant t}\left[\left(\prod_{v=1}^{u-1} b_j(x_{t-v})\right) d_j(u)\max_{i\neq j}\left(p_{ij}\delta_{t-u}(i)\right)\right], \left(\prod_{v=1}^{t} b_j(x_{t-v})\right) d_j(t+1)\pi_j\right\}\end{aligned} \tag{B.30}$$

$$\psi_t(j, 1:2) = \begin{cases} (j,t) \quad \text{, if } \left(\prod_{v=1}^{t} b_j(x_{t-v})\right) d_j(t+1)\pi_j \\ \qquad > \max_{1\leqslant u\leqslant t}\left[\left(\prod_{v=1}^{u-1} b_j(x_{t-v})\right) d_j(u)\max_{i\neq j}\left(p_{ij}\delta_{t-u}(i)\right)\right] \\ \arg\max_{i,u}\left[\left(\prod_{v=1}^{u-1} b_j(x_{t-v})\right) d_j(u)p_{ij}\delta_{t-u}(i)\right], \text{ else} \end{cases} \tag{B.31}$$

for $t = 1, \ldots, T-1$ and $j = 0, \ldots, J$; and

$$\begin{aligned}&\delta_T(j) = b_j(x_T)\times \\ &\max\left\{\max_{1\leqslant u\leqslant T}\left[\left(\prod_{v=1}^{u-1} b_j(x_{T-v})\right) D_j(u)\max_{i\neq j}\left(p_{ij}\delta_{T-u}(i)\right)\right], \left(\prod_{v=1}^{T} b_j(x_{T-v})\right) D_j(T+1)\pi_j\right\}\end{aligned} \tag{B.32}$$

$$\psi_T(j, 1:2) = \begin{cases} (j,T-1) \quad \text{, if } \left(\prod_{v=1}^{T} b_j(x_{T-v})\right) D_j(T+1)\pi_j \\ \qquad > \max_{1\leqslant u\leqslant T}\left[\left(\prod_{v=1}^{u-1} b_j(x_{T-v})\right) D_j(u)\max_{i\neq j}\left(p_{ij}\delta_{T-u}(i)\right)\right] \\ \arg\max_{i,u}\left[\left(\prod_{v=1}^{u-1} b_j(x_{T-v})\right) D_j(u)p_{ij}\delta_{T-u}(i)\right], \text{ else} \end{cases} \tag{B.33}$$

for $t = T$ and $j = 0, \ldots, J$.

3. Termination:

$$P^* = \max_{0 \leqslant i \leqslant J} \{\delta_T(i)\} \tag{B.34}$$

$$\hat{s} = \arg\max_{0 \leqslant i \leqslant J} \{\delta_T(i)\} \tag{B.35}$$

$$\hat{r} = \psi_T(\hat{s}, 2) \tag{B.36}$$

4. Path (state sequence) backtracking:

$$t = T \tag{B.37}$$

$$S^*_{t-\hat{r}+1:t} = \{s\}_{1 \times \hat{r}} \tag{B.38}$$

$$\hat{s} = \psi_t(\hat{s}, 1) \tag{B.39}$$

$$t = t - \hat{r} \tag{B.40}$$

for $t \geqslant 0$.

B.4 Forward-backward Viterbi algorithm for hidden semi-Markov models

Objective: As in regular Viterbi, to obtain the most likely sequence

$$\arg\max_{S_0,\ldots,S_T, S_{t+1} \neq S_t} P(S_0^T = s_0^T \mid X_0^T = x_0^T) \equiv \arg\max_{S_0,\ldots,S_T, S_{t+1} \neq S_t} P(S_0^T = s_0^T, X_0^T = x_0^T) \tag{B.41}$$

General ideas:

We define

$$\begin{aligned} \gamma_t(j) &= \max_{S_0,\ldots,S_{t-1},S_{t+1},\ldots,S_T} P(S_0^{t-1} = s_0^{t-1}, S_t = j, S_{t+1} \neq j, S_{t+1}^T = s_{t+1}^T, X_0^T = x_0^T) \\ &= \max_{S_{t+1},\ldots,S_T} P(X_{t+1}^T = x_{t+1}^T, S_{t+1}^T = s_{t+1}^T \mid S_{t+1} \neq j, S_t = j) \\ &\times \max_{S_0,\ldots,S_{t-1}} P(S_{t+1} \neq j, S_t = j, S_0^{t-1} = s_0^{t-1}, X_0^t = x_0^t) = \beta_t(j)\alpha_t(j) \end{aligned} \tag{B.42}$$

for $j = 0, \ldots, J$; where

$$\beta_t(j) = \max_{k \neq j} \left\{ p_{jk} \max \left[\left(\prod_{v=t+1}^{T} b_k(x_v) \right) D_k(T-t), \max_{1 \leqslant u_2 \leqslant T-1-t} \left(\beta_{t+u_2}(k) d_k(u_2) \prod_{v=t+1}^{t+u_2} b_k(x_v) \right) \right] \right\} \tag{B.43}$$

for $T-1, \ldots, 0$;

$$\beta_T(j) = 1$$

$$\alpha_t(j) = b_j(x_t) \max \left\{ \max_{1 \leqslant u_1 \leqslant t} \left[\left(\prod_{v=t-u_1+1}^{t-1} b_j(x_v) \right) d_j(u_1) \max_{i \neq j} (p_{ij} \alpha_{t-u_1}(i)) \right], \left(\prod_{v=0}^{t-1} b_j(x_v) \right) d_j(t+1) \pi_j \right\} \tag{B.44}$$

for $t = 0, \ldots, T-1$; and due to the right-censoring of the occupancy time in the last visited state, for $t = T$; $j = 0, \ldots, J$:

$$\alpha_T(j) = b_j(x_T) \times \max \left\{ \max_{1 \leqslant u_1 \leqslant T} \left[\left(\prod_{v=T+1-u_1}^{T-1} b_j(x_v) \right) D_j(u_1) \max_{i \neq j} (p_{ij} \alpha_{T-u_1}(i)) \right], \left(\prod_{v=0}^{T-1} b_j(x_v) \right) D_j(T+1) \pi_j \right\} \tag{B.45}$$

We also keep track of the arguments maximizing $\gamma_t(j)$:

$$\psi_t(j, 1:2) = \begin{cases} (j,t), \text{ if } \left(\prod_{v=0}^{t-1} b_j(x_v) \right) d_j(t+1) \pi_j > \max_{1 \leqslant u_1 \leqslant t} \left[\left(\prod_{v=t-u_1+1}^{t-1} b_j(x_v) \right) d_j(u_1) \max_{i \neq j} (p_{ij} \alpha_{t-u_1}(i)) \right] \\ \arg\max_{i, u_1} \left[\left(\prod_{v=t-u_1+1}^{t-1} b_j(x_v) \right) d_j(u_1) p_{ij} \alpha_{t-u_1}(i) \right], \text{ else} \end{cases} \tag{B.46}$$

Thus, at the end ($t = T$) we obtain a probability matrix $\gamma_{(J+1)\times(T+1)}$ and a state sequence matrix $\psi_{(J+1)\times(T+1)\times 2}$ allowing us to reconstruct the complete sequence by going backwards

$$P^* = \max_{0 \leqslant i \leqslant J} \{\gamma_T(i)\} \tag{B.47}$$

$$\hat{s} = \arg\max_{0 \leqslant i \leqslant J} \{\gamma_T(i)\} \tag{B.48}$$

$$\hat{r} = \psi_T(\hat{s}, 2) \tag{B.49}$$

$$t = T \tag{B.50}$$

$$S^*_{t-\hat{r}+1:t} = \{s\}_{1 \times \hat{r}} \tag{B.51}$$

$$\hat{s} = \psi_t(\hat{s}, 1) \tag{B.52}$$

$$t = t - \hat{r} \tag{B.53}$$

Programming procedure:

The complete procedure can be stated as follows (Guédon, 2007):

1. Initialization:

$$\alpha_0(j) = b_j(x_0) d_j(1) \pi_j \tag{B.54}$$

$$\beta_T(j) = 1 \tag{B.55}$$

$$\psi_0(j, 1:2) = (0, 0) \tag{B.56}$$

for $j = 0, \ldots, J$.

2. Recursion:

$$\alpha_{t_1}(j) = b_j(x_{t_1}) \max \left\{ \max_{1 \leqslant u_1 \leqslant t_1} \left[\left(\prod_{v=t_1-u_1+1}^{t_1-1} b_j(x_v) \right) d_j(u_1) \max_{i \neq j} (p_{ij} \alpha_{t_1-u_1}(i)) \right], d_j(t_1+1) \pi_j \prod_{v=0}^{t_1-1} b_j(x \right. \tag{B.57}$$

$$\psi_{t_1}(j, 1:2) = \begin{cases} (j, t_1), \text{ if } \left(\prod_{v=0}^{t_1-1} b_j(x_v) \right) d_j(t_1+1) \pi_j > \max_{1 \leqslant u_1 \leqslant t_1} \left[\prod_{v=t_1-u_1+1}^{t_1-1} b_j(x_v) d_j(u_1) \max_{i \neq j} (p_{ij} \alpha_{t_1- \right. \\ \arg\max_{i,u_1} \left[\left(\prod_{v=t_1-u_1+1}^{t_1-1} b_j(x_v) \right) d_j(u_1) p_{ij} \alpha_{t_1-u_1}(i) \right], \text{ else} \end{cases} \tag{B.58}$$

$$\beta_{t_2}(j) = \max_{k \neq j} \left\{ p_{jk} \max \left[\left(\prod_{v=t_2+1}^{T} b_k(x_v) \right) D_k(T - t_2), \max_{1 \leqslant u_2 \leqslant T-1-t_2} \left(\beta_{t_2+u_2}(k) d_k(u_2) \prod_{v=t_2+1}^{t_2+u_2} b_k(x_v) \right) \right. \right. \tag{B.59}$$

for $t_1 = 1, \ldots, T-1$; $t_2 = T-1, \ldots, 0$; $j = 0, \ldots, J$; and

$$\alpha_T(j) \tag{B.60}$$

$$= b_j(x_T) \max \left\{ \max_{1 \leqslant u_1 \leqslant T} \left[\left(\prod_{v=T+1-u_1}^{T-1} b_j(x_v) \right) D_j(u_1) \max_{i \neq j} \left(p_{ij} \alpha_{T-u_1}(i) \right) \right], \left(\prod_{v=0}^{T-1} b_j(x_v) \right) D_j(T+1) \pi_j \right\} \tag{B.61}$$

$$\psi_T(j, 1:2) = \begin{cases} (j, T-1), \\ \text{if } D_j(T+1)\pi_j \prod_{v=0}^{T-1} b_j(x_v) > \max_{1 \leqslant u_1 \leqslant T} \left[\prod_{v=T+1-u_1}^{T-1} b_j(x_v) D_j(u_1) \max_{i \neq j} \left(p_{ij} \alpha_{T-u_1}(i) \right) \right] \\ \arg\max_{i, u_1} \left[\left(\prod_{v=T+1-u_1}^{T-1} b_j(x_v) \right) D_j(u_1) p_{ij} \delta_{T-u_1}(i) \right], \text{ else} \end{cases} \tag{B.62}$$

for $j = 0, \ldots, J$. Then

$$\gamma_t(j) = \beta_t(j) \alpha_t(j)$$

for $t = 0, \ldots, T$ and $j = 0, \ldots, J$.

3. Termination:

$$P^* = \max_{0 \leqslant i \leqslant J} \{\alpha_T(i)\} \tag{B.63}$$

$$\hat{s} = \arg\max_{0 \leqslant i \leqslant J} \{\alpha_T(i)\} \tag{B.64}$$

$$\hat{r} = \psi_T(\hat{s}, 2) \tag{B.65}$$

4. Path (state sequence) backtracking:

$$t = T \tag{B.66}$$

$$S^*_{t-\hat{r}+1:t} = \{s\}_{1 \times \hat{r}} \tag{B.67}$$

$$\hat{s} = \psi_t(\hat{s}, 1) \tag{B.68}$$

$$t = t - \hat{r} \tag{B.69}$$

for $t \geqslant 0$.

B.5 Constrained forward-backward Viterbi algorithm for hidden semi-Markov models

Here we assume we know that the sequence starts at a given state k and ends at the same state k, which is what happens in our study case (i.e., all fishing trips start and end at cruising).

Objective: To obtain the most likely sequence

$$\begin{aligned} &\underset{S_1,\ldots,S_{T-1},S_{t+1}\neq S_t}{\arg\max} P(S_0=k, S_1^{T-1}=s_1^{T-1}, S_T=k \mid X_0^T=x_0^T) \\ \equiv &\underset{S_1,\ldots,S_{T-1},S_{t+1}\neq S_t}{\arg\max} P(S_0=k, S_1^{T-1}=s_1^{T-1}, S_T=k, X_0^T=x_0^T) \end{aligned} \tag{B.70}$$

General ideas:
We define

$$\begin{aligned} \gamma_t(j) &= \max_{S_1,\ldots,S_{t-1},S_{t+1},\ldots,S_{T-1}} P(S_0=k, S_1^{t-1}=s_1^{t-1}, S_t=j, S_{t+1}\neq j, S_{t+1}^{T-1}=s_{t+1}^{T-1}, S_T=k, X_0^T=x_0^T) \\ &= \max_{S_{t+1},\ldots,S_{T-1}} P(X_{t+1}^T=x_{t+1}^T, S_T=k, S_{t+1}^{T-1}=s_{t+1}^{T-1} \mid S_{t+1}\neq j, S_t=j) \\ &\times \max_{S_1,\ldots,S_{t-1}} P(S_{t+1}\neq j, S_t=j, S_1^{t-1}=s_1^{t-1}, S_0=k, X_0^t=x_0^t) = \beta_t(j)\alpha_t(j) \end{aligned} \tag{B.71}$$

for $j=0,\ldots,J$; where

$$\alpha_t(j) = \begin{cases} b_j(x_t)\max\left\{\max\limits_{1\leqslant u_1\leqslant t-1}\left[\left(\prod\limits_{v=t-u_1+1}^{t-1} b_j(x_v)\right) d_j(u_1)\max\limits_{i\neq j}\left(p_{ij}\alpha_{t-u_1}(i)\right)\right], d_j(t+1)\pi_j\prod\limits_{v=0}^{t-1} b_j(x_v)\right\}, \\ \text{if } j=k \\ b_j(x_t)\max\limits_{1\leqslant u_1\leqslant t}\left\{\left(\prod\limits_{v=t-u_1+1}^{t-1} b_j(x_v)\right) d_j(u_1)\max\limits_{i\neq j}\left[p_{ij}\alpha_{t-u_1}(i)\right]\right\}, \text{ else} \end{cases} \tag{B.72}$$

for $t=1,\ldots,T-1$;

$$\alpha_0(j) = \begin{cases} b_k(x_0)d_k(1)\pi_k, \text{ if } j=k \\ 0, \text{ else} \end{cases}$$

$$\alpha_T(j) = \begin{cases} \max\left\{ \max\limits_{1\leqslant u_1\leqslant T-1} \left[\left(\prod\limits_{v=T+1-u_1}^{T-1} b_j(x_v) \right) D_j(u_1) \max\limits_{i\neq j} \left(p_{ij}\alpha_{T-u_1}(i)\right) \right], D_j(T+1)\pi_j \prod\limits_{v=0}^{T-1} b_j(x_v) \right\} \\ \times b_j(x_T), \text{ if } j = k \\ b_j(x_T) \max\limits_{1\leqslant u_1\leqslant T} \left\{ \left(\prod\limits_{v=T+1-u_1}^{T-1} b_j(x_v) \right) D_j(u_1) \max\limits_{i\neq j} \left[p_{ij}\alpha_{T-u_1}(i)\right] \right\}, \text{ else} \end{cases}$$

$$\beta_t(j) = \begin{cases} \max\limits_{l\neq j} \left\{ p_{jl} \max\limits_{1\leqslant u_2\leqslant T-1-t} \left[\beta_{t+u_2}(l) d_l(u_2) \prod\limits_{v=t+1}^{t+u_2} b_l(x_v) \right] \right\}, \text{ if } j = k \\ \max\{ \max\limits_{l\neq\{j,k\}} \left[p_{jl} \max\limits_{1\leqslant u_2\leqslant T-1-t} \left(\beta_{t+u_2}(l) d_l(u_2) \prod\limits_{v=t+1}^{t+u_2} b_l(x_v) \right) \right], \\ D_k(T-t) p_{jk} \prod\limits_{v=t+1}^{T} b_k(x_v) \}, \text{ else} \end{cases} \tag{B.73}$$

for $0, \ldots, T-2$;

$$\beta_T(j) = \begin{cases} 1, \text{ if } j = k \\ 0, \text{ else} \end{cases}$$

$$\beta_{T-1}(j) = \begin{cases} 0, \text{ if } j = k \\ b_k(x_T) D_k(1) p_{jk}, \text{ else} \end{cases}$$

We also keep track of the arguments maximizing $\gamma_t(j)$:

$$\psi_t(j, 1:2) = \begin{cases} (j,t), \text{ if } j = k \ \& \ \left(\prod\limits_{v=0}^{t-1} b_j(x_v) \right) d_j(t+1)\pi_j > \\ \max\limits_{1\leqslant u_1\leqslant t-1} \left[\left(\prod\limits_{v=t-u_1+1}^{t-1} b_j(x_v) \right) d_j(u_1) \max\limits_{i\neq j} \left(p_{ij}\alpha_{t-u_1}(i)\right) \right] \\ \arg\max\limits_{i\neq j, u_1} \left[\left(\prod\limits_{v=t-u_1+1}^{t-1} b_j(x_v) \right) d_j(u_1) p_{ij} \alpha_{t-u_1}(i) \right], \text{ elseif } (j = k \ \& \ 1 \leqslant u_1 \leqslant t-1) \\ \text{ or elseif } (j \neq k \ \& \ 1 \leqslant u_1 \leqslant t) \end{cases} \tag{B.74}$$

Thus, at the end ($t = T$) we obtain a probability matrix $\gamma_{(J+1)\times(T+1)}$ and a state sequence matrix $\psi_{(J+1)\times(T+1)\times 2}$ allowing us to reconstruct the complete sequence by going backwards

$$P^* = \max_{0 \leqslant i \leqslant J} \{\gamma_T(i)\} \tag{B.75}$$

$$\hat{s} = \arg\max_{0 \leqslant i \leqslant J} \{\gamma_T(i)\} \tag{B.76}$$

$$\hat{r} = \psi_T(\hat{s}, 2) \tag{B.77}$$

$$t = T \tag{B.78}$$

$$S^*_{t-\hat{r}+1:t} = \{s\}_{1 \times \hat{r}} \tag{B.79}$$

$$\hat{s} = \psi_t(\hat{s}, 1) \tag{B.80}$$

$$t = t - \hat{r} \tag{B.81}$$

Programming procedure:

The complete procedure can be stated as follows:

1. Initialization:

$$\alpha_0(j) = \begin{cases} b_k(x_0) d_k(1) \pi_k, \text{ if } j = k \\ 0, \text{ else} \end{cases} \tag{B.82}$$

$$\psi_0(j, 1:2) = (0, 0) \tag{B.83}$$

$$\alpha_1(j) = \begin{cases} b_j(x_1) b_j(x_0) d_j(2) \pi_j, \text{ if } j = k \\ b_j(x_1) d_j(1) p_{kj} \alpha_0(k), \text{ else} \end{cases} \tag{B.84}$$

$$\psi_1(j, 1:2) = (1, k) \tag{B.85}$$

$$\beta_T(j) = \begin{cases} 1, \text{ if } j = k \\ 0, \text{ else} \end{cases} \tag{B.86}$$

$$\beta_{T-1}(j) = \begin{cases} 0, \text{ if } j = k \\ b_k(T) D_k(1) p_{jk}, \text{ else} \end{cases} \tag{B.87}$$

$$\tag{B.88}$$

for $j = 0, \ldots, J$.

2. Recursion:

$$\alpha_{t_1}(j) = \begin{cases} \max\left\{ \max\limits_{1\leqslant u_1\leqslant t_1-1} \left[\prod\limits_{v=t_1-u_1+1}^{t_1-1} b_j(x_v) d_j(u_1) \max\limits_{i\neq j} (p_{ij}\alpha_{t_1-u_1}(i)) \right], d_j(t_1+1)\pi_j \prod\limits_{v=0}^{t_1-1} b_j(x_v) \right\} \\ \times b_j(x_{t_1}), \text{ if } j = k \\ b_j(x_{t_1}) \max\limits_{1\leqslant u_1\leqslant t_1} \left\{ \left(\prod\limits_{v=t_1-u_1+1}^{t_1-1} b_j(x_v) \right) d_j(u_1) \max\limits_{i\neq j} [p_{ij}\alpha_{t_1-u_1}(i)] \right\}, \text{ else} \end{cases} \tag{B.89}$$

$$\psi_{t_1}(j, 1:2) = \begin{cases} (j, t_1), \text{ if } j = k \ \&\ \left(\prod\limits_{v=0}^{t_1-1} b_j(x_v) \right) d_j(t_1+1)\pi_j > \\ \max\limits_{1\leqslant u_1\leqslant t_1-1} \left[\left(\prod\limits_{v=t_1-u_1+1}^{t_1-1} b_j(x_v) \right) d_j(u_1) \max\limits_{i\neq j} (p_{ij}\alpha_{t_1-u_1}(i)) \right] \\ \arg\max\limits_{i\neq j, u_1} \left[\left(\prod\limits_{v=t_1-u_1+1}^{t_1-1} b_j(x_v) \right) d_j(u_1) p_{ij}\alpha_{t_1-u_1}(i) \right], \\ \text{elseif } (j = k \ \&\ 1 \leqslant u_1 \leqslant t_1 - 1) \text{ or elseif } (j \neq k \ \&\ 1 \leqslant u_1 \leqslant t_1) \end{cases} \tag{B.90}$$

$$\beta_{t_2}(j) = \begin{cases} \max\limits_{l\neq j} \left\{ p_{jl} \max\limits_{1\leqslant u_2\leqslant T-1-t_2} \left[\beta_{t_2+u_2}(l) d_l(u_2) \prod\limits_{v=t_2+1}^{t_2+u_2} b_l(x_v) \right] \right\}, \text{ if } j = k \\ \max\{ \max\limits_{l\neq\{j,k\}} \left[p_{jl} \max\limits_{1\leqslant u_2\leqslant T-1-t_2} \left(\beta_{t_2+u_2}(l) d_l(u_2) \prod\limits_{v=t_2+1}^{t_2+u_2} b_l(x_v) \right) \right], \\ D_k(T-t_2) p_{jk} \prod\limits_{v=t_2+1}^{T} b_k(x_v) \}, \text{ else} \end{cases} \tag{B.91}$$

for $t_1 = 2, \ldots, T-1$; $t_2 = T-2, \ldots, 0$; $j = 0, \ldots, J$; and

$$\alpha_T(j) = \begin{cases} \max\left\{ \max\limits_{1\leqslant u_1\leqslant T-1} \left[\left(\prod\limits_{v=T+1-u_1}^{T-1} b_j(x_v) \right) D_j(u_1) \max\limits_{i\neq j} (p_{ij}\alpha_{T-u_1}(i)) \right], D_j(T+1)\pi_j \prod\limits_{v=0}^{T-1} b_j(x_v) \right\} \\ \times b_j(x_T), \text{ if } j = k \\ b_j(x_T) \max\limits_{1\leqslant u_1\leqslant T} \left\{ \left(\prod\limits_{v=T+1-u_1}^{T-1} b_j(x_v) \right) D_j(u_1) \max\limits_{i\neq j} [p_{ij}\alpha_{T-u_1}(i)] \right\}, \text{ else} \end{cases} \tag{B.92}$$

$$\psi_T(j,1:2)=\begin{cases}(j,T-1),\text{ if } j=k\ \&\ \left(\prod_{v=0}^{T-1}b_j(x_v)\right)D_j(T+1)\pi_j> \\ \max\limits_{1\leqslant u_1\leqslant T-1}\left[\left(\prod\limits_{v=T+1-u_1}^{T-1}b_j(x_v)\right)D_j(u_1)\max\limits_{i\neq j}\left(p_{ij}\alpha_{T-u_1}(i)\right)\right] \\ \arg\max\limits_{i\neq j,u_1}\left[\left(\prod\limits_{v=T+1-u_1}^{T-1}b_j(x_v)\right)D_j(u_1)p_{ij}\alpha_{T-u_1}(i)\right], \\ \text{elseif } (j=k\ \&\ 1\leqslant u_1\leqslant T-1) \text{ or elseif } (j\neq k\ \&\ 1\leqslant u_1\leqslant T)\end{cases} \tag{B.93}$$

for $j=0,\ldots,J$. Then

$$\gamma_t(j)=\beta_t(j)\alpha_t(j)$$

for $t=0,\ldots,T$ and $j=0,\ldots,J$.

3. Termination:

$$P^*=\max_{0\leqslant i\leqslant J}\{\alpha_T(i)\} \tag{B.94}$$

$$\hat{s}=\arg\max_{0\leqslant i\leqslant J}\{\alpha_T(i)\} \tag{B.95}$$

$$\hat{r}=\psi_T(\hat{s},2) \tag{B.96}$$

4. Path (state sequence) backtracking:

$$t=T \tag{B.97}$$

$$S^*_{t-\hat{r}+1:t}=\{s\}_{1\times\hat{r}} \tag{B.98}$$

$$\hat{s}=\psi_t(\hat{s},1) \tag{B.99}$$

$$t=t-\hat{r} \tag{B.100}$$

for $t\geqslant 0$.

Appendix C

Details on the computation of accuracy, precision, recall and F1 indicators

This appendix refers to indicators of model performance for inferring behavioral modes (Chapter 4).

Figure C.1 illustrates a sequence starting at time 0 and ending at 11, where the true behavioral modes are given and compared with the inferred ones. Accuracy is the percentage of steps where the inferred states correspond to the real ones, so it is equal to $7/11 \times 100\% = 63.6\%$.

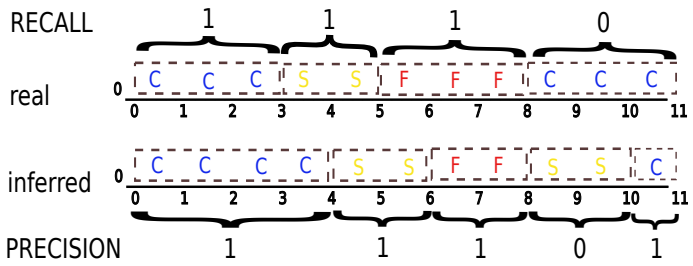


Figure C.1: Example of a sequence with its real and inferred behavioral modes. 1's and 0's in recall/precision represent a positive or null recall/precision corresponding to each behavioral mode, respectively. C=cruising, S=searching and F=fishing.

Precision is the percentage of inferred segments where the inferred behavioral mode corresponds to the true one. An inferred mode m starting at time t and ending at time $t+u$ is said to correspond to a true mode if there is a true mode m in the time interval $[t, t+u]$ with an associated duration of at least $u/2$. The $u/2$

threshold responds to the limitations of the time resolution of the sequence ($\sim$ 1 per hour) and the mean duration of the modes ($\sim$ 2 hours for fishing and searching). Precision, as well as recall and F1, are computed individually for each behavioral mode.

In the example of Fig C.1, there are two inferred segments for cruising. According to the definition above, the two inferred cruising segments correspond to true cruising segments. Thus,

$$\text{Precision(cruising)} = (1+1)/2 \times 100\% = 100\%$$

Likewise, for searching and fishing we have:

$$\text{Precision(searching)} = 1/2 \times 100\% = 50\%$$
$$\text{Precision(fishing)} = 1/1 \times 100\% = 100\%$$

Recall is the percentage of real segments where the true mode is correctly inferred. A true mode m starting at time t and ending at time $t+u$ is said to be correctly inferred if within the time interval $[t, t+u]$ a behavioral mode m has been inferred with an associated duration of at least $u/2$. According to this definition, we compute for our example:

$$\text{Recall(cruising)} = 1/2 \times 100\% = 50\%$$
$$\text{Recall(searching)} = 1/1 \times 100\% = 100\%$$
$$\text{Recall(fishing)} = 1/1 \times 100\% = 100\%$$

Since F1 is defined as the harmonic mean of recall and precision, we get:

$$\text{F1(cruising)} = 0.667$$
$$\text{F1(searching)} = 0.667$$
$$\text{F1(fishing)} = 1$$

Appendix D

Details on the simulation study: HMM vs. HSMM for different time-resolution data

This appendix describes the parameters used for the simulation study described and discussed in section 4.4.2.

We considered only one observed variable in our simulation, in order to keep it simple. The probability distribution of this variable conditioned on one state is Gaussian, with $\mu = 11$ and $\sigma = 3$. Its distribution conditioned on the other state is truncated normal with $\mu = 6$ and $\sigma = 5$, and bounded on $[0, 25]$.

The one-second rate state sequence was simulated considering a logistic distribution with parameters $\mu = 6.89$ and $s = 1.21$ for the duration of one state, and a generalized extreme value (GEV) distribution with parameters $\xi = 0.37$, $\sigma = 0.4$ and $\mu = 2.1$ for the duration of the other state. For the groundtruthed data study, logistic and GEV distributions are also used for modeling the duration of two of the behavioral modes.

Appendix E

Hybrid models and covariates for inferring behavioral modes

E.1 Hybrid models

A weakness of the Markovian models seems to be in their observation model. Observed features are often multidimensional, and the appropriated multivariate probability density function (pdf) is difficult to find. A classical approach is to use Gaussian mixtures, but then the number of mixtures must be fixed. Another approach is to assume mutual independence between observed variables and fit an univariate probability distribution to each one of them. This last approach was taken. It became difficult to fit pdfs and to find some that passed the goodness-of-fit tests. Yet, a different approach could be taken. Hybrid models, stated as Markovian models that rely on the definition of an observation likelihood from the output of a discriminative model, are proposed in this section. Regarding the results obtained with the hidden semi-Markov models (HSMMs; section 4.3), the searching modes were the most difficult to infer. Therefore, hybrid models were evaluated in terms of both global and searching-mode inference accuracy.

E.1.1 Hybrid HSMM/ANN models

- It has been shown (Richard and Lippmann, 1991) that $P(S_t = j \mid X_t = x_t)$ independently at each time t is computed by artificial neural networks (ANNs) when minimizing MSE.

- The observed likelihood
  $b_j(x_t) = P(X_t = x_t \mid S_t = j) = P(S_t = j \mid X_t = x_t)P(X_t = x_t)/P(S_t = j)$

- $P(X_t = x_t)$ is constant

Thus, $\max P(X_t = x_t \mid S_t = j) \equiv \max P(S_t = j \mid X_t = x_t)/P(S_t = j)$. It is usual to only use the numerator $P(S_t = j \mid X_t = x_t)$ term, which is obtained from ANNs, because a prior on states occurrence might be difficult to establish. Another possibility would be to approximate $P(S_t = j)$ as the proportion of j in the training dataset. We did both: with and without $P(S_t = j)$ priors.

Another way of using ANNs for the observation term of HSMMs is shown as following:

$$\hat{b}_j(x_t) = \exp(\beta \log(P(S_t = j \mid X_t = x_t)))$$

where $\beta > 0$ and was numerically estimated. Its value was chosen as the one that maximized hybrid-model performance.

Likewise, the combination of observed variables chosen (see the set of possible observed variables in section 4.2) was the one that maximized hybrid-model performance. Two indicators obeying two different criteria were used for assessing model performance. The first one was global accuracy. The second one was a weighted accuracy equal to the mean between fishing, searching and cruising accuracies.

We thus present results for:

1. ANNs (as in Chapter 4)
2. ANNs + HSMMs where $\hat{b}_j(x_t) = P(S_t = j \mid X_t = x_t)$ (Hybrid1)
3. ANNs + HSMMs where $\hat{b}_j(x_t) = P(S_t = j \mid X_t = x_t)/P(S_t = j)$ (Hybrid2)
4. ANNs + HSMMs where $\hat{b}_j(x_t) = \exp(\beta \log(P(S_t = j \mid X_t = x_t)))$ (Hybrid3)

For each one of the hybrid models, ANN parameters, β (for the 4^{th} case) and observed variables were chosen based on each of the indicators described above. So for each case we show only the combination of observed variables corresponding to each of those indicators (Table E.1). In addition, the indicators of performance defined in section 4.2.3, global accuracy and F1 for each behavioral mode, are shown for comparison purposes. Number of replicas are also shown.

Model	ANN	HSMM	Hybrid1		Hybrid2 (prior)		Hybrid3 (β)
Acc.	Global	Global	Global	Weight	Global	Weight	Weight
Set	sp, Δsp_{-1} Δsp_{+1}, $\Delta\theta_{+1}$	sp, Δsp_{+1}	sp, Δsp_{+1}, $\Delta\theta_{+1}$, θ	sp, Δsp_{+1}, $\Delta\theta_{+1}$	sp, $\Delta\theta_{+1}$	sp, Δsp_{+1}, $\Delta\theta_{+1}$	sp, Δsp_{+1}, $\Delta\theta_{+1}$
G Acc	79.2	80.3	80.3	80.2	78.1	78.0	
FS	56.9	66.7	60.5	59.8	59.8	59.4	< 60
FF	75.4	77.0	82.9	81.5	77.5	82.2	
FC	82.1	89.1	88.6	89.2	89.1	89.8	
♯ rep	20	20	20	5	5	5	1

Table E.1: Performance of ANNs, HSMMs and hybrid HSMM/ANN models for their corresponding best subsets of observed variables (Set), and accuracy criterion for parameter optimization (Acc). G Acc: global accuracy; FS: F1 for searching; FF: F1 for fishing; FC: F1 for cruising; ♯ rep: number of replicas.

Although the fishing and cruising modes were better inferred by hybrid models, they did not surpass the performance of HSMMs neither for global inference or for searching-mode inference.

E.1.2 Hybrid HSMM/SVM models

Support vector machine (SVM) optimization functions can also be translated to a posteriori probabilities (Ganapathiraju *et al.*, 2000). We did the same calculations for hybrid HSMM/SVM models as for hybrid HSMM/ANN models (Table E.2).

As for HSMM/ANN models, although the fishing and cruising modes were better inferred by hybrid models, they did not surpass the performance of HSMMs neither for global inference or for searching-mode inference.

E.2 Covariates

The idea was to use covariates available from the VMS data. Here is a list of the considered covariates:

- Hour of the day
- Cosinus of hour (because hour is a circular variable)

Model	SVM	HSMM	Hybrid1		Hybrid2 (prior)		Hybrid3 (β)	
Acc.	Global	Global	Global	Weight	Global	Weight	Global	Weight
Set	sp, Δsp_{-1}, Δsp_{+1}	sp, Δsp_{+1}	sp, Δsp_{+1}, $\Delta\theta_{+1}$	sp, $\Delta\theta_{-1}$	sp, $\Delta\theta_{+1}$	sp, $\Delta\theta_{+1}$	sp, $\Delta\theta_{+1}$	
G Acc	79.0	80.3	79.5	79.3	79.1	79.4	80.7	
FS	56.9	66.7	58.7	60.8	61.3	62.4	59.0	< 60
FF	75.4	77.0	82.4	80.0	81.8	79.7	77.4	
FC	79.8	89.1	88.9	90.5	89.9	89.6	90.8	
♯ rep	20	20	20	5	5	5	1	1

Table E.2: Performance of SVMs, HSMMs and hybrid HSMM/SVM models for their corresponding best subsets of observed variables (Set), and accuracy criterion for parameter optimization (Acc). G Acc: global accuracy; FS: F1 for searching; FF: F1 for fishing; FC: F1 for cruising; ♯ rep: number of replicas.

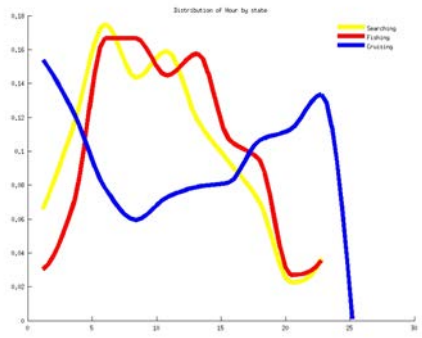
- Cumulated time from moment of departure (in hours)
- Cumulated distance from moment of departure (in nm)
- Distance to the coast (in nm)
- Distance to departure port (in nm)
- Distance to arrival port (in nm)

Empirical probability distributions for each of the covariates conditional to each state are shown in Figure E.1. It should be noticed that covariate densities distributed alike when fishing and searching. We also computed bivariate correlations between covariates (Table E.3). Only Cumulated time and Cumulated distance were strongly correlated.

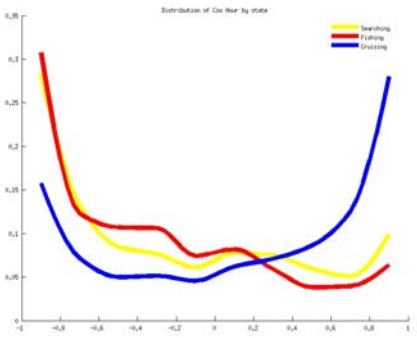
Since there is no theoretical physic model relating covariates Y to states or to observed variables, hybrid modeling approaches were used for introducing covariates into the model. The first one consisted in introducing them directly into the SVM or ANN observation model. That way,

$$\hat{b}_j(x_t) = \exp(\beta_1 \log(P(S_t = j \mid X_t = x_t, Y_t = y_t))) \tag{E.1}$$

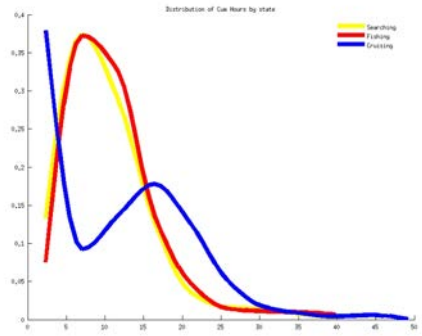
The second approach consisted on using the computed probabilities $P(Y_t = y_t \mid S_t = j)$ and defining the observation model as:



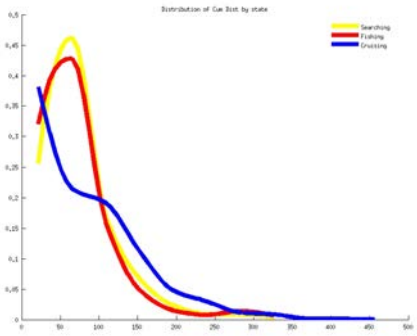
(a) Hour of the day



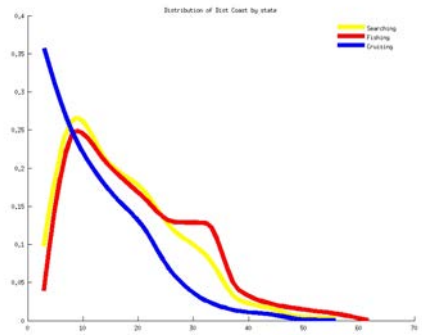
(b) Cosinus of hour



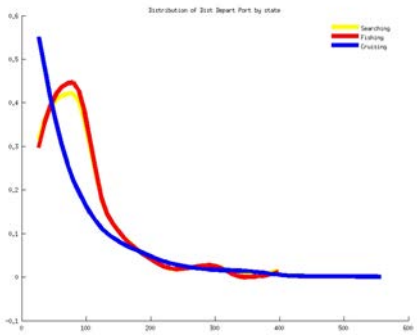
(c) Cumulated Hour



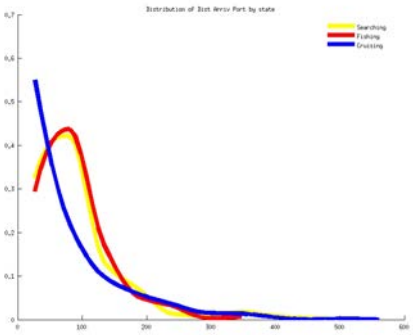
(d) Cumulated Distance



(e) Distance to the coast



(f) Distance to Departure Port



(g) Distance to Arrival Port

Figure E.1: Distribution of covariates conditional to each state. Solid yellow line: searching; solid blue line: cruising; solid red line: fishing.

	Hour	CosH	CumH	CumDist	DistCoast	DepPort	ArrPort
Hour	1	−0.07	0.28	0.21	−0.08	0.05	−0.04
CosH		1	−0.17	−0.1	−0.15	−0.09	−0.1
CumH			1	**0.89**	0.04	0.48	−0.08
CumDist				1	0.09	0.67	0.02
DistCoast					1	0.29	0.34
DepPort						1	0.21
ArrPort							1

Table E.3: Partial correlations between covariates. The strongest correlation is in bold. CosH: cosinus of hour, CumH: cumulated time from moment of departure, CumDist: cumulated distance from moment of departure, DistCoast: distance to the coast, DepPort: distance to departure port, ArrPort: distance to arrival port.

$$\hat{b}_j(x_t) = \exp(\beta_1 \log(P(S_t = j \mid X_t = x_t))) \times \exp(\beta_2 \log(P(Y_t = y_t \mid S_t = j))) \quad \text{(E.2)}$$

for which β_2 was estimated in the same way as β_1 (see section E.1).

For examining the benefits of using covariates, we gave more importance to achieving improvements in the inference of the searching mode rather than global inference. For that reasons, the weighted accuracy indicator is the only criterion used for choosing observed variables, covariates, ANN/SVM parameter values, and for estimating β_1 and β_2. An inferior limit of 0.1 was fixed on $b_j(x_t)$, so that very small values would not strongly affect the likelihoods computed through dynamic programming (appendix B). However, this limit was removed since it worsened the results. From all the covariate analyses, the best result we obtained was for an HSMM/ANN hybrid model with sp, Δsp_{+1} and $\hat{\theta}$ as observed variables and Cumulative time as covariate, using the approach in equation (E.1). After 20 replicas, the average result was 80.2% of accuracy, 60.7%, 81.1% and 88.1% of F-score for searching, fishing and cruising, respectively. It was not better than the one obtained for the HSMM/ANN Hybrid1 model using global accuracy for parameter calibration (Table E.1). This was probably due to the fact that no covariate gave information allowing improvement on searching discrimination (Figure E.1), thus adding more confusion into the model.

To sum up, HSMMs remained the best models for inferring the behavioral modes. Since more sophisticated models have not achieved the expected improvements, it

is more likely that improvements could be done by increasing the resolution of the Vessel Monitoring System data, as discussed in section 4.4.3.

Appendix F

List of publications and communications during the thesis

F.1 Publications and articles in review

- **Joo**, Bertrand, Tam and Fablet, 2013. Hidden Markov models: the best models for forager movements? PLOS ONE. 8 (e71246).

- Bertrand, **Joo**, Arbulu Smet, Tremblay, Barbraud and Weimerskirch, 2012. Local depletion by a fishery can affect seabird foraging. Journal of Applied Ecology. 49 (1168-1177).

- **Joo**, Bertrand, Chaigneau and Ñiquen, 2011. Optimization of an artificial neural network for identification of fishing event positions from Vessel Monitoring System data. Ecological Modelling 222 (1048-1059).

- Bertrand, **Joo** and Fablet, in review. Generalized Pareto for a pattern-oriented random walk modelling of organisms' movement.

- **Joo**, Bertrand, Bouchon, Segura, Chaigneau, Demarcq, Tam, Simier, Gutierrez, Gutierrez, Fablet and Bertrand, in review. Ecosystem scenarios shape fishing spatial behavior. The case of the anchovy fishery in the Northern Humboldt Current System.

- Maufroy, Chassot, **Joo** and Kaplan, in review. First large-scale examination of spatio-temporal patterns of drifting fish aggregating devices from tropical tuna fisheries of the Indian and Atlantic Oceans.

F.2 Communications in conferences

- **Joo**, Fablet, Tam, Bertrand. Hidden semi-Markov modeling for behavioral modes in foraging movement. Oral communication at the 'IIIrd International Statistical Ecology Conference' held in Krokkleiva, Norway; from July 3rd to 6th, 2012.

- Bertrand, **Joo**, Fablet. Do foragers move Lévy or not? Let Generalized Pareto decide. Oral communication at the 'IIIrd International Statistical Ecology Conference' held in Krokkleiva, Norway; from July 3rd to 6th, 2012.

- **Joo**, Tam, Bouchon, Atiquipa, Fablet, Bertrand. Fishing, searching or cruising? A statistical approach for modeling activities from purse-seine anchovy vessels during fishing trips. Oral communication at the 'IIIrd Conferences of Sciences of the Sea' held in Lima, Peru; from June 25th to 29th, 2012.

- **Joo**, Bertrand, Marin, Fablet, Oliveros-Ramos. Random-walk approach for modeling animal movement. Oral communication at the 'XXVth International Biometric Conference' held in Floripa, Brazil; from December 5th to 10th, 2010.

- Bertrand, Weimerskirch, Tremblay, Castillo, Silva, **Joo**, Passuni. Seabirds and fishers competing for the same prey off Peru. Oral communication at the '1st World Seabird Conference' held in Victoria, Canada, from September 7th to 11th, 2010.