Habilitation à Diriger des Recherches

Towards a computational ecology of plankton

Par
Jean-Olivier IRISSON

devant un jury composé de
Andrea DOGLIOLI (rapporteur)
Cédric JAMET (examinateur)
Marina LEVY (examinatrice)
Ketil MALDE (rapporteur)
Maria Grazia MAZZOCCHI (rapportrice)
Frédéric PRECIOSO (examinateur)

À mes étoiles.

Remerciements

La plupart des lecteurs de ce document s'arrêtera à ces pages de remerciements ; je leur suis déjà gré de s'intéresser suffisamment à mon travail pour l'avoir ouvert. Quelque uns feuillèteront les pages qui suivent et s'attarderont sur une figure ou une note de marge ; j'espère qu'ils y trouveront un argument ou un résultat qui leur sera utile par la suite. Trois personnes cependant le liront en entier : Andrea Doglioli, Ketil Malde et Maria Grazia Mazzocchi, je vous remercie vivement d'avoir accepté de le faire et espère que vous trouverez le récit de mes aventures passées palpitant et la description de mon projet futur exaltante ! Cédric Jamet, Marina Levy et Frédéric Precioso viendront compléter ce jury d'exception et je les remercie d'avoir accepté d'en faire partie. Sur une note plus sérieuse, je suis vraiment impatient d'avoir votre avis à toutes et tous sur mon parcours et mon projet. L'Habilitation à Diriger des Recherches est un point d'étape important et rare dans la carrière d'un enseignant-chercheur ; la mienne a été quelque peu sinueuse sur le plan thématique et vos points de vue me seront précieux.

Ensuite, je tiens à remercier les personnes qui ont constitué mon encadrement scientifique durant toutes ces années.

Cela a commencé par mes professeurs à l'École Normale et, en particulier, Régis Ferrière, qui m'a introduit au monde de l'écologie scientifique (alors assez absent des programmes de prépa) et à l'élégance des mathématiques comme outil d'abstraction du monde vivant.

Cela a continué avec le personnel du CRIOBE et Serge Planes, mon directeur de thèse, grâce à qui j'ai entamé des recherches sur la dispersion et le comportement des larves de poissons qui ont occupé les 10 premières années de ma carrière scientifique.

Durant cette période, Claire Paris a eu une influence importante : elle a contribué à l'encadrement de ma thèse, m'a accueilli comme premier post-doc dans son laboratoire, a coordonné avec moi le premier projet d'envergure dont j'ai pu obtenir le financement une fois en poste à Villefranche et a finalement pris en post-doctorat le premier étudiant en thèse que j'ai encadré ; la boucle est bouclée.

Depuis 10 ans déjà (aïe), mon environnement scientifique est le Laboratoire d'Océanographie de Villefranche (LOV). Et je m'y sens extrêmement bien. J'ai trouvé un laboratoire à taille humaine mais diversifié thématiquement ; dans lequel les gens, globalement, s'entendent et partagent

aussi volontiers des idées que des rires ; où je côtoie des chercheur·se·s à la pointe de leur discipline et qui m'inspirent mais qui n'ont pas hésité à m'intégrer à leur travail voire à me demander des conseils, à moi, nouvel arrivant et océanographe relativement novice. Cette ambiance de travail est rare et précieuse. Je remercie tout le personnel du laboratoire d'y contribuer et, en particulier, Antoine Sciandra, Rodolphe Lemée et Fabrizio d'Ortenzio, les directeurs passé et actuels du LOV, d'avoir su l'entretenir.
Plus récemment, les équipes Processus dans les Écosystèmes PélagiqueS (PEPS) et, maintenant, COMPutational PLankton Ecology (COMPLEx) ont occupé une place importante dans mon activité étant donné que j'en ai pris la responsabilité. Merci à tous leurs membres pour leur dynamisme et pour avoir accompagné ma transition vers de nouvelles thématiques de recherche au sein de l'équipe, en particulier l'écologie du plancton et son approche par imagerie, qui constituent maintenant l'immense majorité de mon travail.

J'essaie de m'inspirer des qualités de toutes ces personnes pour l'encadrement des étudiant·e·s dont j'ai la responsabilité. Ils et elles ont été nombreux·ses et j'espère avoir au moins partiellement réussi. Ces étudiant·e·s sont finalement les personnes avec lesquelles j'ai le plus interagi et elles ont donc contribué, très largement, à l'ensemble du travail présenté dans ces pages. Je leur en suis reconnaissant et espère qu'elles ont gardé un bon souvenir des semaines/mois passés en ma compagnie. Parmi elles, Robin Faillettaz a un place particulière : il a été mon premier doctorant, dans le cadre de mon premier projet financé et ses compétences ainsi que ses qualités humaines ont fait que tout s'est très bien passé. Dans les trois années à venir, ce seront Cédric Dubois, Laure Vilgrain et Thelma Panaïotis, les trois doctorant·e·s que je co-encadre actuellement, qui vont prendre cette place particulière. Je me sens extrêmement chanceux d'avoir travaillé et de continuer à travailler avec des étudiant·e·s aussi motivés et sympathiques. Merci !

Enfin, mon travail ne serait pas satisfaisant s'il n'était pas équilibré par une vie personnelle bien remplie. Estelle, Gabin et Louise, c'est vous qui la remplissez, de joie. Je sais que vous devez souvent vous dire que l'équilibre n'est pas en votre faveur et je m'en excuse. Oui, ce travail est aussi accaparant qu'il est passionnant et mon esprit est trop souvent ailleurs quand il devrait être avec vous. J'espère que vous verrez dans ce document la manifestation concrète de ces heures passées loin de vous ; sachez que sans vous, sans toi Estelle, rien de cela n'aurait de sens. Je vous aime.

Acknowledgements

Most readers of this document will stop at these acknowledgements; I am already grateful that they were interested in my work enough to have opened these pages. Some will skim through the following pages and linger on a figure or a few margin notes; I hope that they will find an idea or a result that will be useful to them later. Three people, however, will read it in its entirety: Andrea Doglioli, Ketil Malde, and Maria Grazia Mazzocchi. I thank you very much for agreeing to do so and hope that you will find the tale of my former adventures to be thrilling and the description of my future project exhilarating! Cédric Jamet, Marina Levy, and Frédéric Precioso will complete this exceptional jury and I thank them for agreeing to be part of it. On a more serious note, I am really impatient to have your opinion on my past activities and on my research project. The Habilitation à Diriger des Recherches is an important and rare milestone in the career of a university professor in France (researchers are evaluated more regularly); mine has been somewhat convoluted in terms of research topics and your comments will be very much welcome.

Then, I would like to thank the people who created the scientific environment within which I worked during all these years.
It started with my professors at the École Normale and, in particular, Régis Ferrière, who introduced me to the world of scientific ecology (that was almost absent from the Classe Préparatoires curriculum at the time) and to the elegance of mathematics as a tool for abstracting the living world.
This continued with the staff of the CRIOBE and Serge Planes, my thesis director, with whom I started studying the dispersal and behaviour of fish larvae, which occupied the first 10 years of my scientific career.
During this time, Claire Paris played an important role: she contributed to the supervision of my thesis, hosted me as the first post-doc in her new laboratory, contributed to coordinate with me the first major project which I got funded after getting my current position in Villefranche and finally selected the first PhD student I supervised for a post-doctoral position in her lab; life has come full circle.
For 10 years already (ouch…), my scientific environment has been the Laboratoire d'Océanographie de Villefranche (LOV). And I feel extremely good there. I found a human-sized yet diversified laboratory; within

which people get along, overall, and share ideas as easily as they share laughter; where I sit next to researchers at the cutting edge of their discipline, who inspire me, yet who did not hesitate to integrate me in their projects or even ask advice form me, a newcomer and relatively novice oceanographer. This working atmosphere is rare and precious. I would like to thank all the staff of the lab for contributing to it, and in particular Antoine Sciandra, Rodolphe Lemée and Fabrizio d'Ortenzio, the past and current directors of LOV, for maintaining it.

More recently, the Pelagic Ecosystem ProcesseS (PEPS) and, now, the COMPutational PLankton Ecology (COMPLEx) teams have taken an important part in my scientific activity, since I took charge of them. I thank all their members for their energy and for supporting me in transitioning to new research themes within the team, in particular the study of plankton ecology through imaging, which is now my main activity.

I try to draw inspiration from the qualities of all these people for the supervision of the students I host. These students were numerous and I hope that I succeeded for some of them at least. They were the people with whom I interacted the most and, therefore, they contributed largely to the work presented in these pages. I am grateful to them and hope they remember the weeks / months spent in my company as a good experience. Among them, Robin Faillettaz held a special place: he was my first PhD student, as part of my first funded project; his skills as well as his relational qualities made everything run smoothly. In the next three years, Cédric Dubois, Laure Vilgrain, and Thelma Panaïotis, the three PhD students I am currently co-supervising, will take this special place. I feel extremely fortunate to have worked and continue to work with such motivated and nice students. Thank you!

Finally, my work would not be satisfying if it was not balanced by a fulfilling personal life. Estelle, Gabin, and Louise, you are the ones who fulfil it, with joy. I know that you probably often tell yourselves that the balance between work and play is not in your favour, and I apologise for that. Yes, this job is as time-consuming as it is exciting and my mind is too often elsewhere when it should be with you. I hope that you will find in this document a concrete manifestation of these hours spent far from you; know that without you all, without you Estelle, none of this would make sense. I love you.

Contents

List of Figures

Chapter 1

Introduction

Among natural sciences, ecology has always been quite mathematically-oriented, and those mathematical foundations were often infused from other disciplines.

Population dynamics as differential equations

For example, population ecology concerns itself with the description of the variations in the abundance of individuals over time and space. Its theoretical bases were actually laid out in the 19th century, by Malthus, an economist, who defined the concept of population and its exponential growth. The interaction between two populations, such as those of a predator and a prey, were later formalised by Lotka and Volterra, a mathematical chemist and a mathematical physicist. The consequences of these population dynamics on the evolution of biological traits, such as body size, can be represented through the adaptive dynamics framework of J. A. J. Metz, a mathematician by training.

Behavioural ecology as an optimisation problem

The study of why animals do what they do (behavioural ecology) was accelerated by the definition of the optimal foraging theory by MacArthur and Pianka. Under this theory, the foraging strategies of animals should maximise the ratio of energy gained over energy expended, given their current environment. Mathematically, this is a problem of maximisation under constraints that echoes similar problems in other fields, such as the simulation of the choices of economical agents. In economics, the driving force is the market, in ecology it is natural selection. When the optimal strategy for one individual depends on what the other members of the populations are doing, game theory (which was also borrowed from economics) can help explain why some behaviours that are apparently counter-selective are still maintained through evolution, such as altruism. In both cases, the development of these economic theories through ecological problems later perfused back into economics.

Formalisation to compensate data scarcity

These are just two examples of fields of ecology which benefited from mathematical formalisation. These theoretical frameworks helped make the most of a relative scarcity of data, especially experimental data that allows to test hypotheses. Indeed, compared to other fields of biology, experiments are particularly difficult to carry out in ecology. For example, while it is possible to trigger and observe a physiological response to a stimulus *in vitro*, the stimulus and its response in ecology usually occur over days to months and several hundred meters square. This makes them difficult to manipulate and observe.

Numerical ecology deals with distance matrices

Once more data becomes available, the focus turned to correctly describing it. In ecology, the data is often counts of various species. Their variations among locations/times need to be described and, often, related to some potentially explanatory environmental factors. For example, the composition of the algal community is not uniform along depth and can be related to factors such as tidal level and light intensity. The mathematical tool of choice for ecologists in this case is the distance matrix: the differences in species composition among all couples of samples are quantified using a multivariate distance metric (Euclidean, Bray-Curtis, etc.). The distance matrix hence computed can be (i) represented and interpreted in a simplified space, through Principal Component Analysis for example, and (ii) related to environmental variables through regression on these environmental variables (in Redundancy Analysis for example). Variations around this theme are numerous and gave rise the to field of *numerical ecology*, which has succeeded thanks to well laid out theoretical foundations (in Legendre & Legendre, Numerical Ecology) and increasingly available software implementations (such as packages for the R language).

Computational methods repeat many simple operations…

…and tune them through trial and error

Currently, large amounts of data are made available to ecologists, thanks to the development of news instruments (e.g. digital imaging sensors) and new lab techniques (e.g. next generation sequencing). Concurrently, computers have become orders of magnitude faster in a couple decades. This has led to the rise of *computational ecology*, whereby mathematical equations and analytical solutions have been partly replaced by algorithmic solutions that can only be resolved through massive amounts of computation. Good examples of such algorithms are bootstrap aggregating (= bagging) or gradient boosting. Both are based on the repetition and averaging of simple operations (a classification, a regression, etc.) and the key to their performance is often some implementation "hack" such as dropping half the data randomly at each iteration to increase the generality of the solution. Such techniques can only succeed in a situation where computational power and data are in ample supply but, when they are, they tend to outperform all other approaches. The current poster child for these techniques are surely deep Convolutional Neural Networks (CNNs), typically applied on images. To summarise them grossly, they amount to the weighted combination of huge amounts of matrix multiplications and maxima, with weights initialised at random and progressively tuned to give the expected answer on training data. Here also, implementation "hacks", such as various ways to combine convolutions and pooling or to transmit data from one layer to the next through activation functions, have led to significant performance improvements. And overalls, these CNNs have set a new bar in many image classification problems. This general idea of “try a bit of everything and keep what works” is central to computational approaches and while it may not be particularly fancy, it is extremely efficient.

Research at the intersection of ecology and computer sciences

I have used all of those techniques in my research, from mathematical modelling to data mining through numerical techniques and now computational ones. This latter shift has occurred mostly because I have always been interested in computer and programming and this put me in a position where I was able to exploit machine learning tools. To be honest, this was never a part of a grand strategy, anticipating the fact that computing skills would become essential for scientific research, like they have in many fields. No. I just like computers. I like writing a piece of code that solves a problem efficiently. I like combining it with others to support a string of reasoning and provide an interpretable result. I find this elegant and empowering, hence stimulating. The potential novelty of the ecological result could be just icing on the cake. But then again, the icing is the most important part of a cake: it is what you choose it by and what whets your appetite! So in the following, I will present my various research endeavours according to the ecological questions they are tackling. At the same time, I will strive to highlight how the basic ingredients underlying these different research activities are often the same: the application and development of mathematical/numerical/computational techniques appropriate to solve the problem at hand.

Chapter 2

Previous activities

2.1 Behavioural ecology of fish larvae

The vast majority of marine organisms produce larvae who spend a few hours to a few months living in the water column of the open ocean (the pelagic domain; Figure 2.1). Those larvae are often quite different from the adult form in terms of appearance, diet, activity, etc. Early on in the 20th century, this larval phase was recognised as critical for the renewal of stocks: larvae are numerous and a slight change in their survival probability has a multiplicative effect on the future abundance of adults[1]. In addition, for coastal organisms whose adults live in close relationship (demersal) or fixed to (benthic) the substrate, this pelagic phase is the sole opportunity for larger scale dispersal[2]. Hence it conditions connectivity among coastal populations which is crucial for the efficacy of marine protected area networks, for spatial stock management, and for the distribution of genetic diversity.

The larval phase is critical for stocks and connectivity…

Early on, these larvae were considered to be plankton, i.e. organisms that drift freely with ocean currents[3]. Research focus was therefore put on the environmental conditions (currents, influx of nutrients, timing of the bloom[4], etc.) that facilitate their feeding and keep them close to suitable nursing grounds, hence improve their survival. Numerical models of their pelagic drift with currents were use to compute their potential spread and infer connectivity scales[5]. But experiments first performed on coral reef fishes in the mid 1990's, and later generalised to other species, showed that larvae do have significant behavioural abilities[6]. This had important implication for such models.

…and larval behaviour can determine its outcome

It is in this context that I started my first research internship and then my PhD thesis. The first review of this body of research was actually published in 2006, right in the middle of my PhD; it defined fish larvae as nekton (i.e. organisms that actively aggregate and/or swim), rather than plankton[7].

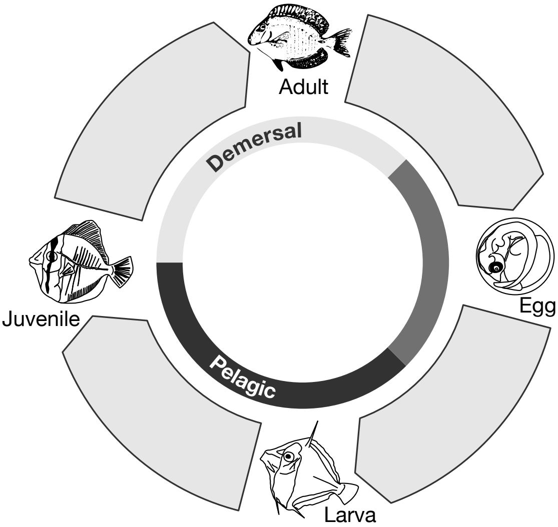


Figure 2.1 The life history of a typical demersal fish: adults produce eggs that can be benthic of pelagic but, in almost all cases, larvae are pelagic before they metamorphose into juveniles that settle on the coast.

2.1.1 Fish larvae can swim

To assess their swimming abilities, individual fish larvae are put in a flow-through tunnel in which the current speed is controlled. This speed is progressively increased until the fish cannot hold its position in the tunnel[8]. This measurement, called critical speed, is standardised and comparable among species, and has been proven to be a good way to estimate both natural, in situ, swimming speeds and forced swimming abilities. When measured on various species of, mostly, coral reef fishes, it revealed that larvae swam at speeds on the order of or greater than coastal current speeds (10 to 20 cms^{-1}; Figure 2.2); they cannot be considered as plankton under these conditions[7].

Coral reef fish larvae swim fast

But the vast majority of those measurements were performed on coral reef fishes and it was argued that these results were specific to these particular, very warm, waters[7]. Indeed, when cold temperate species were studied, they displayed much lower critical speeds. However, cold temperate environments are dominated by Gadiformes (e.g. cod, *Gadus morhua*) and Pleuronectiformes (e.g. sole, *Solea solea*) while tropical environments are dominated by Perciformes, so the comparison is not straightforward.

Warm temperate environments, like the Mediterranean Sea, are also dominated by Perciformes (e.g. seabreams such as species of the genus *Diplodus*) and very little data was available on these species, none in the Mediterranean Sea[9]. We measured the critical swimming speed of 153 settlement-stage larvae in the bay of Villefranche and it ranged between 10 and 40 cm s^{-1} (Figure 2.3). For species that were comparable with trop-

We proved the same for Mediterranean fish larvae

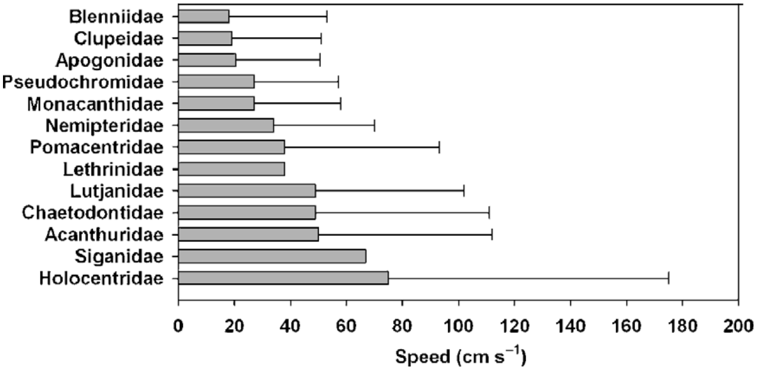


Figure 2.2 Critical speeds of settlement-stage reef-fish larvae of 12 families from the Indo-Pacific and Caribbeans (from Leis 2006[7]).

ical counter parts, the speeds of the Mediterranean species were slightly lower. Yet, they were on the the order of current speeds in the region: the fastest local current (the North Current) flows along the shore at 25-35 cm s^{-1} on average[10]. Mediterranean fish larvae are therefore competent swimmers, at least at the end of their pelagic episode, and, like tropical fish larvae, cannot be considered as passive.

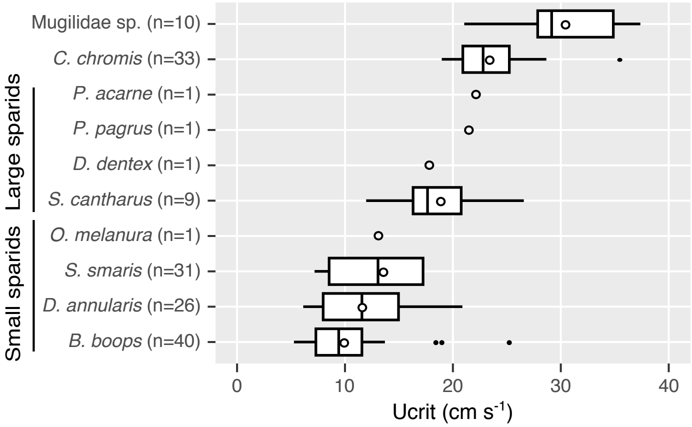


Figure 2.3 Critical swimming speeds of Mediterranean settlement stage larvae, in cm s^{-1}. The mean values (empty circle) are overlaid on top of standard boxplots.

2.1.2 Fish larvae can orient

Swimming is oriented

If larvae swam fast but at random, that would only increase their spread but not fundamentally change connectivity patterns. However, observations of the swimming direction of tropical fish larvae in situ showed their behaviour was much more complex. To gather such observations, settlement-stage fish larvae were caught in the Great Barrier Reef, released away from shore, and followed by two scuba divers, who recorded their bearing with a compass[6]. Although somewhat crude, this method worked very well for some species and provided unequivocal evidence that most larvae keep a bearing when swimming, and orient cardinally. It also suggested that this orientation varied with time of day.

An in situ device allows to quantify this orientation

To observe this behaviour with less human presence, yet in a more controlled environment, which would allow to manipulate the cues that larvae could use for orientating, we designed a new in situ observation instrument (Figure 2.4)[11]. Its principles are: (i) larvae are placed in a circular chamber and the instrument rotates slowly, meaning that, to keep a bearing, the larvae have to move along the circular wall; (ii) the whole instrument is locked in its surrounding water mass by a sea anchor, to avoid creating a current in the chamber which would influence the swimming direction of the larva; (iii) the behavioural chamber can be transparent or opaque, made of mesh or sealed, etc. to test various orientation cues.

After several design iterations it allowed to prove that…

The initial version sported an aluminium frame which integrated the sea anchor; we used a video camera that looked down on the experimental chamber and on an analog compass to record the device's rotation. The data processing, which I wrote, involved extracting stills from the video, correcting them to account for severe light changes and vibrations, and finally semi-automatically detect the position of the larva on successive frames[12]. This proof of concept showed that larvae of the family Pomacentridae also oriented significantly, along the coast of Florida.

…larvae respond to odours…

…orient better in groups…

In the next version, the frame was made of acrylic to be transparent to light and sound (and reduce vibration), the device was turned upside down so that larvae were directly exposed to natural light, and the sea anchor was separated from the frame. Electronics-wise, we used a still DSLR camera coupled to a time lapse device as well as a digital compass. I adapted the software to these new, more precise, inputs. This new device allowed to detect, for the first time in situ, that fish larvae react to odours emanating from neighbouring coral reefs by changing their swimming regime and direction[13]. It was also compared to the "scuba-following" method and provided consistent results[14]. Finally, by observing several larvae placed simultaneously, we showed that groups orient better than single individuals[15] (Figure 2.5). The same was observed for migrating birds, for example, with explanations ranging for probabilistic (averaging out of individual errors[16]) to cognitive (one individual "knows better" than the others[17]). However, it was the first time that such a behaviour was observed this early in ontogeny, in organisms with probably reduced cognitive abilities; this favours purely probabilistic explanations.

The current version of the instrument is essentially the same but smaller, using a shutter-less camera, an Arduino-based compass and a new model of drogue. Software-wise, it evolved from a byzantine collection of shell and Java scripts into a more robust package for the R language.

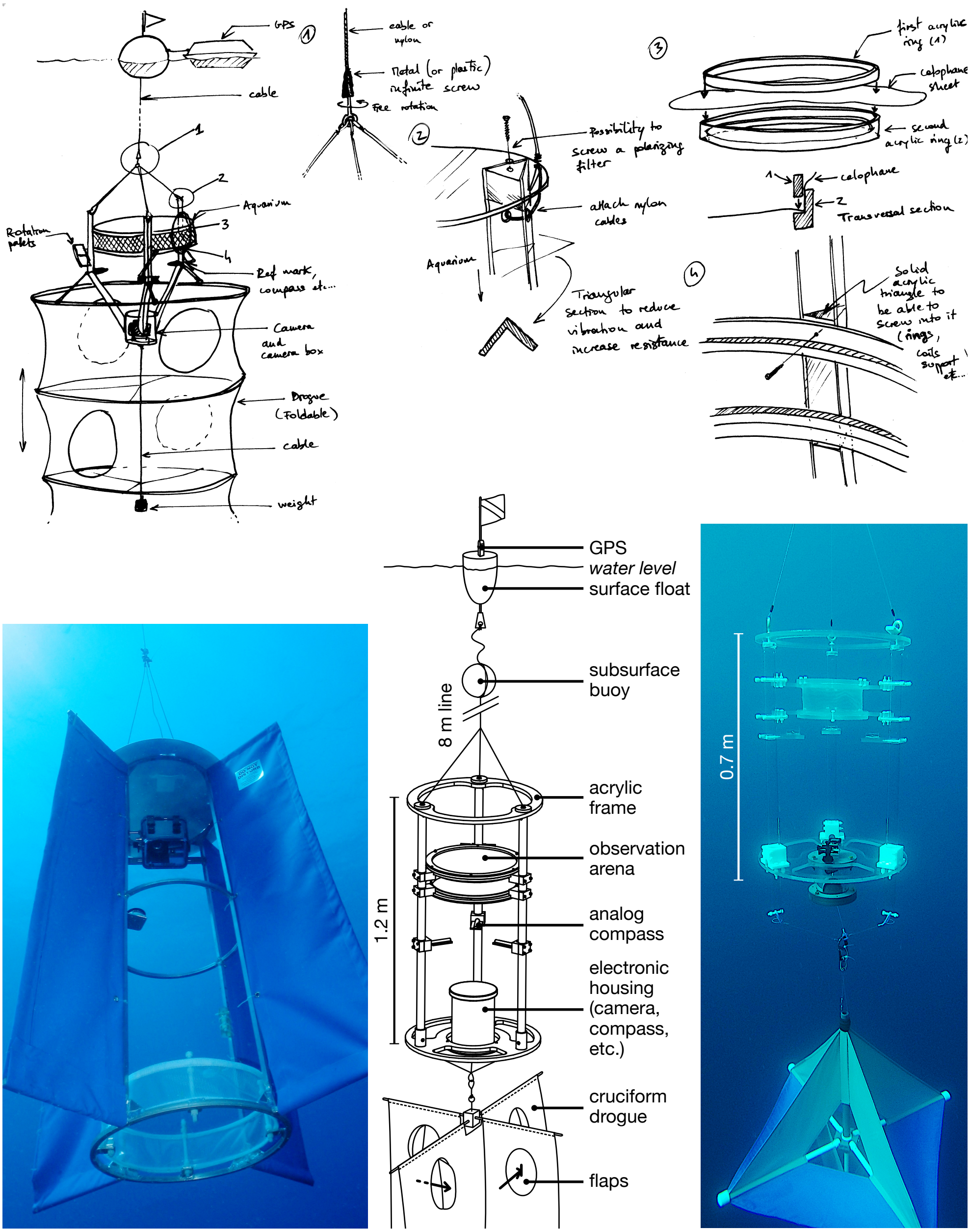


Figure 2.4 Several iterations on the design of the DISC instrument. Bottom left: first prototype with aluminium frame, integrated drogue and a downward pointing video camera. Top: my drawings for a circular, acrylic-based design with a separate holey-sock type drogue. Bottom centre: final version of the circular design with a DSLR camera housing. Bottom right: current smaller version of the instrument with a GoPro camera and a star-shaped drogue.

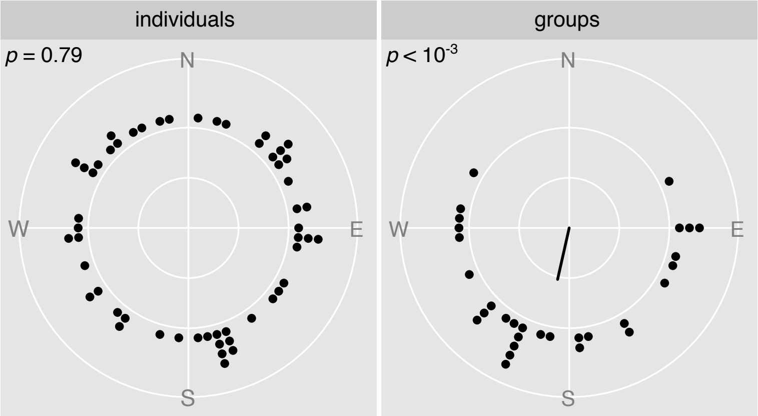


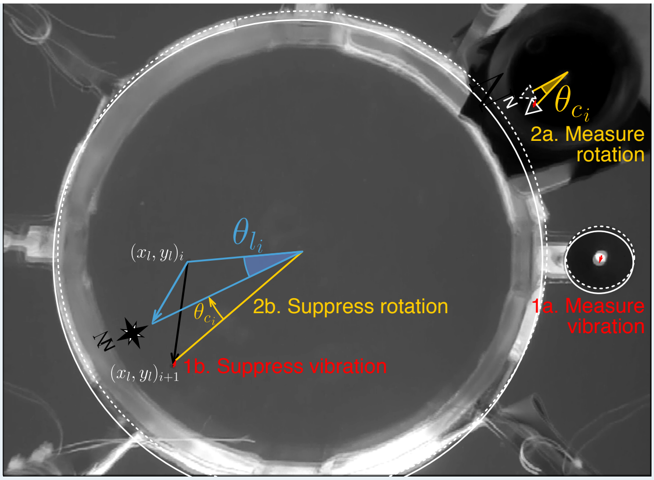
Figure 2.5 Cardinal orientation of individual and groups of *Chromis viridis* larvae. Each dot is the average bearing in one experiment (either of one fish or of a group of them). For single individuals (left), bearings are spread all around; for groups (right), bearings are significantly concentrated, denoting orientation, to the South.

…and orient relative to the sun

Its reduced bulk allowed us to deploy several instruments at once in the Mediterranean Sea and collect the first orientation data in four Mediterranean species. The large number of experiments, carried out throughout the day and over several months allowed to show that larvae oriented with respect to the sun's azimuth, hence providing a potential mechanism for large scale orientation, with no geographical point of reference[18].

Image processing in a circular reference

The devices presented above capture images. In those images, the information lies in the position of the larva within the circular chamber, particularly in its angle relative to the centre, which is considered as its bearing. Because the camera rotates with the instrument, a compass is required to provide a frame of reference and rotate images so that North is in the same direction in all (and points up). Actually, rather than rotating the images themselves, the detected positions of the larvae are rotated around the centre of the chamber. As shown below, on the superposition of two successive frames of version 1 of the instrument, this involves detecting the chamber position and its centre, shifting it between frames to correct for vibration, detecting the position of the larva as (x_l, y_l) on the image and converting it to polar coordinates around the centre of the chamber (ϕ_l, ρ_l), detecting the centre and needle of the compass, computing the rotation of the needle between frames $i+1$ and i : $\theta_{c_i} = \phi_{c_{i+1}} - \phi_{c_i}$ (or compute this from the record of a digital compass), and subtracting that angle from the angular component of the larva coordinate $\theta_{l_i} = \phi_{l_{i+1}} - \phi_{l_i} - \theta_{c_i}$. The data then becomes the sequence of angles θ_{l_i}.



Angles have the particularity that $0 = 2\pi$. While the statistical theory for such circular data exists[19], it is certainly less developed than regular, linear, statistics; implementations are often lacking in particular. Furthermore the θ_{l_i} are recorded in a sequence and therefore not independent from each other. To compensate for this non-independence, initially, 1000 bootstrapped subsets of the data were taken at various intervals and the Rayleigh statistic (the norm of the vector sum of the angles) was computed to assess the concentration of positions around a given bearing. Based on these tests, we now sub-sample images directly, at 5 to 10 seconds intervals. To further process this data, I implemented various functions to complement the circular package for R: to assess the dispersion of angles (circular standard deviation, Wallraff test) and compare two sets of angles (Watson-Wheeler test, Watson-Williams test).

2.1.3 Active swimming changes the outcome of Lagrangian dispersal models

Modelling can help understand the hard-to-sample dispersal phase

Fish larvae cannot be physically followed during the days to months they spend in the pelagic realm. Numerical models are often used to gain an understanding of this hard-to-sample phase[20]. In very broad terms: oceanic currents and hydrological conditions (temperature, etc.) are simulated through approximations of the Navier-Stokes equation in a discretised 3D domain; then, virtual larvae are released in this simulated environment and their individual displacement by ocean currents as well as ecological history is computed.

In a best-practices research report, we examined how such models can be used to simulate connectivity among coastal populations[21]. In addition to a realistic current field, the minimum ingredients were concluded to be good knowledge of spawning times and locations, of the maximum duration of the pelagic period, and of the potential settlement grounds for larvae. The second step would be to add vertical migration by fish larvae, which alters trajectories because currents are different near the surface and at depth, and spatially explicit mortality. The research needs

More effort on small scales and behaviour are needed

identified in 2009 were (i) a better definition of oceanic currents near spawning grounds, because the initial direction of dispersal has important consequences on the rest of the trajectory and (ii) more consideration of biological processes, in particular through the collection of more data on in situ mortality rates and behavioural abilities of larvae, which were extremely scarce. Since then, hydrodynamical models have progressed but mortality still remains a big unknown.

Extreme, individual-level behaviour must be considered

As for behaviour, it was considered at length in a dedicated section of the same report[22]. We highlighted that considering average behavioural abilities is not sufficient, because the extreme, most capable, individuals are likely the only ones that actually survive. Nor is it appropriate to infer behaviour from taxonomically-related species, because considerable variability exists among species. For each behavioural trait (vertical position, horizontal swimming, orientation, schooling, etc.) we designed simple tests to check its potential influence in a simulation of interest. Then we explained how to collect data, implement the behaviour in the model, and parameterise it according to the data collected. When these simple tests were conducted, they systematically showed a large potential influence of larval behaviour on dispersal outcomes, and underlined that, for most behaviours, data was lacking to correctly set parameter values. This is in part what prompted the research described in sections 2.1.1 and 2.1.2 above.

Modelling swimming larvae increases settlement…

After implementing the swimming and orientation observations collected in Villefranche into a model of the Ligurian sea, we showed that active swimming over the last four days of the pelagic phase was sufficient to completely change the probability of recruitment (Figure 2.6)[23]. Swimming at only 2.5 cm s^{-1} already doubled the recruitment rate; for the fastest swimming larvae, it multiplied it by ten. The influence of current patterns (such as eddies) was still perceivable though, meaning that models that combine a precise description of the hydrodynamics and include well parameterised behaviour are indeed necessary.

Starting from this parameterisation, we also modelled the full larval period (from spawning to settlement) for small and large Sparidae (Figure 2.3) to investigate how connectivity patterns between Marine Protected Areas (MPAs) change when behaviour is introduced[24]. Small Sparidae spend 13 to 19 days as larvae, which swim at about 6 cm s^{-1} at the end of their larval phase. Large Sparidae develop for longer (28 to 38 days) and swim faster in the end (10 cm s^{-1}). In both cases however, introducing active swimming in the model increased the intensity of connections (Figure 2.7), because it increased overall recruitment (as above), but also modified the patterns of retention around the release zone. In most cases, adding swimming increased retention or, in other words, decreased the spatial scale of connectivity (Figure 2.8).

…and modifies connectivity

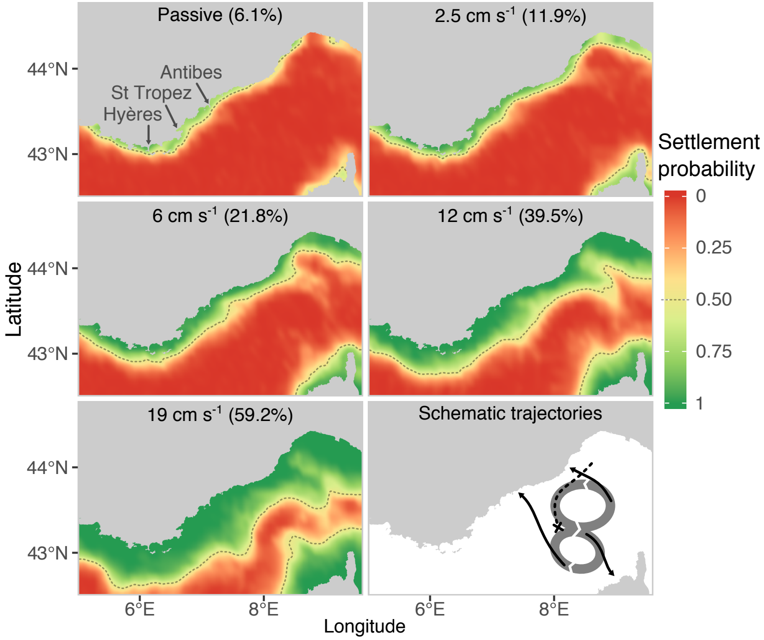


Figure 2.6 Maps of the probability for larvae to settle on the coast in the next four days, averaged over one month of simulation in June (period of maximal larval abundance). Larvae are assumed to have large scale orientation and swimming abilities. They swim towards the coast at speeds inferred from the measurements in Section 2.1.1. In the bottom right, the main current features over the month of simulation (two stable eddies) are schematised and their influence on larval trajectories is represented by the black arrows. This effect can be seen in the probability maps: it deforms the 0.5 isoline.

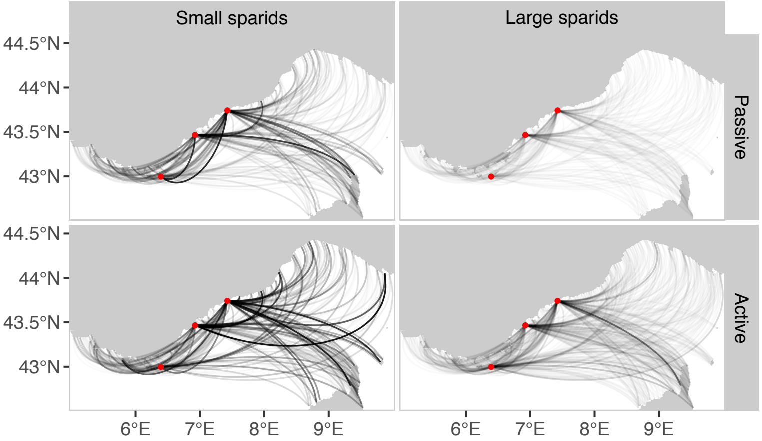


Figure 2.7 Simulated connectivity network from three MPAs in the north-western Mediterranean Sea. The intensity of the curves represents the intensity of the connections. Connections are more intense when active swimming towards the coast is added (bottom row).

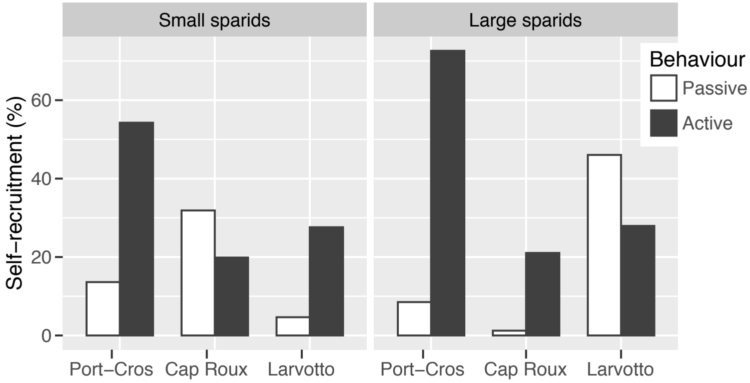


Figure 2.8 Self-recruitment rate in each MPA: proportion of the larvae settled in the MPA that were spawned in it. Self-recruitement often increases when swimming behaviour is added.

Implementing swimming in a Lagrangian model

A Lagrangian particle tracking model computes the trajectory of a particle as it is carried by currents. Its general equation is

$$\frac{d\vec{x}}{dt} = \vec{u} + \vec{u}'$$

where $\vec{x}$ is the position (potentially 3D), $\vec{u}$ is the modelled current vector (same dimension as $\vec{x}$), and $\vec{u}'$ is a random current component, much smaller than $\vec{u}$, that can take several forms; typically a random walk. Indeed, the current field is computed on a discrete grid and some features of the current (very small eddies, turbulence, etc.) cannot be represented at the scale of that grid; these features would spread the particles more and this spread is captured by adding this random diffusion term.

Now, swimming particles could be modelled with

$$\frac{d\vec{x}}{dt} = \vec{u} + \vec{u}_s + \vec{u}'$$

where $\vec{u}_s$ is the swimming vector. However, larvae do not swim deterministically, always in the same direction, over the duration of the model integration time step (typically around 1 hour). They rather swim towards a general direction, with more or less precision. This process can be represented by a *correlated* random walk, with two parameters: one determining the variability in speed, the other in orientation (or turning angles). Because $\vec{u}'$ is already a (uncorrelated) random walk term, the model is often simplified to

$$\frac{d\vec{x}}{dt} = \vec{u} + \vec{u}'_c$$

where $\vec{u}'_c$ is a correlated random walk process, the parameters of which represent both physical diffusion by sub-gridscale features and orientated swimming. Given the swimming speeds highlighted above (which are on the order of $\vec{u}$ rather than $\vec{u}'$), the physical diffusion term is often negligible.

2.1.4 Modelling optimal behaviour helps understand ecological processes

Despite the efforts documented above, data on fish larvae swimming and orientation is still scarce. In consequence, many questions remain regarding the mechanisms behind the observed behaviours: how are the different cues available for orientation used together? How do larvae choose their swimming speed? What determines the balance between directional swimming and foraging? In situations such as those, when the overall ecological setting is known but the detail of the behaviours are not, the theory of optimal behaviour has been an efficient way forward[25].

Optimal behaviour models capture distal mechanisms…

Let us take vertical positioning in the water column as an example. When the sun rises, many organisms migrate downwards. Often, the "proximal" cause (i.e. the how) of this behaviour is that organisms sense light and move away from it (they are photophobic). But the "distal" cause (i.e. the why) is that they risk being eaten by visual predators if they stay in an illuminated area, yet they still need to come to the surface because it is where most of the food is. Vertically migrating fish larvae certainly choose their swimming orientation and speed but they probably do so in direct reaction to stimuli from their environment, like light; it is unlikely that they consciously solve the trade off between staying near the surface, eating more but risking predation vs. migrating down, stop eating and hiding from surface predators. This trade off is in fact solved by natural selection: in a situation where photophobic and non-photosensitive organisms coexist, the photophobic organisms move down, suffer less predation, have more opportunities for reproduction, and therefore their proportion in the population increases over time, until only the "optimal" (or at least better) behaviour remains.

…that result from natural selection

When selective pressure is high, such models work

To predict what an organism will do in a given environment, one does not have to know all the proximal mechanisms, which can be very intricate, one "only" has to find which behaviour would yield highest survival and/or reproduction, and assume natural selection will have selected it. This is the theory of optimal behaviour. Of course, many caveats remain: constraints exists on what behaviours are feasible, all possible trade offs have to be considered, measuring survival and reproduction over the whole lifespan of the organisms of interest is almost impossible, etc. Yet, in situations where the trade offs can be simplified, the application of this theory has proven to be very effective[26].

And selective pressure is very high during the larval phase

The larval phase of coastal fishes is such a situation. There is no reproduction at this stage, so the only goal is surviving with maximal energy reserves, to be ready for the next step of the life cycle. Survival is very low (1 in a million larvae survive the pelagic phase on average[27]) and this is partly due to the fact that larvae who do not find a suitable habitat at the end die[28]. This puts enormous selective pressure on any mechanism that increases survival, in particular that allows finding a settlement habitat.

Models of optimal swimming…

…show that modest swimming speed can influence recruitment

Therefore, we built several models in which a dynamic environment was set up, larvae were parameterised with realistic maximum swimming speed, energy reserves and expenditures, and the optimisation target was to recruit back on the starting habitat (which is known to be suitable) with maximal remaining energy[29]. The trajectories of passively drifting and optimally swimming larvae were compared, over 25 days of dispersal, in two typical coastal environments where eddies are present and therefore passive retention is possible: an island and a promontory. Figure 2.9 shows 10 typical passive and active trajectories in the promontory case, which illustrate that only 1% of the passively drifting larvae were brought back to the suitable habitat while 95% of active tropical-type fish larvae returned. When cold-temperate larvae were simulated, with much lower swimming abilities, 72% of them still reached the target. Indeed, the examination of optimal swimming decisions showed that larvae always swam at very low speeds (2 cm s^{-1} on average, while the maximum speed of tropical larvae went as high as 30 cm s^{-1}). They swam mostly at the beginning of their dispersal and exploited the backward flow of eddies during the rest of the 25 days, hence saving energy. This suggested that even very moderate swimming abilities can have large consequences on dispersal trajectories and highlighted that initial dispersal is critical for the energetic cost of returning to the shore in the end.

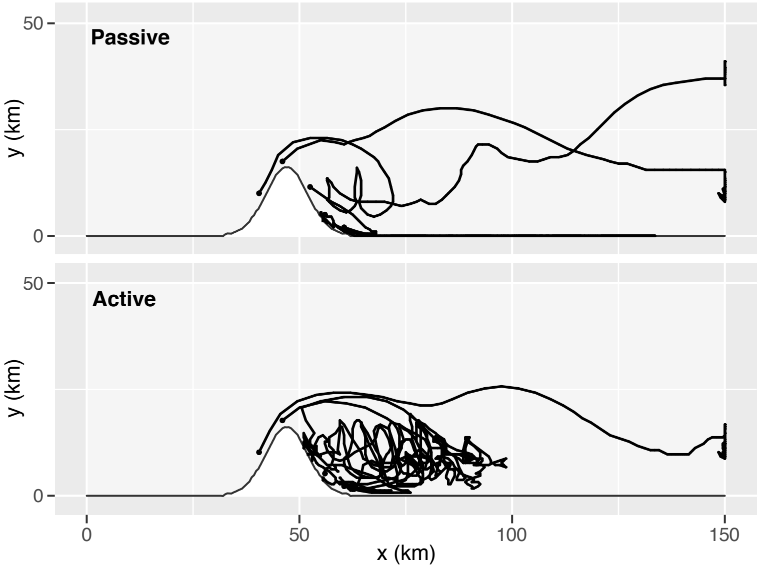


Figure 2.9 Top view of simulated dispersal trajectories of passive (top) and optimally swimming (bottom) larvae over 30 days. The domain is simulated with a uniform flow from the left side. Larvae are either entrained with (top) or trapped by (bottom) eddies that detach from the promontory.

A temperature increase influences swimming abilities, hence recruitment

The same model was used to simulate the effect of a 2ºC increase in water temperature, that accelerates the development of swimming abilities, reduces the duration of the larval period but also increases daily mortality rates[30]. In all situations, recruitment rates increased, mainly because larvae were subject to slightly higher mortality rates but over a shorter time and also because the earlier development of swimming abilities allowed them to better exploit the low flow regions in the lee of the topography and be retained near appropriate settlement sites (Figure 2.10). This is consistent with several other models predicting reduced connectivity scales under simulated climate change[31].

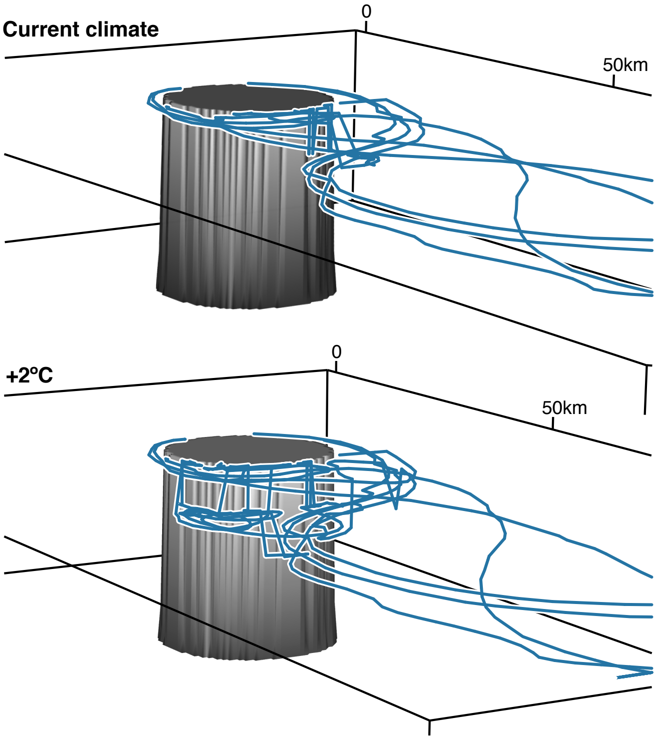


Figure 2.10 Simulated trajectories around an idealised island for optimally swimming cold-temperate fish larvae, in the present conditions (top) and after a 2ºC increase (bottom). The earlier onset of swimming allows a larger proportion of simulated larvae to move down, in the lee of the island, towards regions of low flow, and be retained close to suitable settlement locations, hence increasing the recruitment rate.

Dynamic programming to solve optimisation problems

To put the theory of optimal behaviour in practice, we need a way to find this optimal behaviour, i.e. the one optimising a quantity of interest in a given environment. In our case, we want to find the sequence of swimming decisions that allows to reach the coast with maximal energy reserves at the end of the larval phase. The brute-force way of solving this is to compute every single swimming decision at each time step for each larva, record the outcome of each, and, at the final time step, select the best sequence. But the computational cost of this is prohibitive, hence the use of dynamic programming.

Let us consider the simplified case of a unidimensional space (x, the distance from the coast) with larvae that have energy e. Larvae can take two decisions: (s) swim shoreward by δ_x^s and loose energy δ_e^s or (f) forage and gain energy δ_e^f but drift offshore by δ_e^f because of the current. We start by setting the "gain" of a larva at the final time step, $t = T$

$$G(x_T, e_T) = e_T \mathbf{1}_{\{x_T=0\}}$$

which means that the gain is e_T, the final energy, if the larva is on the coast at the end ($x_T = 0$) and 0 if not. Now, at each time step, the gain depends on what decision the larva takes (s or f) and where this decision will take it, in state space (x, e). Because we model optimal behaviour, we want the decision that maximises the gain

$$G(x_t, e_t) = \max(\, G(x_{t+1|s}, e_{t+1|s}) \,,\, G(x_{t+1|f}, e_{t+1|f}) \,)$$

in our simplified case, we can write it out as

$$G(x_t, e_t) = \max(\, G(x_t - \delta_x^s, e_t - \delta_e^s) \,,\, G(x_t + \delta_x^f, e_t + \delta_e^f) \,)$$

With this equation and the final gain, we can recursively compute the optimal decision at any time step. For example, at the last time step $T-1$, if the larva is within δ_x^s of the coast it should swim to it, if not both decisions provide gain zero. At the previous time step $T-2$: if the larva is within $[0, \delta_x^s - \delta_x^f]$ of the coast it should forage because it will get more energy and still be within δ_x^s of the coast at time $T-1$ (i.e. non-zero gain); if it is within $]\delta_x^s - \delta_x^f, 2\delta_x^s]$ it should swim to reach a non-zero gain position at time $T-1$; if is beyond $2\delta_x^s$, the decision does not matter because the gain is zero either way. This defines the gain landscape for $T-3$, etc.

This is Bellman's dynamic programming equation[32]. It still requires to compute the outcome of every decision at every time step, but only the optimal decision is kept and the optimal decision at time step t only depends on the gain at time step $t+1$, so it avoids the combinatory complexity of the brute force approach. It solves the optimisation problem elegantly and provides the decision sequence in addition to the optimal trajectory. Examining this decision sequence allowed to show that moderate swimming speeds were sufficient to strongly influence trajectories and to understand by which mechanism the changes in ontogeny dues to warmer waters influenced trajectories.

It has caveats though. The gain field has to be computed at each time step. So the state space has to be discretised and the outcome of every decision computed at each grid point. Therefore, the dimension of the state, the fineness of its discretisation, and the number of decisions influence computational cost in a multiplicative way, which imposes limits to their size. Furthermore, the decision field is not continuous (e.g. sharp changes between swimming and foraging occur in the case above) and, therefore, cannot be interpolated. So the approach is probably limited to theoretical study cases (such as the ones above) rather than large realistic simulations with a fine grid and many decisions. Yet, the whole premise of the optimal behaviour theory is to gain an understanding of the "why" of behavioural decisions rather than to predict their exact realisation.

2.2 Plankton and fish larvae distribution

In addition to looking into the behaviour of fish larvae and its potential consequences through modelling, I studied their in situ distribution and how it can be related to environmental factors, including the distribution of their potential planktonic preys.

2.2.1 Mesoscale features require detailed analyses

Grid sampling around an atoll…

We sampled fish larvae and plankton on two grids of stations around the atoll of Tetiaroa, with the goal of detecting its "island mass effect", i.e. the increased productivity in the leeward side of the island due to the eddies it creates within an otherwise steady flow[33]. The large grid comprised 36 stations distributed in a 50×50 km square, took three days to sample, and was repeated four times (Figure 2.11); the small grid were 16 stations distributed radially in the immediate vicinity of the atoll, took one day to sample, and was repeated twice.

…highlighted patchy distributions of planktonic groups…

…with no clear correlations

At both scales, the distribution of phytoplankton biomass inferred from fluorescence suggested an island mass effect but was shifting between repetitions of the sampling, which was not surprising since meteorological conditions (wind speed and direction in particular) also drastically changed. The distributions of zooplankton and fish larvae were patchy, very different among taxa and repetitions of the sampling (Figure 2.11), and apparently not related to environmental conditions or to each other. Even though sampling was perturbed by unstable meteorological conditions, the relationships between fish larvae, zooplankton, and water masses conditions should still have been somewhat preserved, even if the communities were shifting geographically. In this case, such relationships were not apparent when looked for at community-level, in depth-integrated samples, at mesoscale (i.e. scales on the order of kilometres).

Higher resolution sampling showed clearer patterns…

We also sampled fish larvae and plankton at stations spaced by 5km on average, sometimes closer, along a transect across the front created by the North Current in the north-western Mediterranean Sea (a front is a confrontation of two different water masses). In that case, the mesoscale frontal feature had an obvious effect on the concentration and community composition of fish larvae. Larvae of coastal fishes, in particular, seemed constrained to the coastal zone (Figure 2.12). The associated planktonic populations have not yet been fully examined and this is part of my research project (Section 4.3).

Finally, in another study[34] (for which I contributed to the data analysis only), zooplankton was sampled at very high resolution (km on the

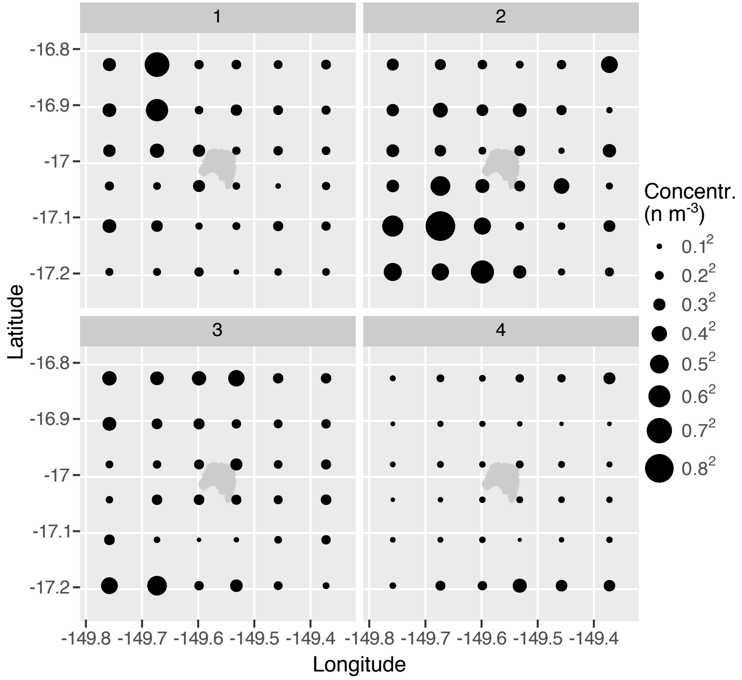


Figure 2.11 Distribution of reef fish larvae around the atoll of Tetiaroa over four repetitions on the same sampling grid.

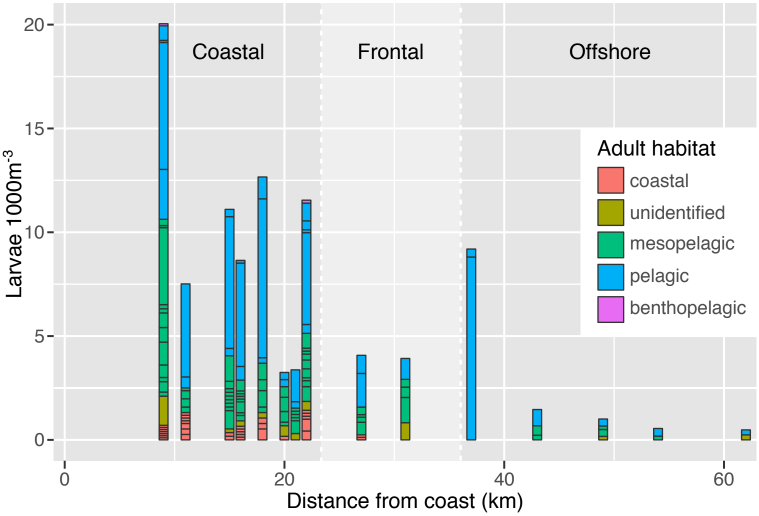


Figure 2.12 Concentration of fish larvae in depth-integrated samples from 100 m depth to the surface taken with a 680 μm meshed net during the VISUFRONT cruise. The transect starts on the coast (left), crosses the coastal zone, enters the main frontal area where the current is, and then moves out of the current into an offshore region.

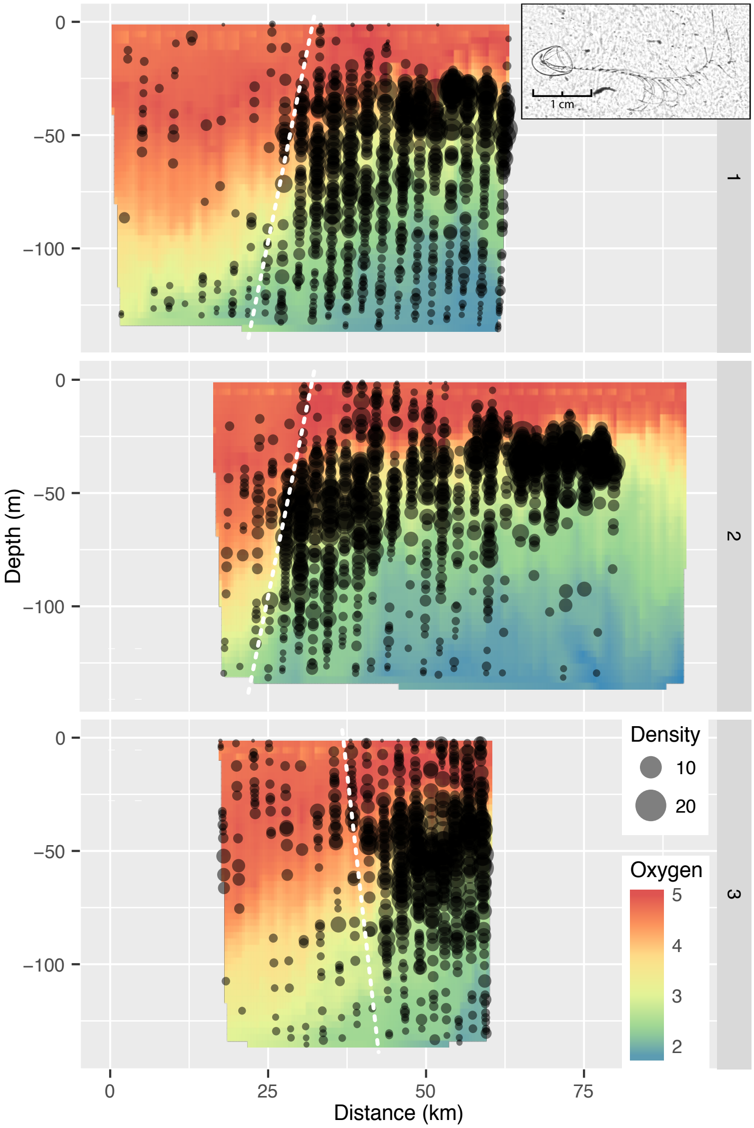


Figure 2.13 Distribution of *Sphaeronectes sp.*, a siphonophore, (black dots of area proportional to the concentration) overlaid on an interpolated map of oxygen concentration, over three ISIIS transects in the California Current. The dashed white line shows the approximate position of a front. Oxygen, temperature, depth, fluorescence, and salinity, explain 46% of the variance in the concentration of *Sphaeronectes*.

horizontal, cm on the vertical), by an in situ imaging instrument, along three transects crossing a sharp salinity-driven front in the Southern California Bight. At those scales, the relationships between the concentrations of gelatinous zooplankton and environmental variables were very clear, with R^2 of multivariate models often above 40% and up to 93% depending on the species.

…particularly at m scale, in 3D

So, overall, the effect of environmental forcing on zooplankton distribution can be very strong, as is well known at basin or worldwide scale[35]. But, in dynamic mesoscale structures, it appears that very high resolution data, resolving the details of the structures as well as the differences between taxa, is necessary to capture this effect. Such data is still scarce because sampling zooplankton is not yet as streamlined and automated as measuring hydrological or biogeochemical variables.

Submesoscale resolution is mandatory

2.2.2 Vertical distribution changes finely with age and time of day

As explained in the previous section, the vertical position of larvae is particularly important because it conditions their probability to find food or encounter predators, and because it alters their drifting trajectories since currents are often slower at depth than near the surface[36].

Fish larvae were sampled in a depth-stratified manner around Tetiaroa, using a Multiple Opening/Closing Net and Environmental Sensing System (MOCNESS: a frame containing several plankton nets that can be opened and closed at specific depths). By splitting the catches among preflexion (young), flexion (older), and postflexion (almost ready for settlement) stages, we could show that the depth distribution of many families spreads deeper when larvae grow older. This is often called "ontogenetic vertical *migration*", suggesting that the distribution shifts towards deeper waters. By using repeated sampling at many different depths, we were able to get more details on the distribution and show that the behaviour is clearly a spread, rather than a shift (except for Labridae); young larvae were often limited to the first 50 m while older ones exploited a larger range of depths (Figure 2.14).

Larvae spread, rather than migrate, down during ontogeny…

When these observed distributions were input in a Lagrangian dispersal model parameterised around an isolated island within a steady flow, as in Section 2.1.4, the dispersal distances were lower for vertically moving larvae than for larvae drifting always close to the surface, as expected. However, they were not much lower, which was surprising compared to the literature. This is probably because the other models represent the expected shift towards deeper, slower currents. When parameterised with the actual observations of a spread, old larvae were able to move within the full water column, hence experiencing all current speeds, not just the slow, deep, speeds.

…with limited consequences for dispersal

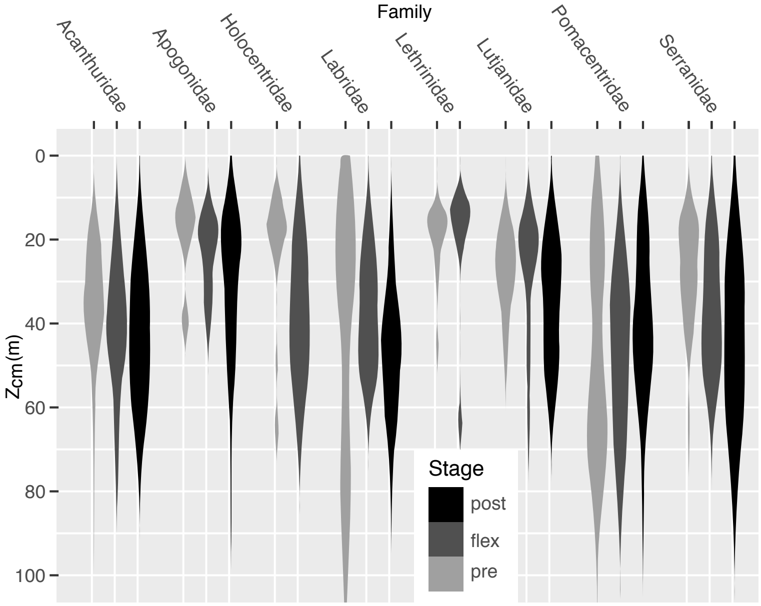


Figure 2.14 Violin plots of the vertical distribution of the depth centres of mass computed at each station for larvae from several families at three developmental stages. In most families, older larvae are spread over a wider depth range than young ones

Fine scale vertical distribution from coarse scale sampling

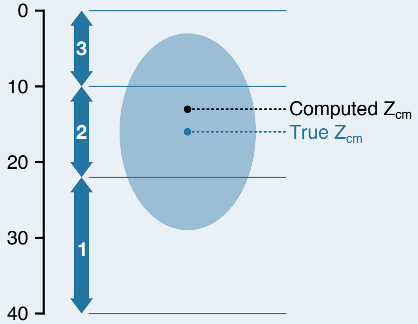
Given its ecological importance and the prevalence of diel vertical migration, many studies have investigated the vertical distribution of plankton[37]. In most cases, depth-stratifying nets (such as the MOCNESS[38]) were used. These yield concentrations organisms within depths bins typically spanning from 20 to 200m.

To compare distributions among conditions (e.g. between night and day), one can use distribution tests on the concentrations, such as the Kolmogorov-Smirnov test. However, the power of such tests is often directly proportional to the number of bins; and this is rarely more than 10 for plankton nets. In addition, this test is sensitive to aggregation in the underlying distribution and plankton is usually distributed in aggregated patches.

An alternative has been to compute the depth centre of mass (z_{cm}) of the plankton patch in each sample and then compare of those centres among conditions. The z_{cm} is the average of the depths of each net weighted by the concentration in it:

$$z_{cm} = \sum_i \frac{c_i}{\sum c_i} z_i$$

where c_i is the concentration in net i and z_i the depth of the middle of the layer sampled by net i. However, if the depth bins are not equally tall, this often-used formula is wrong, as shown in the simple example below.



The quantity of plankton is the same in nets 1 and 3 and the true z_{cm} is right between them. But, because net 1 samples a larger portion of "empty" water, the concentration in net 1 is lower than in net 3. Weighting the depths by concentrations biases the z_{cm} towards the depth of net 3. Instead, integrated abundances should be used

$$z_{cm} = \sum_i \frac{a_i}{\sum a_i} z_i \text{ where } a_i = c_i \cdot h_i$$

In addition, if the depths bins are the same at all stations, the estimates of z_{cm} are biased towards the middle of those bins and have a precision equal to the bin size. Let us take the extreme case of a thin layer of plankton, 1m thick, at 85 m depth, sampled by a net going from 100 m to 50 m. Its depth will always be estimated at the middle of the net bin: 75 m. If the bins are randomised among samples, or at least varied, the average of z_{cm} across samples would tend towards 85 m, rather than 75 m; the precision of the average therefore becomes finer than the thickness of the bins.
To get correct and accurate estimates of the depth distribution of fish larvae around Tetiaroa from only four nets bins between 105 m and the surface, bin limits were shifted by 5 m between stations, yielding four sets of bins that there distributed across stations, and integrated abundances were used to compute z_{cm}s. To our knowledge, this was the first sampling and analysis of this kind, despite the flaws in the usual procedures highlighted above.

Diel vertical migration can occur over very fine scales

Along North Current transect, in the north-western Mediterranean Sea, plankton was sampled by the same in situ imaging system as in the California Bight. It provides cm-scale vertical positions of the organisms. While the data is still being processed, enough has been sorted to provide accurate estimations of the distributions of various planktonic groups during the day and at night. This highlights the well known diel vertical migration of plankton, such as copepods. What is more interesting here is that this very fine scale data allows to detect migrations on the order of a couple meters, as is the case for Doliolids (Figure 2.15). Such fine scale behaviours were previously completely erased by the sampling precision of vertically stratified nets.

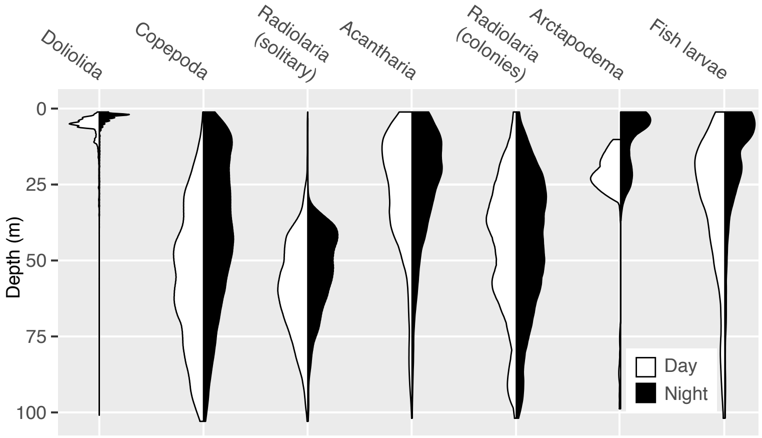


Figure 2.15 Day (white) and night (grey) distribution of various planktonic taxa collected over two transects across the North Current, in the north-western Mediterranean Sea. Most taxa are distributed shallower at night.

2.3 Plankton and nekton biogeography

2.3.1 Niche models predict basin-scale distribution of marine organisms

Patterns in environmental variables are more stable over large spatial scales than in the mesoscale structures studied above. For example, the latitudinal gradients in temperature and light dominate the variability of the surface layer of the oceans[35,39]. Therefore, the relationships between such habitat variables and the presence or abundance of a given species are often strong over these large scales. Ecological niche models[40] describe these macroecological relationships: the set of habitat conditions in which a given species is more present or abundant. They are also known as habitat suitability models.

Niche model capture macroecological patterns…

Data on the distribution of biological organisms is often scarce, with entire regions of the oceans that are not sampled because they are difficult to access. Environmental variables have better coverage, thanks to remote observation from satellites and unmanned instruments, accurate models, etc. Therefore, habitat models can help fill the gaps in our knowledge of the distribution of species. The model is fitted in regions where biological samples are numerous. It captures the conditions in which the species is abundant and those in which it is rare. The model predictions are then extrapolated to under-sampled regions. When the right environmental conditions are met, the species is assumed to be abundant.

…by correlating species observations with environmental conditions

We used this approach to describe the distribution of krill (Euphausiacea) and fish (Actinopterygii) for the Biogeographic Atlas of the Southern Ocean[41,42](http://atlas.biodiversity.aq). A set of 108 environmental variables layers (geomorphology, hydrology at various depths and sea-

They highlighted how fronts structure the distribution of Antarctic krill

sons, etc.) was extracted from a combination of in situ, satellite, and modelled information. A database of presence and absence records of each of the 13 species of krill and 374 species of fish was assembled thanks to the efforts of the CCAMLR. For species with enough records, niche models were built, allowing to describe their potential circumpolar distribution, while samples in this regions are often limited to the most accessible areas. The models often highlighted the dominant influence of the polar front for pelagic species and the sharp changes in the community of benthic, but also pelagic, fishes at the break of the Antarctic continent's shelf. For example, the distribution of antarctic krill (*Euphausia superba*, Figure 2.16) was clearly limited by the Atlantic polar front in the north; the model allowed to project its distribution in the Ross Sea (lower left quadrant of the map), where records are scarce.

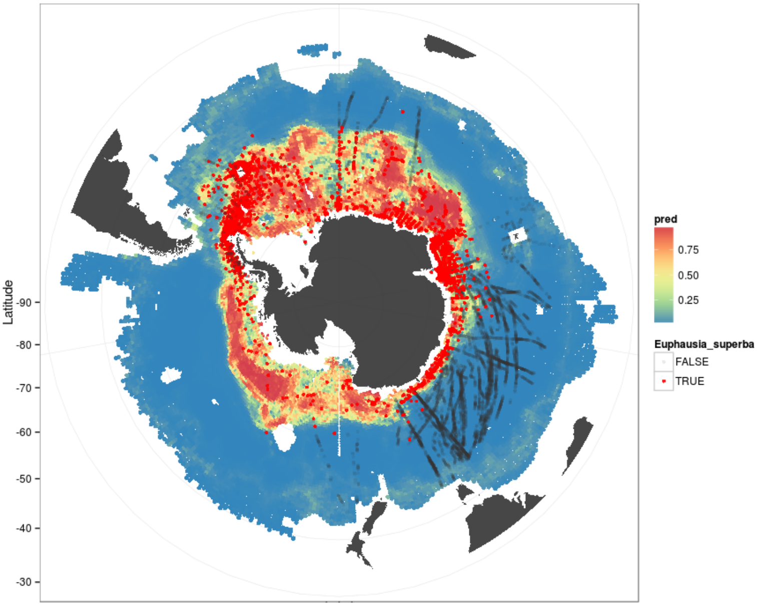


Figure 2.16 Presence (red) and absence (black) records of *Euphausia superba* overlaid on its probability of presence predicted mainly from silicate concentration at the surface, temperature at 200 m, and oxygen at 50 m. The model explained 63% of variance in the presence/absence record. Predictions are not extrapolated to regions where those environmental variables exceed the range in the observations, hence the ragged northern edge of the map.

Combine niche models output…

…and expert opinion…

Similar work was carried out in the Mediterranean Sea, where we modelled the distribution of 1250 species[43]. The modelled output were used in a regionalisation study (see below) and to investigate the provision of ecosystem services. Species were assembled into groups important for the functioning of ecosystems or the provision of some services. The contributions of each group to the services defined by the Millennium Ecosystem Assessment was estimated from expert knowledge through and in person and online survey (http://mesi.obs-vlfr.fr). Finally, a map of an index of provision of services over the Mediterranean Sea was computed by combining the modelled distribution and the estimated contributions. While the output generally made sense, it highlighted some corner cases that made us question the mathematical soundness of the synthetic index. In addition, this work was very divisive within the community: some applauded the effort while others considered it ridiculous, unscientific, and actually insulted us. It was never published. At least it allowed the Master student involved in the project to continue towards a PhD thesis in environmental economics.

…to map the provision of ecosystem services

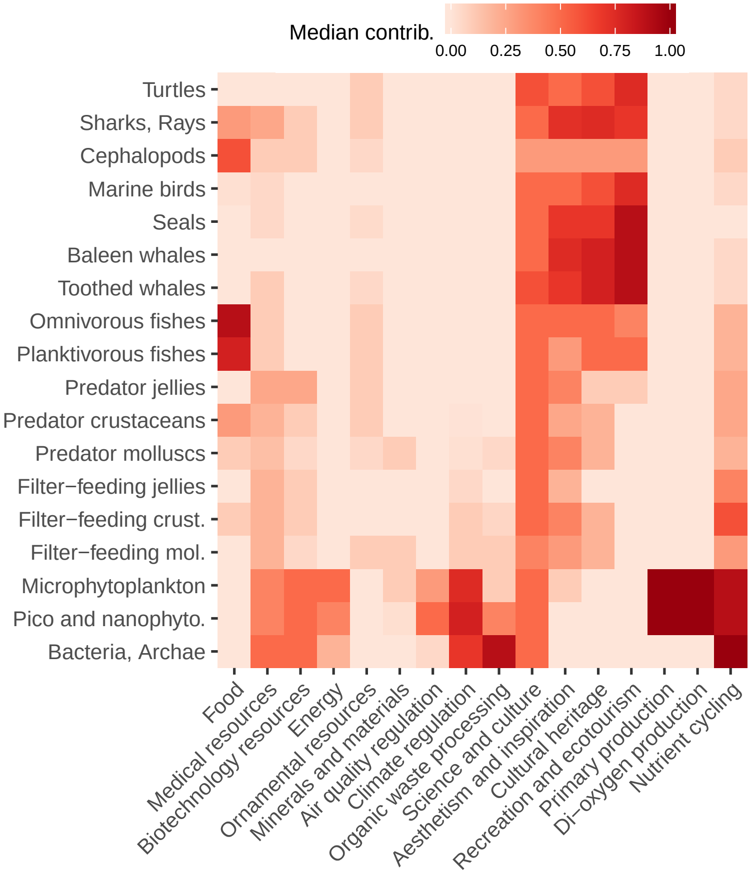


Figure 2.17 Matrix of the contributions of functional groups (rows) to ecosystem services (column) gathered through the survey (median of the answers of 70 participants weighted by their self-estimated knowledge of the groups).

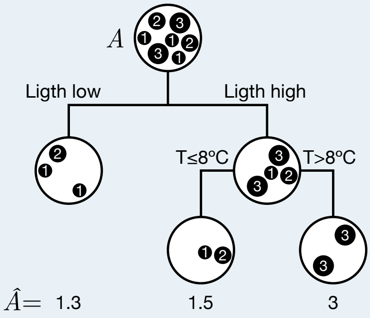
Finally, because climate models can predict the fields of environmental conditions in the future, a niche model fitted on the present distribution of a species can project its future distribution based on future fields. This comes with many caveats however. First, the uncertainties of the climate and niche models combine. When this was examined for copepods assemblages in the Mediterranean Sea, the choice of niche model had more influence on the predicted assemblage that the greenhouse gas emissions scenario, which shows how much uncertainty remains in the niche modelling[44]. Second, biological organisms may evolve and adapt to their changing environment rather than stay with the same preferenda and move geographically (or disappear). For species with short generation times, such adaptations can be fast. But to know whether and how this would happen, very long term experiments of artificial evolution have to be carried out and those are rare, so adaptation is neglected for lack of a better hypothesis. Third, even if organisms do not have time to adapt, their realised niche (the conditions in which they are currently observed) is different from their fundamental niche (the conditions in which they could be observed). This difference comes from biological interactions (presence of predators, etc.) or biogeographic barriers that do not allow the species to colonise all suitable habitat. The niche that matters for climate projections is the fundamental niche, which represents the physiological conditions the species can withstand. However, the realised niche is classically used because it is easier: outlining the fundamental niche requires physiological experiments while the realised niche can be deduced from existing data through the niche modelling process explained above. In a currently ongoing project, we are investigating the difference between using the realised or fundamental niches to project the future distribution of a toxic dinoflagellate (*Ostreopsis ovata* cf. *ovata*) along the coasts of the Northern Mediterranean, where it is already responsible for harmful algal blooms.

Fundamental, rather than realised, niche should be used for climate projections

Gradient boosting to model species-environment relationships

Species-environment relationships are often non-linear. For example, species have a temperature preferendum and cannot survive when temperatures are either too low or too high. They also involve complex interactions among environmental factors; for example, food concentration may become limiting only when temperature is high because metabolic rates are higher.

Regression trees are able to capture non-linear relationships and represent interactions naturally.



To model the effect of environmental factors on the abundance (A) of a species, the abundance records are recursively partitioned into homogeneous subsets according to the value of environmental factors (light and temperature here). The model estimate for the abundance in each subset ($\hat{A}$) is the average of the abundance records of the subset. The partitions are made as to minimise the sum of the within-subsets variances in abundance. By growing deep trees, one can model complex relationship shapes. The asymmetry in the tree, which occurs naturally, represents interactions; here temperature only has an effect when light levels are high.

Such models are versatile but may have limited predictive power. In addition, their predictions are stepwise (in the example above, predicted abundance changes abruptly when temperature crosses 8ºC) while the true relationships are smoother. Gradient boosting solves both problems: instead of growing one deep tree, it grows a long sequence of smaller ones, each working on the residuals of the previous. In addition, each tree is not fully "learned", meaning that its predictions are scaled by a shrinkage (or learning) rate, leaving more information in the residuals. This allows for a very large number of small trees (usually in the thousands), which results in smoother relationships (thousands of partitions instead of dozens) and a better fit to the data. In addition, to avoid overfitting and keep some generality in the model, only a fraction of the data (typically 50%) is used to grow a tree and a new fraction is randomly chosen for each tree. Therefore, only the strong relationships, still apparent in all 50% subsamples of the data, are learned.

To go beyond the predicted distribution and understand the correlative mechanisms modelled, one can use partial dependence plots: plots of the effect of changing one variable while holding the others constant. They are univariate approximations of a multivariate model but are often sufficient to show which temperature or salinity range a species prefers, for example.

Boosted Regression Trees (BRTs) with bootstrapped partial dependence plots were used to predict the niche of Antarctic species and document their preferenda. In addition, I setup a a web application that allowed anyone from the consortium of the Atlas of the Southern Ocean to upload occurrence data and test BRT modelling on their species of interest (https://github.com/jiho/atlasr); a screen capture of the user interface for the choice of settings is displayed on the next page. Technical choices were reduced to a minimum and set to meaningful defaults; most of the choices left were ecological (which environmental variable to consider, within which limits to predict, etc) and easily set by clicking buttons and moving sliders. This allowed experts of each biological group to assess the relevance of models and made for great interactions during consortium meetings. The software was heavily used for fishes (~350 species).

The same technique was used with great success to model fine scale species-environment relationships along the North Mediterranean Current front and California Current front (Figure 2.13), as well as in classification mode, to detect seismic signals (see Section 3.1 later).

Input data

Upload a csv file with columns for latitude and longitude and each species of interest

myctophid_data-csv.csv

Species to model

gymnoscopelus_braueri

☑ Bin input data on 0.1° grid

All data points will be kept but points in the same grid cell will be weighted down

Options

See ?gbm in R. For better predictions, you want a large number of trees and small shrinkage.

Maximum number of trees (the actual number will be estimated through cross validation)

1,000

Shrinkage per tree

0.03

☐ Advanced gbm settings

Number of bootstraps

0

Prediction

☑ Predict habitat suitability

☐ Predict beyond observed environmental range

Minimum variance for prediction

70

When some environmental data is missing, predict only at locations where the available environmental data allows to capture the specified percentage of the variance

Latitudinal limits and step

-80 -30

2

Longitudinal limits and step

-180 180

2

Explanatory variables

☑ Prefer interpolated variables

☐ Allow 'distance from ...' variables

Restrict to depths:

☑ 0
☐ 50
☑ 200
☐ 250
☐ 500

Restrict to season:

☑ summer
☐ winter

Select variables:

☐ bathymetry_slope
☐ bathymetry
☐ caisom_floor_current_speed
☐ caisom_floor_temperature
☐ caisom_floor_vertical_velocity
☑ chl_summer_climatology
☐ floor_temperature_interpolated
☐ geomorphology
☐ gridcell_area
☐ mixed_layer_depth_summer_climatology_interpolated
☐ nox_0_interpolated_summer_climatology
☐ nox_200_interpolated_summer_climatology
☑ oxygen_0_interpolated_summer_climatology
☑ oxygen_200_interpolated_summer_climatology
☐ salinity_0_interpolated_summer_climatology
☐ salinity_200_interpolated_summer_climatology
☐ seaice_gt85
☑ si_0_interpolated_summer_climatology
☑ si_200_interpolated_summer_climatology
☑ ssh_spatial_gradient
☐ ssh
☐ ssha_variability
☐ sst_spatial_gradient
☐ sst_summer_climatology
☐ surface_wind_annual
☐ t_0_interpolated_summer_climatology
☐ t_200_interpolated_summer_climatology

2.3.2 Regionalisation summarises geographical patterns

Regionalisation summarises biogeochemical/ecological patterns

One goal of biogeography is to provide a synthetic view of the spatial patterns of biogeochemistry and biodiversity. This can be done through regionalisation: the division of space into regions within which properties are quite homogeneous and among which they are different. For oceanic biogeochemistry, the most well-known regionalisation is that of Longhurst. But regionalisation efforts have mostly been global. Regional seas, such as the Mediterranean, are often left as one single unit. Yet, management policies, which would benefit from the simplification that regionalisation provides, are usually defined at regional scale.

Between 2009 and 2017, seven studies proposed regionalisations of the Mediterranean Sea. Most of them focused on physical or biogeochemical properties of the surface layer only.

So, first, we exploited 3D climatologies (i.e. average fields) of biogeochemical variables to propose a depth-stratified regionalisation of the Mediterranean[46] (Figure 2.18). Regions were delineated within the epi-, meso-, and bathy-pelagic layers, by combining four statistical clustering methods (see box later in this section). The Mediterranean Sea is dominated by a longitudinal gradient between warm and very oligotrophic (i.e. nutrient poor) waters in the East and colder and less oligotrophic waters in the West. This feature of course transpired in the regionalisations but its influence was very different among layers. For example, the separation of regions in the epipelagic layer was primarily influenced by patterns in temperature, chlorophyll content, thermocline intensity and bottom depth; in the mesopelagic one, some dominant variables differed. In particular, oxygen and phosphate concentrations, linked to the ingress of Atlantic waters in the Mediterranean, had a larger influence. By highlighting the difference between the surface layer and the other, less studied, deeper layers, this study put focus on the importance of acquiring data at depth.

Biogeochemical patterns change with depth in the Mediterranean

Second, we used the 1250 niche models of Mediterranean species[43] as input to the same clustering techniques to provide the first regionalisation of this sea based on ecological communities, rather than biogeochemistry. Species were initially split per trophic level (producers, consumers, secondary consumers, top predators) and then a combined regionalisation was proposed. This ecological regionalisation focuses on the surface layer because this is where most occurrence of species were recorded. When the epipelagic biogeochemical (bgc) regionalisation is compared to this ecological (eco) one (Figure 2.18), it can be noticed that eco-regions often subdivide bgc-regions. For example bgc-region 10 corresponds to eco-regions 14, 11, 13, 16, and 7; bgc-region 1 corresponds to eco-regions 17, 18, and 20; etc. These bgc-regions are in the transition zone between the oligotrophic east and the less oligotrophic west, and have “average” properties. Biological organisms seem to exploit these transitions and this creates different communities within those seemingly “average” regions.

Near the surface, specific communities exploit biogeochemical gradients

Finally, we tried to synthesise all these regionalisation efforts. They were based on different views of the Mediterranean Sea: phenology of the phytoplankton, biogeochemistry at the surface or in three dimensions, connectivity due to currents, ecological composition, etc. All could be of interest and a consensus was needed. To find it, we selected 8 non-redundant regionalisations among the 20 proposed, superposed the frontiers between the regions from each to compute a congruence metric, and smoothed it to remove noise and highlight general patterns. This allowed to delineate consensus regions, defined as having a low congruence core and being surrounded by strong frontiers; those were re-

Different regionalisations of the Mediterranean…

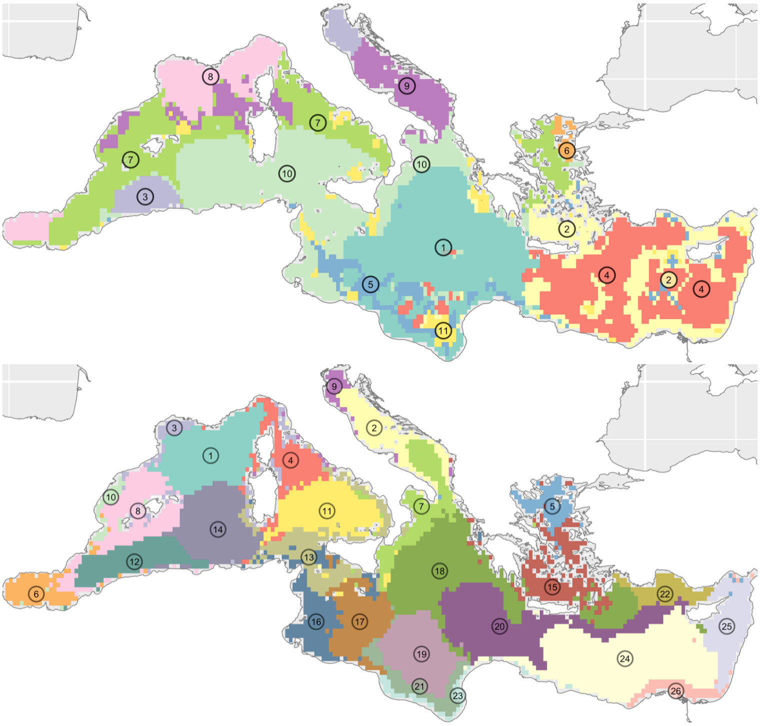


Figure 2.18 Biogeochemical regionalisation (top) and ecological regionalisation (bottom) of the surface layer of the Mediterranean Sea.

…still agree on some consistent regions and frontiers

gions that were found consistent for all variables (Figure 2.19). It also highlighted consensus frontiers, defined as having a high congruence core and some spatial extent; which were regions where all variables change simultaneously. This map shows that, while large areas of the Mediterranean Sea cannot be easily summarised (intermediate congruence values = inconsistency among regionalisations), some patterns still emerge and can form the basis of management units in the open sea: some regions and frontiers are consensual. When cumulated human impacts were mapped on top of these regions, most impacts were found at the regions boundaries, in dynamic zones (that are favourable for fishing for example). Therefore, the regions themselves can still be protected and potentially provide benefits to their immediate, highly impacted, surroundings.

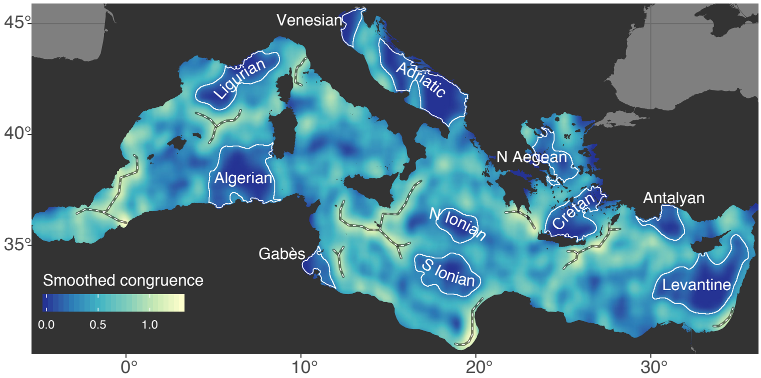


Figure 2.19 Consensus among eight regionalisations of the Mediterranean Sea. White polygons are consensus regions (low congruence core, strong frontiers). Dashed lined are consensus frontiers (large, high congruence core). This map and the underlying regionalisations are explorable in a dynamic web application at https://mermexregio.obs-vlfr.fr.

Congruence among clusters

Clustering algorithms classify each object (here a pixel in the Mediterranean Sea) into one cluster. Clusters are labelled (usually numbered) but these labels have no particular meaning or order. For methods with a random initialisation (such as k-means), they many even change between runs. Therefore, the labels cannot be used to compute, for each pixel, the consensus among various clustering methods. Yet, it is advisable to use several clustering methods simultaneously to avoid producing a result that is specific to the method rather than to the underlying data[47].

To produce the maps in Figure 2.18, four methods are used: k-means, c-means, agglomerative clustering with Ward linkage, and with complete linkage. In addition, for each method, various number of clusters are considered (from 2 to k, k being the optimal number of clusters for each method, computed based on a ratio of intra- to inter-cluster variance; the Figure Of Merit[48]). To find the consensus among all these clustering results, we compute the frontier pixels for each clustering, give them a score of one over a background of zeroes, and sum those frontiers. This gives a single raster layer comparable to an elevation map, with high hills in the regions where many clustering technique place a frontier and low plains where none cut. To define contiguous regions, an algorithm used to define watersheds (also called river catchment basins) was transposed to this clustering congruence “landscape”; it computes the direction in which water would flow if poured on top of the landscape and thus separates basins delimited by ridges. In our case, those basins are regions found to be homogeneous by all clustering techniques and separated by frontiers consensual among them.

This technique was also considered to compute the consensus among regionalisations in Figure 2.19. But we thought that the low agreement among the selected regionalisations (i.e. the fact that there are small hills almost everywhere rather than large plains separated by high ridges) was an important result, which computing contiguous regions would have masked.

2.4 Long-term ecological monitoring

One of the flagship scientific activities in Villefranche-sur-Mer is the long-term monitoring of the pelagic environment, which is easily accessible within or from the Villefranche Bay. The exploitation of this data was part of my job description when I arrived in the laboratory.

2.4.1 Making data interoperable is essential

Consistency of data from different programs…

Villefranche teams collect data on the physics, the hydrology, the biogeochemistry, and the plankton communities of the waters of the bay and around the offshore BOUSSOLE/DYFAMED site, as well as atmospheric deposits near shore, optical properties of the atmosphere and surface waters at BOUSSOLE, and biogeochemical properties at global scale via autonomous Argo floats. Because all these datasets were collected by different teams, they were, naturally, spread within the lab, with very different formats, conventions, processing and backup procedures, even within a single series. However, for researchers to be able to exploit this wealth of data as well as to propagate it consistently to higher level databases, the datasets needed to be interoperable. Furthermore, this data represents considerable effort which should be properly protected and publicised.

…through a common processing pipeline

Therefore, over four years during which I was coordinating these various observational data collection efforts, I set up a central, dedicated data server. Its role is to store all observation data, process it through scripts, visualise it in interactive web applications, and store it in incremental backups (Figure 2.20). The data is still stored in silos per team, to ensure each team has full control on its activities and avoid that mistakes by one affects the data of others, but the processing and visualisations are generic, automated, and shared across teams. This is what I spent most time on.

Use standards for names and data formats

All data, in its rawest possible form, is re-processed every night to check its integrity (the whole process is logged) and make it easy to update something over the full archive. For example, CTD (Conductivity, Temperature, Depth) casts are processed by automated scripts, in separate batches defined by the different configuration of sensors, but targeting a consistent output over the whole archive (1m depth bins, same derived variables, consistent units, etc.). In many othercases, data series are concatenated from multiple source files; then, the names and units of variables as well as the quality codes are all converted to a single internal convention, compatible with international standards such as the Climate Forecast (CF) conventions[a], and obvious outliers are flagged. Once this initial processing is done, each dataset is written to disk, as a whole, in various formats: `RData` for visualisation in R, `feather` for data exchange with python, `csv` for archiving, `sqlite` to enable database queries. Cur-

---

[a] http://cfconventions.org/

rently, this represents a volume of ~7M data records of ~60 variables over 25 source datasets.

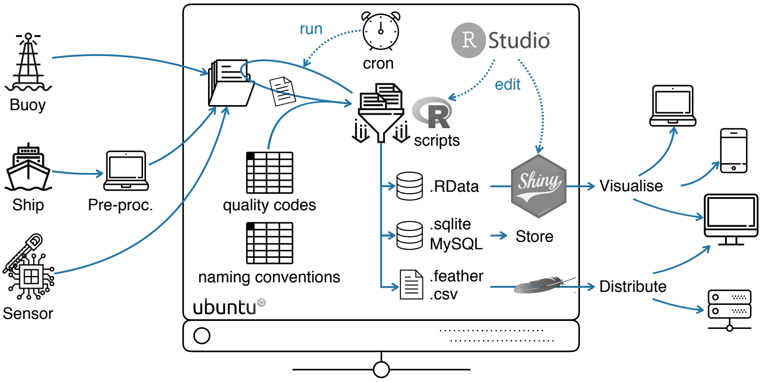


Figure 2.20 Flowchart of the data processing pipeline on Villefranche's data server. Data comes from sensor on the left, is stored on a single file server, processed every night to make it homogeneous and interoperable, written to disk in this cleaned form, and then visualised on http://obs-vlfr.fr/data/view/ and distributed.

Homogeneous data eases visualisation

The second step done locally is to visualise this data through interactive plotting applications (http://obs-vlfr.fr/data/view/). In those web applications, users can click buttons or move sliders and the plot updates to reflect their selection of variables, dates, displayed quality codes, etc. I took particular care to keep the interface of those applications consistent and responsive, to be most useful to their end-users. Because the data is homogeneous at this step, generic interface components to select data and options could be defined and re-used in several applications. Three techniques are used to keep the applications fast: (i) almost all data is loaded in Random Access Memory (RAM) on application startup, (ii) long data series are binned at various precisions (day, week, two weeks, etc.) and the application switches data sources depending on the date span to be plotted (daily data when plotting 2 months, monthly data when plotting 20 years), (iii) reactive programming principles are used to compute the minimal update to the plot, e.g. the query of the dataset for the selected date range is separated from the selection of quality codes so that, if the users only changes the selection of quality codes, the data does not have to be re-queried from the full dataset but only re-filtered to keep the new selection of quality codes. Two kinds of applications are available: some are externally available, to display data and publicise it; some are internal only, for quality control, easy database extraction, etc. Currently 26 applications are available (17 external, 9 internal).

All open source

This data processing and visualisation stack is fully implemented with open-source software, except for CTD processing which has to rely on SeaBird Data Processing (for their proprietary sensors).

2.4.2 The Mediterranean Sea is a climate change hotspot

Among these datasets is the long term monitoring of the hydrological properties of the water at point B, in Villefranche's bay. Temperature and salinity are monitored weekly since 1957 but, since 1995, as part of the national SOMLIT program, a wealth of variables are measured at 6 depths (1, 10, 20, 30, 50, 75 m) and along a full profile using a CTD.

Fill missing values by exploiting correlations

This dataset is, unfortunately, not entirely complete: some depths and some weeks are missing. Those missing values were filled by exploiting the correlation among variables at various depths. For example, temperature and salinity often covary; if salinity is missing, temperature can be used to estimate an approximate value of salinity. In addition, if salinity at the surface is available but not at 20 m, both temperature at 20 m and salinity at the surface can be used to constrain the estimate of the missing salinity value. This was implemented using an iterative Principal Component Analysis procedure[49]:

1. a first PCA is computed by replacing missing values by zeroes (which is neutral in a PCA with variables scaled to zero mean and unit variance);
2. estimates of these missing values are computed by using the first n principal components of the PCA (i.e. taking into account the correlations among all variables at all depths and removing some of the noise because n is smaller than the number of input variables in the PCA);
3. a new PCA is computed with these estimates;
4. steps 2 and 3 are repeated until the estimates for the missing values are stable.

This allowed to reproduce realistic seasonal cycles in variables/depths that are missing over long periods. This imputed data is not to be analysed as such but allows other procedures which do not tolerate missing values to run over the whole series.

Extract trends and seasonality…

With these complete series, long term dynamics were examined at 20 m depth, because it avoids the high variability of surface waters and still captures the seasonal stratification cycle which is characteristic of the region. The series at 20 m were decomposed into a long term process, a seasonal cycle, and residuals using Seasonal Trend decomposition by Loess (STL)[50]. This procedure iteratively fits Loess polynomials over a large window (5 years here) to capture long term trends, a smaller one (3 months here) to capture seasonal patterns, and another one per week and across years to smooth the changes in seasonal cycle across years.

…to highlight earlier summers…

The seasonal component of some variables displayed characteristic points: the peak of chlorophyll a concentration marks the spring bloom, the height and breadth of the temperature peak characterises summer, etc. All these characteristic points shifted earlier in the year along the series, by about 3 weeks between 1995 and 2013 (Figure 2.21), showcasing a notable advance of the summer period, consistent with other evidence around the globe[51].

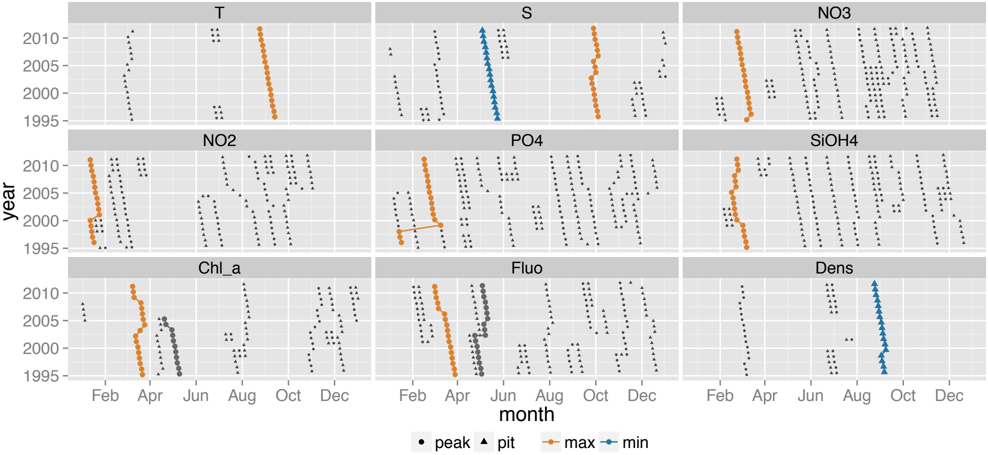


Figure 2.21 Date of characteristic points (peaks and pits) in the seasonal component of several variables; the overall maximum and minimum are highlighted in colour when relevant; the second spring peak of Chlorophyll a concentration is displayed in bold until it disappears in 2005. All points shift towards earlier dates throughout the time series (NB: the shift is real but its smoothness is exaggerated by the STL method).

…and significant warming and oligotrophication

The long term component was examined visually and tested for a monotonous linear increase or decrease (Figure 2.22). Temperature significantly increased, by approximately 0.7ºC in 18 years. In the meantime, nitrate concentration increased while chlorophyll a, particulate organic carbon, and particulate organic nitrogen concentrations decreased significantly, outlining an oligotrophication process. In addition, a ~5 y cycle, robust to changes in the long term smoothing window size, was apparent in most variables, particularly salinity. This cycle is likely related to large scale meteorological forcing similar to the North Atlantic Oscillation (NAO).

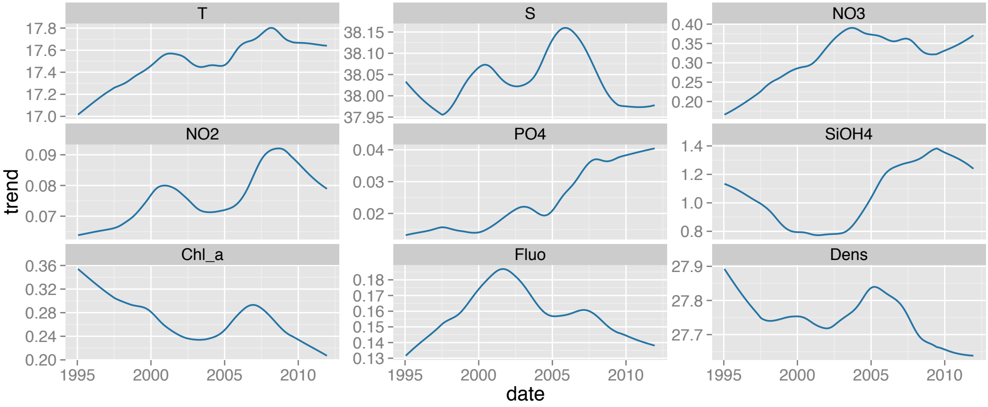


Figure 2.22 Long-term component decomposed for several variables.

Overall, these results confirm the characterisation of the Mediterranean Sea as a climate change hotspot[52], with trends towards warmer and more oligotrophic waters in the North-Western basin.

2.4.3 Planktonic diversity increases along an oligotrophy gradient

Semi-automated imaging of long term collections…

Another flagship monitoring series of Villefranche is that of zooplankton concentration in the bay, where plankton nets are towed at least weekly since 1966 (in recent years, nets were even collected twice a day and pooled into a weekly sample to average short-term variability). Preserved plankton samples are scanned using a specialised flatbed scanner[53] and, therefore, images rather than physical samples, are analysed. The processing of these images is automated: grey levels are normalised, individual objects are extracted and measured, objects are classified taxonomically by a supervised classification algorithm and classifications are all reviewed and corrected by human operators.

…allows to study morphological diversity

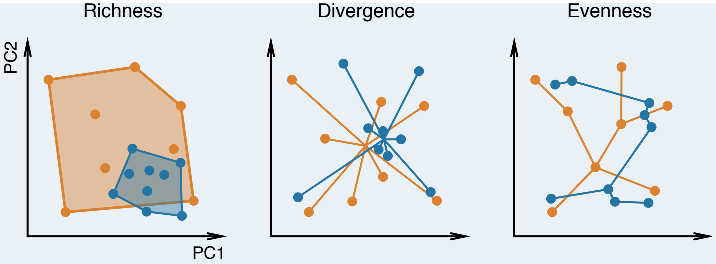
Changes in taxonomic composition can therefore be analysed along the series. While this approach has lead to very informative insights, we wanted to focus on another facet of biological diversity and to take advantage of the individual-level measurements of all organisms (which, so far, have mostly been used to inform the supervised classification). We studied morphological diversity of zooplankton, which can be described with the existing measurements of size, shape, and grey levels (that quantify transparency) for each organism. Indeed, morphology is often related to function in zooplankton (transparency is a strategy to avoid predation in transparent pelagic waters, appendages are often related to feeding, etc.).

An approach inspired from functional ecology to describe morphological diversity

Functional diversity (i.e. the diversity of functions operated by the different organisms of the ecosystem) is often described with the following framework[54]:

1. Organisms belonging to each species are counted in several locations or, in our case, at several dates.
2. Functional “traits” values are associated with each species (because it is most of the time impossible to do this for each organism); functional traits are characteristics such as size, feeding mode, home range, etc. that influence the fitness of an organism as well as how it contributes to the functioning of the ecosystem.
3. The complexity of this multivariate species-traits space is reduced through a dimensionality reduction technique (e.g. PCA).
4. Three indices are computed for each date, which each reflect a different aspect of diversity; their computation is based on the species position in the reduced space of step 3 as well as their abundance at that date estimated in step 1.

In a simplified setting of equal abundance and two dimensional reduced space, the definition of these indices can be summarised as follows.



Each point is a species characterised by several functional attributes, summarised by its coordinates along PC1 and PC2. For each of the three indexes of Villeger[54], two situations (two dates here) are represented; the orange date has a higher value of each index than the blue one.

Richness (left) represents the breadth of functions present at that date, quantified as the area of the convex hull of points in the reduced space.

Within a richness level (similar richness for orange and blue here), divergence (centre) represents whether the species present are all quite different (high divergence, orange) or mostly similar with a few extremes (low divergence, blue); this is quantified based on the distances from the barycentre of points in the reduced space.

Evenness (right) shows wether the change from one species to the next is regular (high evenness, orange) or if there are sudden jumps (low evenness, blue) and is quantified as the length of branches of the minimum spanning tree connecting points in the reduced space

In our case, we have multivariate measurements of morphological traits of *each* organism at each date; therefore, we do not have to reduce ourselves to the species-level approximation classically used functional diversity studies.

The morphological space is constructed by performing a PCA on the values of those ~40 measurements for *all* organisms. Prior to the PCA, the variables are normalised by log/square-root/box-cox transformation and the 0.1% most extreme individuals are removed to better represent the bulk of the community. Then, only a few principal components are selected, based on the Kaiser-Guttman representativity criterion[55].

Computing the diversity indices on all individuals is impractical and the results can be still influenced by only a few individuals, even after removing outliers. Instead, individuals are regrouped into 200 clusters using a k-means procedure and the diversity indices are computed based on the positions of the 200 cluster centroids and the total concentration of plankton in each cluster.

However, among dates, the volume filtered by the net and the fraction of that sample that gets scanned change. To account for these changes in sampling effort, each organism is weighted to represent its concentration in situ; the PCA and k-means are computed taking these weights into account.

Finally, to represent the morphological space of the PCA graphically or to get a synthetic image for each cluster, we designed a way to "morph" several images together. The steps are: (i) find the best fitting ellipse on each organism; (ii) determine the darkest side of the object: the side of the ellipse where the centroid (barycentre weighted by the grey values) is; (iii) rotate the objects to align their major axis horizontally and put their darkest side (usually the head) on the left; (iv) align the centroids of all images; (v) average all images to create the "morphed" version. Because the original images never superpose exactly, the resulting morphed object is often lighter than any of the original organisms. To correct for this, an optional step (vi) is to correct the gamma (i.e. contrast) of the morphed object so that its average grey value matches the mean of the average greys of the original organisms. The appropriate gamma value is found iteratively, using a numerical optimisation procedure.

This approach was applied to the series collected with a WP2 net (200 µm mesh size) from 2009 to 2018 in Villefranche Bay. This amounts to ~590,000 organisms.

Size and transparency dominate morphological variations

The morphological space highlighted that the largest difference among planktonic organisms along the series is size (along PC1), followed by lightness, interpreted as transparency (along PC2; Figure 2.23). This was consistent whether we used the original 43 features measured on the objects or 412 features computed by a Convolutional Neural Network trained to classify plankton images. Features computed by such a CNN are learned by the network itself, starting from a random initialisation; they are completely objective and cannot be biased towards size or grey levels, like our original features possibly are. Therefore, the fact that these two aspects (size and grey) emerged as discriminating seems to be a property of images of planktonic organisms rather than of our way of measuring them.

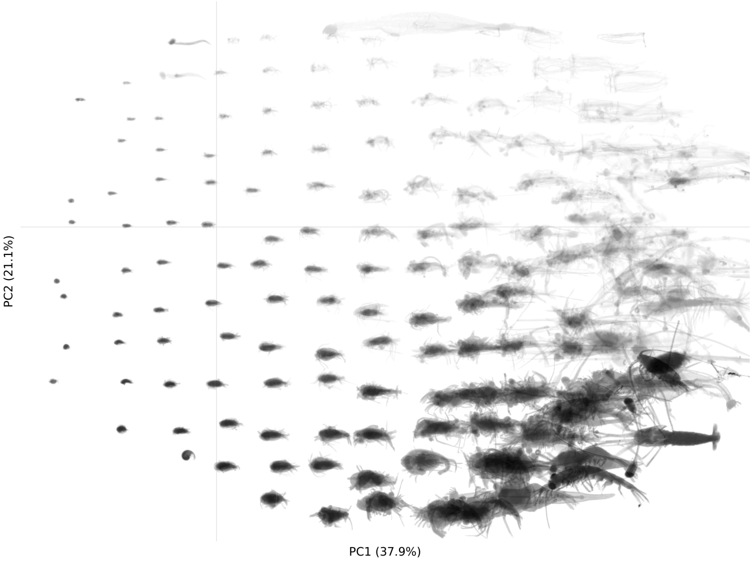


Figure 2.23 First two principal components of the morphological space computed on WP2 samples from 2009 to 2018. The images displayed are "morphed" composites build using the procedure described in the box above, from up to 8 images picked close to the (PC1,PC2) plane at each point.

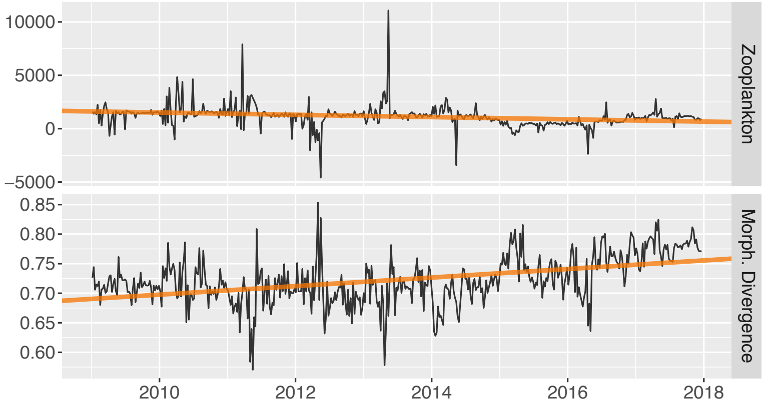


Figure 2.24 Time series of zooplankton concentration and morphological divergence. The black line is the deseasonalised original data; the red line is the result of a Generalised Least Square regression fitted to the data, which is significant in both cases.

Morphs are not (only) taxa

In that space, five principal components were deemed relevant and this reduced space was used to compute the 200 morphs and the three diversity indices. The morphs were morphologically consistent, as expected, but rarely taxonomically consistent: some were dominated by one type of organism (often copepods, due to their shear number) but many were a mix of several taxa that resemble each other, highlighting the fact that this approach captures more than just taxonomic diversity.

The time series of the coordinates of each of the five components and of the three indices were decomposed into a seasonal signal and a long term trend+residuals using the STL procedure described previously.

Morphological diversity decreases during blooms…

The seasonal component showed that all indices of morphological diversity of plankton decrease during spring, the most eutrophic period of the year, and increase during summer, the most oligotrophic one. This seems related to the dominance of copepods during the spring, who have a consistent morphology.

…and increases inter-annually

Over the long term, the general context is that of a warming, more oligotrophic sea, as explained previously. In terms of zooplankton, the overall concentration of organisms decreased but their morphological divergence increased (Figure 2.24). This is does not seem related to any taxonomic group in particular.

Oligotrophication leads to diversification

All these observations are consistent with (i) an oligotrophication process affecting zooplankton too and (ii) the theory that oligotrophic ecosystems should give rise to more diverse communities[56]. Indeed, they cannot be dominated by only a few generalist taxa: organisms have to be specialised to exploit the scarce resources. Examples validating this theory are rare, beyond the fact that the most diverse marine communities occur in the oligotrophic inter-tropical ocean[57]. For zooplankton morphology, this theory seems consistent with both the seasonal patterns, where the most common taxon dominates the most eutrophic period, and longer term trends, where scarcity of resources is associated with decreasing concentrations but increasing diversity.

2.4.4 Spatio-temporal patterns of fish settlement are regionally consistent

I also started a data series in Villefranche, monitoring the settlement of coastal fish larvae. The ecological importance of the settlement phase has already been highlighted in Section 2.1. Along the French Mediterranean coast, where fisheries are local and self-regulated, understanding the processes driving high settlement rates locally would be particularly relevant for forecasting stocks over a few years and guiding their exploitation.

Intense sampling of fish larvae settlement

Settlement was monitored weekly, from mid 2012 to 2016, using light traps[58] set before sunset and collected within a couple hours of sunrise. Concurrently with this regular series in Villefranche, other fish larvae collection programs were ongoing, using the same technique but with less regular sampling, in 12 other sites along the French Mediterranean coast (Figure 2.25). The data of all sites was analysed together, to provide a coast-wide picture. This represents 27,800 fish larvae caught and identified, often to species, over ~7000 samples[59].

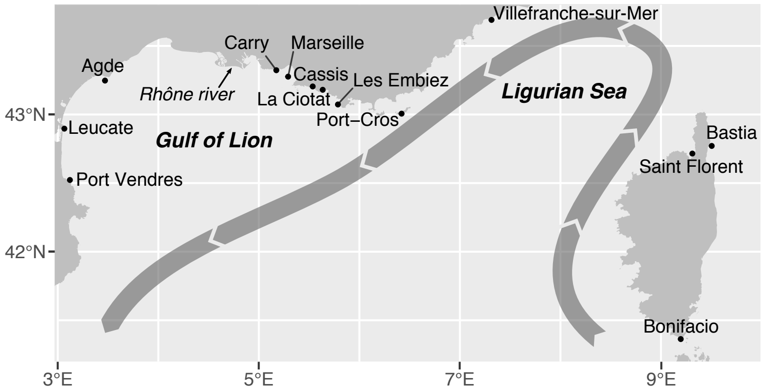


Figure 2.25 Sampling sites along the French Mediterranean coast. The dominant coastal current (the North Current) is pictured with arrows and the mouth of the primary river (the Rhône) is highlighted.

To compensate for the different number of night fished and number of light traps set among sites, catches were analysed in terms of Catch Per Unit of Effort (CPUE) = number of larvae caught per trap per night. Catches were very low overall, with a median CPUE of 0, and were concentrated in a few high catches events in each site.

Those high catches events occurred mostly from May to August and, to a lesser extent, in the fall (Figure 2.26); they were also more likely to occur around the new moon, which is consistent with the findings in the more studied tropical regions[60]. Temporal autocorrelation was weak, and significant over two days at most, showing that those high settlement events are very short pulses. The strong seasonality in the settlement time of each species was consistent basin-wide and over the four years studied.

Settlement is very episodic and seasonal…

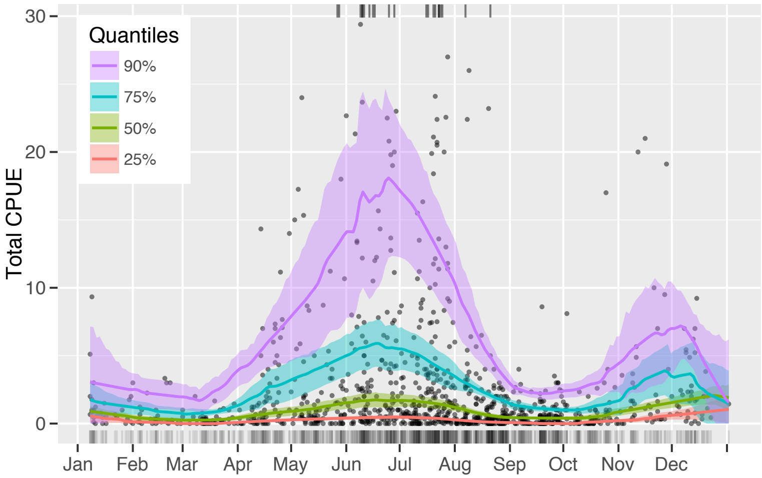


Figure 2.26 Catch of fish larvae per site and per night fished over all sites and all years. Sampling effort is represented as a rug plot at the bottom. Local quantile regressions for quantiles 25%, 50%, 75%, and 90% are displayed in colour, together with their 95% confidence interval. A few particularly high catches are outside of the range of the plot and marked by ticks on the upper side.

Spatially, larval supply was higher at some sites and, on average, higher west of the Rhône river than east of it. Species assemblages were geographically consistent with some site-specific associations. At similar sampling effort, species richness was lower in Carry compared to all other sites and, for higher effort levels, lower in Bastia compared to Villefranche and Leucate (Figure 2.27). This seemed related to the geomorphology of sites, with rocky sites (e.g. Villefranche) displaying higher diversity than sandy ones (e.g. Bastia)

…and consistent over small spatial scales…

Overall, fish settlement in the north-western Mediterranean occurred in short term pulses, that were highly consistent seasonally. While this seasonality was coherent at the scale of the study, assemblages were not and geographical patterns emerged, highlighting the importance of local processes and the need for local management.

…requiring local management

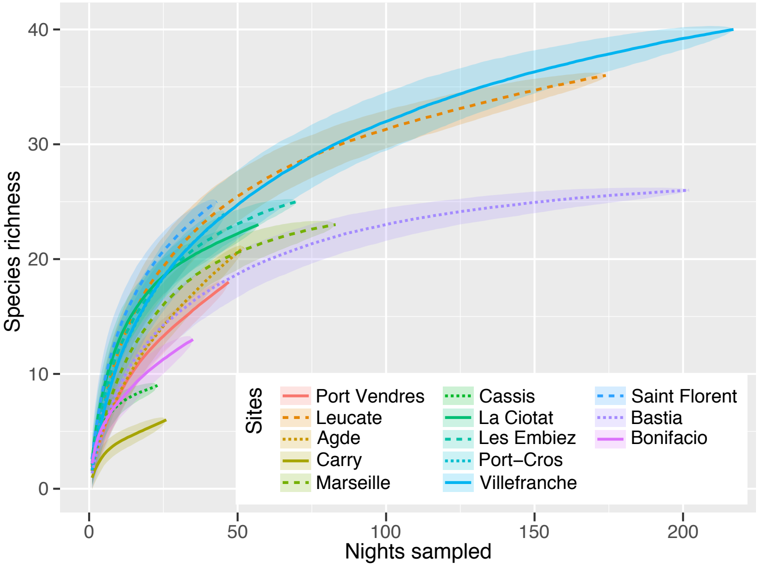


Figure 2.27 Species accumulation curves (number of species caught for a given sampling effort level) at all sites, together with their bootstrapped standard deviation.

Quantile-based methods to study episodic events

The study of episodic events is often hindered by the fact that the mean and median are meaningless and do not capture those rare but important events. In contrast, studying high quantiles (q75, q90) can provide information on such events[61]. This is why all inferential analyses of the study of fish settlement in the Mediterranean are based on quantiles. In particular we implemented:

- locally linear quantile regression with confidence intervals, by fitting a linear quantile regression to a neighbourhood of the point of interest, computing the 95% confidence interval by bootstrapping the points within that neighbourhood, and joining the regression line and confidence intervals at the boundary of each neighbourhood (Figure 2.26).
- quantile-based analysis of variance, by comparing a quantile regression with and without a term of interest and testing significance through an analysis of weighted absolute residuals on 1000 permutations of the data (following Chen et al.). This allowed to compare CPUE values among sites, regions, years, etc. and to highlight, in particular, that catches where higher west of the Rhône than east of it.
- quantile-based autocorrelograms (or quantilograms), which work similarly to autocorrelograms but, while classic autocorrelograms consider the pairwise correlations of anomalies to the mean for each temporal lag, quantilograms consider the pairwise correlations of the *signs* of the anomaly to the selected quantiles (i.e. whether the value in the future is above or below the value

now[63]). This approach is coarser but more robust for series with large episodic events and allowed to detect the short autocorrelation scale of high settlement events.

Overall, this quantile-based approach proved very efficient in handling extraordinary yet ecologically-relevant events, which more classic parametric or non-parametric approaches would have missed. However, because it is less common, it required a significant research and implementation effort, which is made fully available to the community at https://github.com/jiho/medplanet.

Chapter 3

Entracte

3.1 Statistics hotline

Beyond the main projects summarised in Chapter 2, I am often solicited for advice regarding the statistical analysis of data. This usually starts with "Do you have five minutes? I have a small question about my data", continues with a two hours discussion, and ends with me writing code over a few days to clean up the data and analyse it. This basically came with my position (the person I took over from when he retired had this role) and I am very glad for it. Indeed, it allows me to work on diverse problems, each with a somewhat well-defined question, for which I have no other worry than extracting the most out of the data. It is an appreciable break from trying to find ways to frame my personal research as world-saving for grant applications.

Helping others with the analysis of datasets…

These random questions have led me to explore the lipid content of a dominant mesopelagic Antarctic fish[64], the possibility to use cement factory outflow gases to grow microalgae[65], the changes in pH and pteropods concentration in Villefranche's bay[66], the physiological response of corals to various stresses[67], the mode of governance of the fishing fleets of the Atlantic coast from Brittany to the Azores, the delay in diagnostic of osteomyelitis[68], the dive behaviour of elephant seals in Antarctica[69–71], the distribution of Rhizaria vs. Copepoda in the world's oceans[72], the orientation of the mitotic spindle in ascidian embryos[73], the characteristics of ocean-based solutions to address climate change[74], and, currently, the fine-scale distribution of meiofauna in deep-sea sediment, the correlation between genetic and geographic distance in indo-pacific fishes and giant clams, and the size distribution of zooplankton along a latitudinal gradient in Tara samples. All this makes for good brain stretching!

…of very diverse nature!

Among those activities, one was a prolonged collaboration, with the goal of classifying seismic signals (P-waves, etc.) based on their power spectrum. The raw seismic signal was decomposed into six wavelet scales and a library of signals were manually classified as P-wave, T-wave, ship noise, or air gun shots.

Classifying seismic signals…

…with a simple algorithm for in situ instruments…

The first goal was to design a classification technique that was simple enough to be later embedded on autonomous floats; those floats should come back to the surface as soon as a P-wave is detected and transmit their position and time of detection. To do so, we examined the distribution of the power of P-waves at each scale, computed the probability for the power of a candidate signal to be within that distribution, and combined those probabilities by giving more weight to the scales at which P-waves were known to be different from the three other signals (the difference in the distributions of power was quantified by the Kolmorov-Smirnov statistic). This algorithm, which amounts to computing a p-value, proved efficient and was implemented on the floats[75].

We then simulated how a fleet of such floats would change the distribution of the detection of P-waves across the world and increase the definition of Earth's tomography[76].

…with advanced techniques in the lab

Finally, signals in the record of moored hydrophones needed to be classified among P-waves, T-waves, ship-noise, and iceberg-noise. In that case, the data was downloaded to a computer and computational power was therefore not limiting. Seven wavelet scales were extracted and Gradient Boosted Classification Trees were used to classify signals in the four types (among which T-waves were much more numerous than the others). The classifier reached 99% correct classification on an independent data subset and, in particular, yielded good results for the rare but interesting P-waves. This is a good example of how the expertise of a technique I acquired in one domain (niche models) can be of use in a completely different one.

3.2 Open-source code

I code R packages…

…and contribute to open source projects

My activities lead me to writing a lot of computer code. I strive to make it reusable by others by publishing it in the form of open-source collections of scripts or packages, mostly for the R language (through my GitHub account[a], Figure 3.1). In addition, I use a lot of open-source tools and I am immensely thankful for the time people spend coding them. When I can, I try to give back some of my time to the open-source community, by reporting bugs, fixing them if possible, or contributing new functions to existing libraries. Here is a summary of my significant contributions.

3.2.1 Data visualisation

Tools for plotting…

`ggplot2`, a packages for data visualisation) is one of the most popular R packages; it is downloaded ~200,000 times per week. I debugged the early versions and contributed a few functions (for colour scales and mapping coordinates) as well as graphical themes, which got me acknowledged in the book[77].

---

[a] https://github.com/jiho/

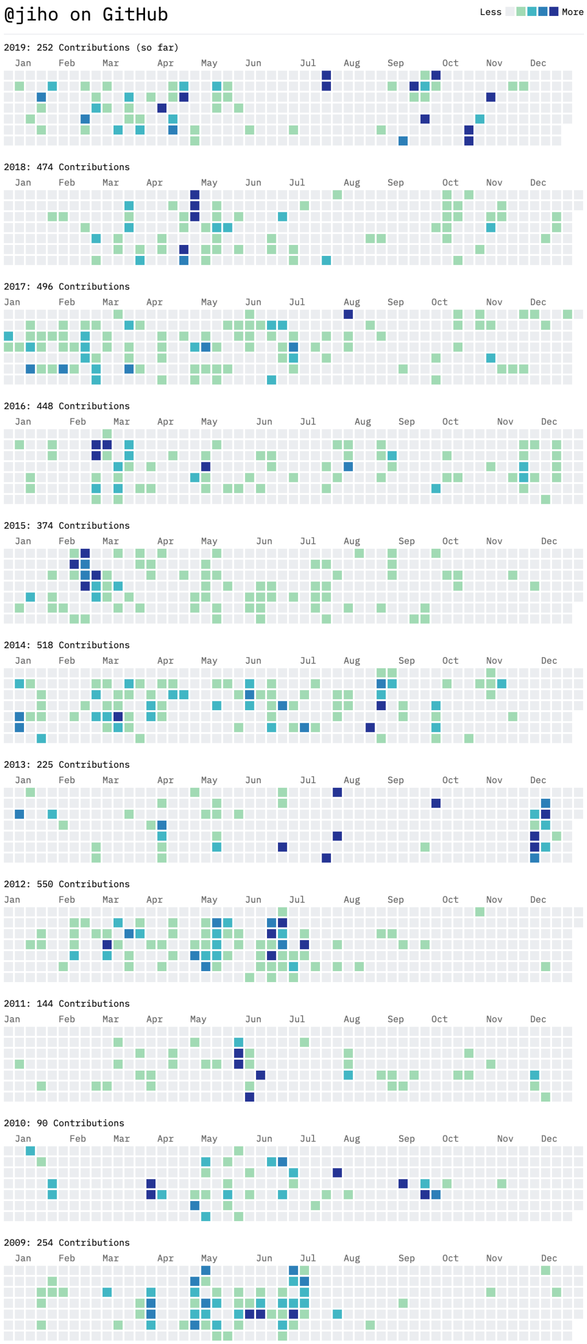


Figure 3.1 Timeline of my GitHub contributions since 2009. I teach mostly in Sep, Oct, Nov, and Jul and it shows as downtimes in coding (except this year when I reviewed one of my packages to use it during a class).

Inkscape is a vector graphics software (similar to Adobe Illustrator) initially written for Linux, for which I contributed the MacOS port.

…and choosing good colour scales

`chroma` is a R package I wrote for dealing with colours. It was born from an experiment I ran during an annual lab meeting, in which I polled lab members about a few graphs, to show them some classic mistakes in data representation, including the problems that rainbow-like colour scales can introduce. `chroma` can parse colours in various specifications, manipulate them, compute colour differences and distances, and, of course, define colour scales and palettes that are perceptually correct.

I also contributed functions to the `marmap` and `ggmap` packages, for plotting maps, as well as to the `ggrepel` package for adding non-overlapping labels to plots.

3.2.2 Data manipulation and extraction

Tools for shuffling data around…

`plyr` was another hugely popular data manipulation package by the author of `ggplot2`, Hadley Wickham. He now wrote a faster replacement (`dplyr`) but I helped Hadley debug the early versions of `plyr`. The `broom` package provides functions that collect objects resulting from various statistical analyses and coerce the results into nicely formatted tables. I contributed functions that deal with objects resulting from multivariate analyses (PCA, CA, etc.) as well as x,y,z lists for 3D plots.

…extracting coastlines…

`shoreliner` is a package I wrote to query and cut the Global Self-consistent, Hierarchical, High-resolution Geography (GSSHG) database of NOAA, that provide coastlines at various resolutions over the whole world. A web interface to cut a piece of coastline to user-specified bounds is available at http://coastline.obs-vlfr.fr.

…and fetching data from NetCDF or image databases

I wrote a few custom packages to ease the type of data manipulation I do often: `ncdf4helpers` allows to cut 2D slices of data within n-dimensional NetCDF files with a natural syntax (giving bounds in terms of latitude, longitude, or depth rather than indexes of dimensions), the package `zooprocessr` reads files and projects created by `ZooProcess` (the software that analyses ZooScan, UVP, and FlowCam plankton images in Villefranche), `ecotaxar` reads files and queries the database of EcoTaxa (the software that replaced ZooProcess+PlanktonIdentifier for plankton image classification in Villefranche), `apeep` processes the enormous datasets produced by the In Situ Ichthyoplankton Imaging System.

3.2.3 Data analysis

Tools to handle circular data…

I contributed several functions to the `circular` package for R, which deals with data that repeats itself along a circle: angles or times. This package is largely used in the code of `discr`, the software I wrote to perform the data acquisition and analysis for the DISC instrument, used to study the orientation of fish larvae (Section 2.1.2).

To analyse biogeographic patterns, I wrote `atlasr`, which performs niche modelling based on boosted regression trees and was used for the

Biogeographic Atlas of the Southern Ocean[78] (Section 2.3.1). I also wrote the R implementation of the Non Probabilistic Potential Ecological Niche (NPPEN) method for species distribution modelling[79], which proved hundreds of times faster than the original MATLAB implementation thanks to a numerical trick (package `nppen`). To analyse the distribution statistics of organisms responsible for harmful blooms as part of the EU project CoCliME, I wrote the package `coclimer`.

…to perform biogeography analyses…

To ease the use of the methods I developed for sound waves classification in the context of seismic tomography (Section3.1) by people unfamiliar with R, I wrote the `soundclass` package. Similarly, through of my efforts to handle observation data in Villefranche (Section 2.4.1), I wrote the `castr` package to facilitate the processing of data series such as CTD casts or monitoring time series.

…to process data series…

Finally, exploring the morphological diversity of plankton (in Section 2.4.3) required quite a bit of coding, in Fortran to perform weighted k-means in a fast and memory efficient way even with millions of records and in Python for image processing (so as to produce Figure 2.23 for example), all wrapped in R packages `wkmeans` and `morphr`.

…and to explore morphology from images

3.2.4 Miscellany

I am in charge of the website of the Master's degree in Oceanography (Sciences de la Mer): http://sdm.sorbonne-universite.fr. It has the peculiarity of being hosted on a server in Villefranche but relies on lists of students, classes, internships, etc. handled by the secretariat of the Master in Paris, mostly in the form of spreadsheet documents (i.e. not as a queryable database). To accommodate this situation, I wrote the package `yssr`: a static site generator that generates the HTML files for the website based on a hierarchy of folders and a set of spreadsheet documents that the secretary can easily understand and modify and that is shared between Villefranche and Paris. Unorthodox, but it works.

And a bit of web development

3.3 Teaching activities

This memoir is about research, but half of my time is spent teaching, so I cannot omit it. In addition, I think that I am a better professor than researcher; at least, teaching comes to me more naturally so I enjoy doing it. Actually, I could not envision my research career without some teaching (maybe not as much as the ~200h I am currently doing, but some!).

First because the qualities honed over time as a professor (how to explain difficult concepts to people unfamiliar with them, how to organise thoughts logically, how to start from a simple example and generalise, etc.) are all useful for typical research endeavours such as preparing a conference presentation or drafting a paper.

Teaching shapes the thinking process…

Second, teaching provides an immediate feedback, based on the attention of students, that research does not. In research, we know whether an experiment/cruise was worth it only months afterwards, when the data is analysed; we know whether others appreciate it maybe a few years later, once it has been presented at conferences or published in a paper that gets cited. One has to be extremely confident in the research undertaken to stay motivated every day over such long periods of time; I find that teaching provides a good shorter-term balance to research's long-term objectives.

…and forces to deeply understand the topics taught

Finally, because I am lucky enough to teach mostly at Masters degree level, my teaching is often closely related to my research: I teach Lagrangian modelling, fish larvae ecology, general behavioural ecology, scientific cruise planning, and a lot of statistics (multivariate, spatial, time series, etc.) together with some introductions to programming. I often notice that, even though I thought I had a particular numerical method figured out, my understanding only because complete when I had to teach it and explore every corner of it to anticipate or answer questions from students. In addition, teaching about my research activities allows me to get in touch with prospective interns and helps me to consider what they know (and do not know!) to propose solid, yet realistic, internship topics. This has been instrumental for many of my research projects, which have benefited from the work of very skilled interns.

Figure 3.2 Teaching *is* enjoyable, even in moments like these.

Everyone should teach

Overall, I find teaching essential, even just from the point of view of research, and I think that scientific research in France would improve if everyone was doing both. My CNRS colleagues, who do not have to teach, often argue that they would be bad at it, or are not interested by it. I argue back that the qualities that make a good researcher are so intertwined with those that make a good professor that I do not think it is possible to be good at research and very bad at teaching. Also, because of the immediate feedback, it is faster to become better at teaching than better at research. And finally, it is impossible to know wether something is truly uninteresting until after having tried it for some time; those who try teaching a bit usually find that they are less bad at it than they thought they were and that it is more rewarding than they thought it was. So overall, I stand by my point: every researcher should teach!

3.4 Community and management activities

Beyond my research and teaching activities, I contribute a fair bit of time to the overall functioning of the scientific research community. Here are my main activities.

Evaluating cruise proposals is interesting…

From 2014 to 2018, I was a panel member for the Commission Nationale de la Flotte Côtière, which evaluates the applications for ship time on the six "coastal" oceanographic ships of the French scientific fleet. This meant evaluating 4 to 6 applications for each of the bi-yearly panel meetings and participating to the meetings. This activity was intellectually stimulating, gave a good overview of the various types of cruises planned on oceanographic ships but proved frustrating on occasions. First, because the evaluation process was sometimes not fully taken into consideration by the proposers (applications did not improve over time) and/or by the instances running and programming the activities of the ships (because some ship routes were pre-planned and other excluded, irrespective of the scientific evaluation). Second, because of the relative decoupling of the evaluation of the whole scientific project and that of the cruise; this sometimes led to granting ship time to projects that eventually did not get funded and for which the possibilities of exploiting the cruise were therefore impaired. I think everything should be factored in the same evaluation process, of the full scientific project including the cruise, and experts from this commission should be solicited to do so.

…but can be frustrating

Coordinating ocean observations at Villefranche was challenging…

As already mentioned, from 2013 to 2017, I coordinated the ocean observation activities of the Observatoire Océanologique de Villefranche-sur-Mer (OOV). This mostly meant meeting with all the parties involved and trying to make existing observation programs work smoothly; this was complemented by some budget management and human resources advice but over a limited scope. The main innovations I introduced were (i) the centralised data server that allows the automatic processing of data, its visualisation as well as its backup (described in Section 2.4); (ii) increased transparency, internally at OOV, regarding how these activities work, with a fully public consolidated budget and an estimation of the time involvement of every technician, engineer, and researcher or professor in the ocean observation activities (in terms of days per year spent on each task). This proved instrumental in justifying how significant those activities were. Both of these innovations are still in use currently.

…but has lasting effect

Contributing to the Société Française d'Écologie et d'Évolution…

Since 2014, I am a member of the board of the Société Française d'Écologie et d'Évolution (SFE^2). The goal of this scientific society is to promote the sciences of Ecology and Evolution within academia and towards a larger public. Every year, the society distributes three awards (for young researchers based on a PhD paper, for mid-career researchers, and for lifetime achievements of advanced researchers) and funds two grants (for field work within a PhD and for applied research actions joining a academic team and a non-academic entity to achieve an ecological conservation goal). Every two years, it contributes to the organisation of the SFEcologie conference, once on its own (~900 attendees), the next time together with another European ecological society (the British BES, the german-speaking GFÖ, etc.; ~1500 attendees). Throughout the year, it co-organises, sponsors, or at least distributes information regarding events (debates, conferences, etc.) around the themes of ecology and evolution in France, through it website, Twitter, and FaceBook accounts. It also runs the SFEcodiff service, which allows anyone to post job/internship/PhD offers in the fields of interest of the society; those posts are manually moderated and then distributed on the website and to a mailing list of >9000 subscribers. As a board member, I participate to the bimonthly meetings that allow to run these various activities (discuss the awards, evaluate and discuss the proposals, coordinate the events the SFE^2 participates to, etc.). Day to day, I am more particularly in charge of the website and the SFEcodiff service, which I partly run, partly subcontract to a company. In addition, I motivated the board to publish opinion pieces[b] that are of significance for academia or for the general public. I contributed to two of them: one on the absurdity of the current scientific publishing system and one on the urgency to act in the face of climate change and biodiversity loss (that will be forwarded to the press and government in the near future). Beyond the fact that I am happy to contribute to the actions of SFE^2 because I find them important, it is intellectually stimulating for me to be embedded in a general "Ecology" society, as opposed to an "Oceanography" one; it keeps me in touch with a broad field that, I feel, encompasses marine ecology/biological oceanography.

…broadens my scope

Leading a team…

Finally, since 2016 I lead one of the research teams of the Laboratoire d'Océanographie de Villefranche (LOV). From 2016 to 2018 it was the Processus dans les Écosystèmes PélagiqueS (PEPS) team (10 researchers/professors, 3 emeritus, 4 engineers, 2 invited researchers). Since the restructuring of LOV in 2019, and for the next five years, it is the COMPutational PLankton Ecology (COMPLEx) team (6 researchers/professors, 2 emeritus, 2 engineers, 1 invited researcher). Both teams focussed on pelagic ecosystems. PEPS had two main groups (zooplankton ecology in situ and phytoplankton ecology, mostly studied in the lab) and a few extra activities. COMPLEx is re-centred on zooplankton ecology in an effort to refocus all teams within LOV. Leading the team means, at minimum, handling the budget allocated to it by the lab (~20k€ for PEPS, ~10k€

---

[b] https://www.sfecologie.org/actions/positions/

for COMPLEx) and participating to monthly, 1/2 day management meetings with the laboratory director and the other team leaders. In addition, for PEPS, we tried to organise scientific+management monthly meetings, but it proved difficult to keep everyone interested in the scientific part given the wide range of topics covered within the team. For COMPLEx, we decided on shorter, weekly meetings for management and organise a 1/2 day scientific meeting every month and a half, focused on the activities of the current post-docs, PhD students, and interns, so that they can all benefit from the advice of the whole team rather than only their advisor. In addition to these scheduled activities, team leading activities included writing the summary of PEPS activities and prospective of COMPLEx activities for the evaluation of the lab by the HCERES (which occurs every five years), drafting plans for allocating lab and office space at the level of the three building of OOV (which is very time-consuming yet of prime importance on the very space-constrained site of Villefranche), and dealing with day-to-day questions that may arise within the team (e.g. team-level strategy for hiring of new staff). While clearly the most time-consuming of my community activities, I find leading the COMPLEx team very enjoyable, chiefly because we are a tight-knit group who get along well with each other, yet have strong opinions and frequent thought-provoking debates. This leads to a true "team spirit". Our way of functioning requires significant coordination effort but I feel it pays off in terms of productivity and solidity of the work accomplished, compared to the typical organisation I knew before, with single researchers working with their post-docs and students. I hope I can help keep this team spirit alive for the next four years of operation of COMPLEx.

…is time consuming…

…but extremely rewarding

Chapter 4

Present activities and research project

4.1 Why machine learning for plankton ecology?

Digital images are data for scientific research…

Digital images have become an integral part of our lives: they dominate our news feeds and capture our most personal moments. In science too, digital imaging has been a game changer. For example, medical imaging is often indispensable from diagnostic to treatment, space is largely explored through images from telescopes and, in the other direction, satellites provide us with increasingly detailed images of the earth. Ecological sciences have also benefited from digital imaging: difficult to access animals are recorded with camera traps, field surveys are carried out through photo or video recordings, etc. This is particularly true in the aquatic realm, where direct observation by humans is more difficult than on land. On such ecological images, the first aim is often to identify which species/taxa, are present and, sometimes, to extract individual level traits, such as size or colour.

…in particular in aquatic ecology

Quantitative imaging provides high data volumes…

An important provider of ecological information in the form of images are *quantitative imaging* instruments: high frequency image acquisition followed by the systematic extraction of information yields large number of data points that can then be used to estimate density of organisms, diversity indices, etc. Several instruments have been developed specifically for *plankton* imaging (Figure 4.1). Planktonic organisms are often small (µm to mm), transparent, fragile, and very unevenly distributed. Sampling them directly is difficult and incomplete (because it destroys the gelatinous organisms for example). Sorting the resulting samples is long. It difficulty scales to the high resolution needed to investigate plankton patchiness in hydrological structures and the large number of stations required to cover their pan-oceanic distribution. Imaging devices deployed in situ with specific lighting techniques (such as shadowgraphs) can provide a more complete and higher resolution description of planktonic communities, albeit at a lower taxonomic resol-

…and is particularly effective for plankton ution. Still, because most organisms are transparent, imaging can also reveal individual level *traits*, such as gut fullness or reproductive status, that provide additional insight into ecological processes.

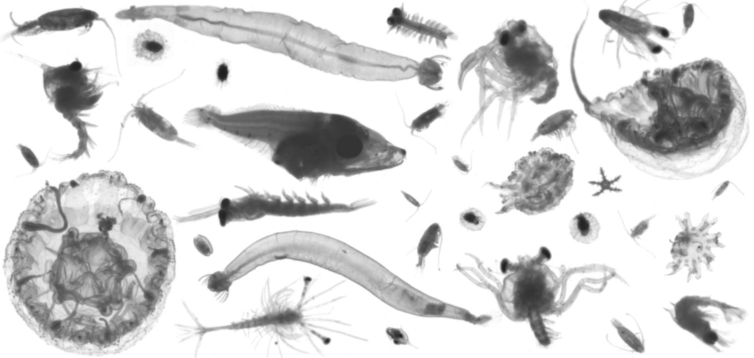


Figure 4.1 Montage of plankton images from quantitative imaging instruments, on their usual uniform white background

Plankton is a key component of the biosphere…

Obtaining such detailed information on plankton is particularly important because is it a keystone components of Earth's biosphere. Plankton is responsible for half of the global primary production[80]. Dead plankton and excreted matter sink to the seabed and store massive amounts of carbon, hence regulating climate through this biological carbon pump[81]. Plankton is also a critical component of many marine food webs[82]; it directly supports some of the largest fisheries on Earth[83] and some emblematic species such as corals. Finally, many taxa have strict environmental requirements while they live a relatively short life embedded within a given water mass; this makes them very sensitive indicators of change in those water masses[84], useful for ecosystem monitoring programs[85]. Despite its importance, recent studies have still highlighted large gaps in our knowledge of plankton. Extensive genomic sampling in the Tara Oceans expedition uncovered around 40 million genes, the vast majority of which are new to science[86]. This data also suggests that the contribution of some planktonic groups to the biological carbon pump was grossly underestimated[87]. In situ imaging revealed that rhizarians (large and fragile unicellular organisms) dominate over crustacean plankton in wide oceanic regions[72], contrary to what was previously assumed.

…for which our knowledge is still largely incomplete

Machine learning can accelerate the acquisition of data…

Those knowledge gaps are, in no small part, due to the lack of data or the incomplete exploitation of the very noisy data we have. On one hand, sampling plankton is difficult and, on the other hand, imaging instruments yield an overflow of raw data that needs to be interpreted to be usable. *Machine learning* techniques can help solve this problem. To classify organisms into taxa or ecological groups, morphological features are measured on each image, a set of images are identified by humans, then supervised learning algorithms use the features of those known examples as templates to classify the other, unknown, images. In this field, Convolutional Neural Networks (CNNs) have scored higher than any other technique on major image classification challenges[88,89]. In these *deep learning*[90] approaches, images pass through a hierarchical network of unitary transformations, which computes morphological features, and those features are then fed to a classifier (Figure 4.2). The initially random parameters of the network are progressively optimised to simultaneously learn the "best" features and the "best" classifier based on these features. With large enough sets of training examples, CNNs often reach high accuracy. But classic or deep learning techniques also apply to regression problems of course, where one needs to predict a numerical quantity. For plankton imaging, applications could range from the automated extraction of a specific trait from an image (quantity of reserves, size of gonads, etc.) to the detection of relationship between environmental conditions and the concentration of the organisms. In this later case, "computational" approaches such as Random Forests or Gradient Boosted Trees have proved more successful than traditional statistical models[91] (see Section 2.3.1).

…and its interpretation

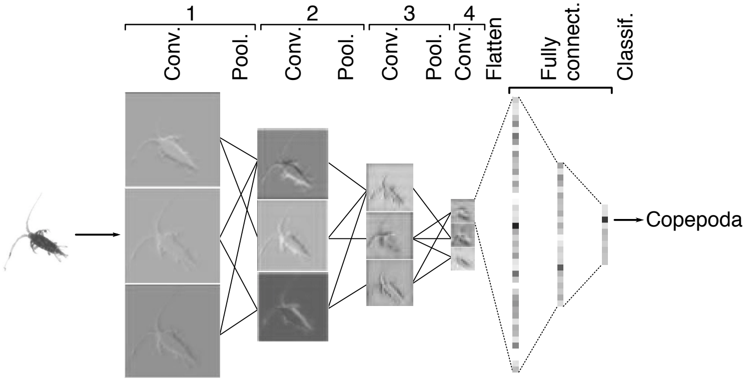


Figure 4.2 Schematic representation of a CNN. The image is transformed through a set of filters, three here, that highlight some of its morphological features (convolution step). Those features are summarised within a window moving over the image, hence reducing its size (pooling step). Convolution and pooling steps are repeated, four times here, on the output of the previous step. Finally, the content is flattened into a vector, the dimension of which is reduced through fully connected layers. The last fully connected layer is mapped to the training categories, here providing a classification. Each layer is linked to the next ones through weights, that are optimised backwards during training.

My research aims at developing such computational approaches to (i) accelerate the identification of planktonic organisms from images and increase the data throughput, (ii) investigate the fine distribution of planktonic organisms inside mesoscale (10-100km) oceanic structures to detect dynamical processes, (iii) extrapolate the effect of such mesoscale structures over larger scale through satellite imagery.

4.2 Identification of plankton from images

The taxonomic classification of plankton from images using machine learning is not a new problem: the first attempt dates back to 1978[92]. Since then, over 200 papers have touched on the topic, with five to ten papers per year from 2000 to 2014 but over twenty per year since 2015. The shift is probably associated with a Kaggle competition on plankton images classification[a] that spiked some interest from researchers beyond the ocean sciences community. It also pioneered the use of CNNs on plankton images.

A well-studied machine learning problem…

Despite this considerable effort, most papers present one-shot experiments on disparate datasets, with no further applications and no means to explicitly compare performance among methods. The only conclusion is that accuracy reaches an upper limit at about 85% on realistic datasets and is frequently around 75%. These numbers are too low to be trustworthy for many ecological applications and imply manual labour to correct labels predicted by classifiers. To this end, we developed the EcoTaxa web application[b] that allows plankton taxonomists to efficiently sort images (Figure 4.3). Its machine learning backend predicts a class and an associated score for each image. Sorting images by score allows taxonomists to focus their effort on the most difficult cases and to reach about 3000 images reviewed per person and per hour (and up to 15,000 for the easiest tasks). Doing it within a centralised web application leverages the effort of the community and allows to learn from a vast amount of classified data. These features enabled the growth of EcoTaxa to its current size of 92M images (40M of them labelled), 700 users from 200 institutions (40 of them active concurrently during weekdays). It is certainly the largest and most diverse database of its kind.

…that has not be solved

We built on the results of the Kaggle competition to define the CNN that contributes to EcoTaxa's image classification system. It is a spatially-sparse network[93] (which discards the uniform background of plankton images) with 12 layers alternating convolutions and fractional max-pooling[94] (which shrinks images slower than classical pooling and therefore extracts more information from the typically small plankton images). The features defined by this network (i.e. the output of the last fully connected layer) are quite generic descriptors of plankton images; they are used, together with handcrafted features imported by the users (such as measures of size, elongation, grey level, etc.), to train a Random Forest classifier and predict the most likely class and score. The same CNN was also used to classify a real-world dataset of 20M images of plankton. It largely improved accuracy compared to the previous approach, based on handcrafted features and Support Vector Machines, and constitutes the first large-scale application of CNNs to plankton images[95].

The challenge is now to improve the *efficiency* of the *whole* imaging pipeline. While the CNN architecture currently used in EcoTaxa seems

---

[a] https://www.kaggle.com/c/datasciencebowl/

[b] https://ecotaxa.obs-vlfr.fr

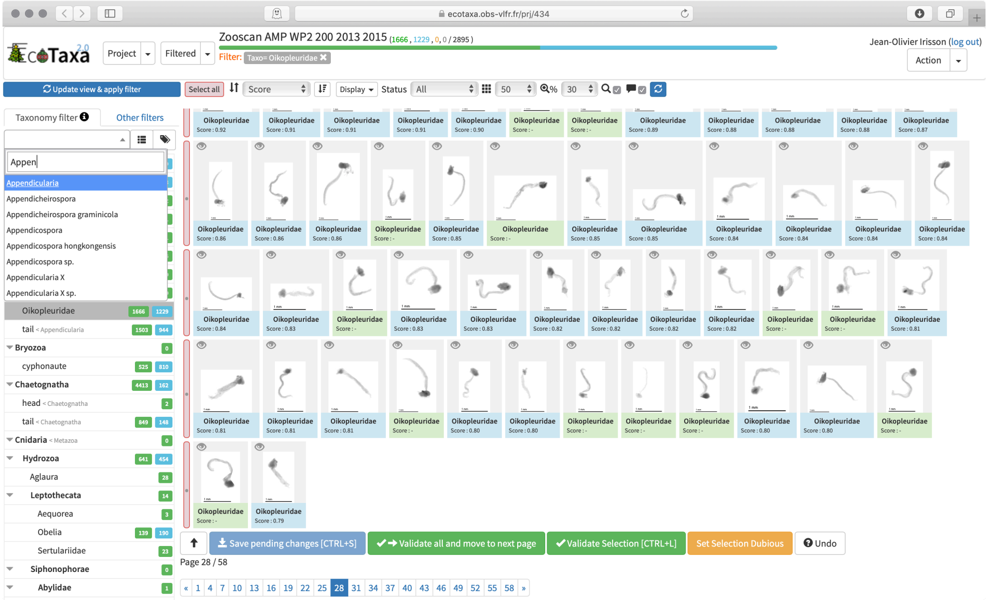


Figure 4.3 Classification page of the EcoTaxa web application. Images are displayed on the right, they can be ordered/filtered/described according to various criteria from the top of the page, and sorted in categories on the left, either by drag and drop or by typing their name (with autocompletion, as showcased). Images reviewed by a human operator are green. Based on those, the machine learning algorithm proposes labels for other images (in blue), that can be validated in batch, when they are all correct, with the buttons at the bottom of the page.

T1.1 – Evaluate current CNN architecture…

adapted to the peculiarities of plankton images, more classic but more modern architectures, such as Inception[96], result in very high accuracy[97]. Therefore, we will compare the current approach with a few modern alternatives and pick the best one. Published datasets extracted from EcoTaxa[98] now allow such objective comparisons. Also, because the score associated with the classification is so important for EcoTaxa operators, classifiers will be trained and ranked according to a metric such as the log-loss (which considers the difference between the predicted probability of being in a class and the actual class) rather than by maximising the overall accuracy, as is usually done.

…and training settings

T1.2 – Improve the database

Classifiers with better accuracy should obviously increase the data throughput. However, most of the slowdowns are currently related to the interactions of users with the data, to validate the identifications proposed by classifiers. To handle the millions to billions or records that plankton image databases host in a swift way, we will test GPU-powered databases, which take advantage of the massively parallel abilities of Graphics Processor Units (GPU; i.e. graphics cards) to query tables, hence reducing request times by orders of magnitude[99] and changing our interactions with large tables. We will explore various choices, including Kinetica[c] and BlazingDB[d], as possible solutions in this space.

T1.3 – API for computer-to-computer interactions

Another purely technical development that will accelerate the whole data processing is the creation of an Application Programming Interface (API) on top of EcoTaxa's database. This API will allow other computer programs to interact with the data, rather than having to click buttons on the current user interface. This will allow to write scripts that continuously stream new images to the database, removing a first slowdown in long-term monitoring projects. Then, images and data could also be accessed from other web applications, in particular citizen science initiatives such as PlanktonID[e], hence increasing the number of users contributing to image classification. Finally, the API will be used to automatically export data to the Ocean Biogeographic Information System[f], managed by UNESCO, which is the largest source of biogeographic information for the marine realm.

T1.4 – Group similar objects through unsupervised clustering

As shown in Figure 4.2, CNNs compute numerical features that summarise the visual appearance of the objects passing through the network (i.e. the content of the last fully connected layer). Organisms that look similar have similar feature vectors and, therefore, those vectors can be used to cluster objects in an unsupervised manner. Using density-based clustering (e.g. HDBSCAN) or neighbourhood-based dimensionality reduction (e.g. t-SNE), dense clusters of extremely similar objects can be identified[97]. Such batches of similar images could be presented to the operator for validation instead of individual images, hence speeding up the validation process. A recent experiment in that direction[g] suggests a 2× improvement in throughput. A similar approach will be tested and implemented in EcoTaxa.

T1.5 – Use phylogeny as a guide to improve classification

Classifiers typically consider all classes as equivalent. However, biological taxa can be organised into a hierarchy, the most classical one being their phylogeny or Linean taxonomy. Even though evolutionary convergence may cause organisms from two very distant branches to look similar, the general case it that phylogenetic differences manifest into morphological ones: distant taxa look different, closely related species look similar. Therefore, the phylogeny should be usable as a guide for the classification of organisms from images. While CNNs can incorporate such hierarchical constraints[100], classification trees (such as those in a Random Forest), which are inherently hierarchical, seem like a more natural fit. We will modify the measure of node impurity to incorporate a penalty proportional to the phylogenetic distance and force high level taxonomic splits (e.g. animals vs. algae) to be resolved early in the tree. Including such external knowledge should improve the overall accuracy. In addition, it should reduce mistakes with large ecological con-

---

[c] https://www.kinetica.com
[d] https://blazingdb.com
[e] https://planktonid.geomar.de/en
[f] https://obis.org
[g] https://morphocluster.geomar.de

sequences (e.g. mistaking a jelly for a crustacean) compared to those that could be less problematic (e.g. mistaking various jellies).

T 1.6 Semantic segmentation of plankton images

“Plankton” images are actually often images of various sort of detrital matter found in the water: fecal pellets, dead carcasses, gelatinous mucus, etc. Those represent ~10M of the ~40M classified images currently in EcoTaxa; the largest biological group is Copepoda with ~5M images. In images from in situ devices, the proportion of those “marine snow” particles is even around 85%. The fall of marine snow to the deep sea is the main contribution of surface plankton to global carbon sequestration, so it is of prime ecological importance; yet it can be efficiently quantified by various instruments while, on data from imaging instruments destined to study planktonic diversity, it constitutes noise. Sorting through this noise wastes important human effort. To reduce this noise, we will explore semantic segmentation approaches, which extract only objects of known types from full images, as opposed to the current approach that extracts all dark objects and then classifies them. Numerous deep networks have been proposed for this task; U-Net[101], SegNet[102] (Figure 4.4), and Mask R-CNN[103] are well-known ones.

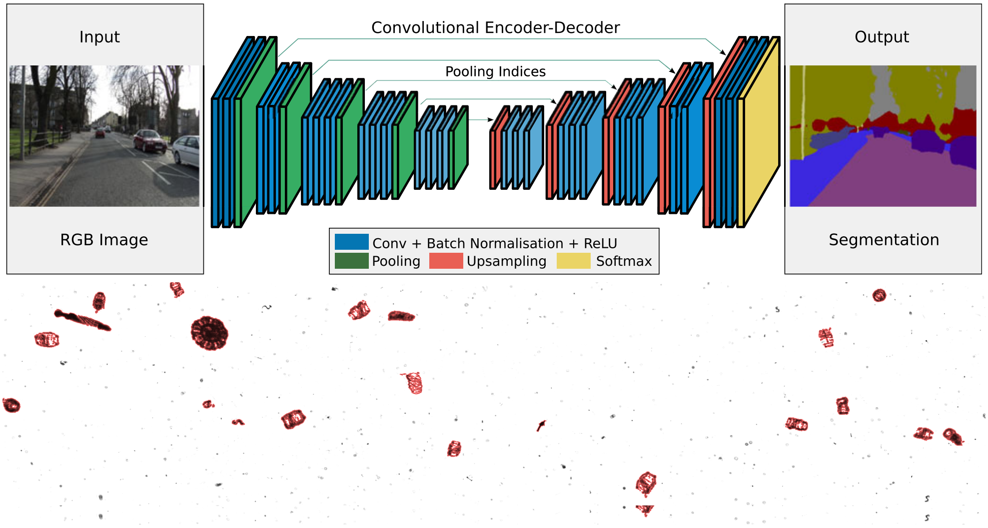


Figure 4.4 Top: The SegNet architecture first simplifies the information in the input image through the usual combination of convolution+pooling layers before upscaling it again to an image of the same dimension as the original where each pixel is classified into an object type by the final layer. Bottom: a piece of frame of an in situ plankton imaging system. The background takes most of the image and the few objects of interest are masked in red; those masks (each with a class associated) would be the ones used to train the segmentation network.

4.3 Submesocale understanding of planktonic communities

Life in the ocean is governed at submesoscale…

As highlighted in Section 2.3, the average distribution of plankton populations is largely shaped by the global latitudinal gradients in light and temperature[104]. Still, local concentrations may vary several folds. These intense patches are associated with short-lived (days) small scale (0.1-10km) submesocale structures, such as filaments and small vortices, that are created by instabilities in mesoscale (10-100km) fronts[105]. The small spatio-temporal scale of these structures makes them hard to describe while they probably have a governing influence on life in the ocean.

…for phytoplankton…

Indeed, they can enhance phytoplankton growth by upwelling nutrients into the depleted surface layer and by increasing vertical stratification, hence residence time in the well-lit, surface, euphotic zone[106]. But, they can also decrease phytoplankton stocks by downwelling cells below the euphotic zone. Furthermore, they interact with mesoscale stirring that creates filaments and thin vertical layers, of highly concentrated organisms. The balance between all these processes depends on the relative depths of the mixed layer, the pycnocline, the euphotic zone, and the nutriclines; all of which vary in space and time. Because of the complexity of these interactions and of their transient nature and small scales, sampling or modelling the effect of sub-mesocale structures is particularly challenging. Still, all existing evidence suggests that they may have a governing role on the distribution of phytoplankton and, particularly, on its diversity[105].

…and zooplankton alike

The reaction of *zooplankton* to these features is even harder to quantify because automated sensors are scarce and zooplankton can move significantly, at least vertically, which may partially decorrelate concentrations from local conditions. Yet, rare cases of intensive sampling show that zooplankton concentrations and community structure change concurrently with small features around mesoscale fronts, both horizontally at km scale[107–109] and vertically at m scale[110]; differences in concentration can reach an order of magnitude. The effect of sub-mesoscale features is therefore visible on zooplankton too. It is actually strong enough to propagate all the way up the food chain to top predators, such as seabirds and tunas[111,112].

High frequency imaging can capture such scales

The understanding of sub/mesoscale structures, and of their influence on phytoplankton, came from our ability to describe them at high enough resolution, via satellites first and autonomous instruments more recently[105]. As Section 2.2.1, we need to describe zooplankton distribution at the same, high, resolution to hope capture the effect of local environmental forcing on its concentration. Imaging has been recognised as a reference method of quantitative observations of plankton[113]. Indeed, only high frequency in situ imaging devices, such as the ISIIS[114], the ZooGlider[115], or the UVP[116], are capable of resolving the distribution of of mm-sized pyto- and zooplankton over scales of cm vertically and km horizontally. They can help us gain insight into these important submesoscale structures.

We used ISIIS to repeatedly sample a mesoscale front created by a coastal jet in the Ligurian Sea (NW Mediterranean) during the VISUFRONT cruise. We sampled cross-front transects with 1km horizontal and 10cm vertical resolution, on a fixed line and while following Lagrangian drifters. Only a fraction of that data has been analyses yet and allowed to describe the vertical patterns in Figure 2.15. Figure 4.5 exemplifies the finescale relationships that this data also hints at: the distribution of radiolarians closely tracks the deep chlorophyll maximum's depth and spread and concentration is lower in the frontal region (30-40km). Even finer scale features such as plankton accumulation on the edges of vertical recirculation vortices of deep, oxygenated waters, seemed to appear for other groups.

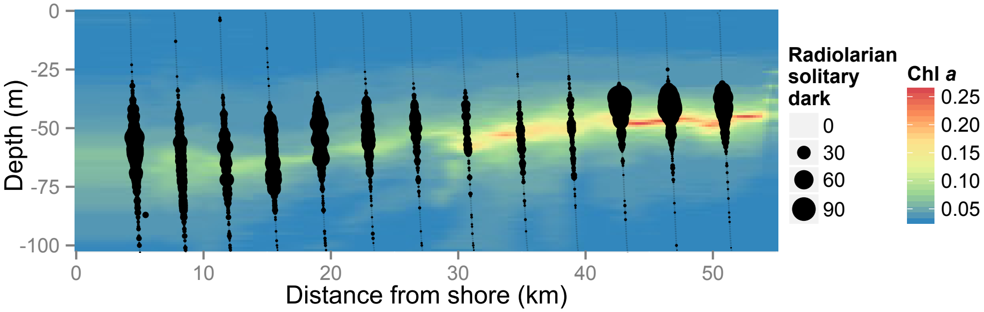


Figure 4.5 Distribution of a specific kind of Radiolarians (black dots) displayed over an interpolated map of Chlorophyll a concentration. Only 1/4th of the ISIIS data has been analysed to detect Radiolarians.

In addition, our recently-developed UVP6 miniaturised camera system will be installed on a glider (autonomous robot that moves in the water by adjusting its buoyancy; Figure 4.6) and will regularly sample a transect close to the VISUFRONT one. This data, collected during the spring bloom and summer oligotrophic period, will allow to expand the conclusions inferred from the summer-only ISIIS data and detect other processes during the dynamical spring period.

T2.1 – Produce high resolution data through automatic classification

An initial processing of the VISUFRONT data produced 20 million automatically-labelled images, which cannot all be verified by taxonomists. Yet, because the data is so rich, we proved that discarding images classified with low scores provides a smaller but more trustworthy dataset, that can still reveal ecological patterns despite a lower recall[117]. In addition, efficient selective validation of intermediate score images can help grow this dataset. We plan on using the improved classifiers developed above (Section 4.2) and this selective validation to produce an accurate picture of the distribution of about 50 planktonic taxa along the front, from the data of both sensors.

T2.2 – Describe morphotypes in addition to taxa

In addition to providing a substrate for classification, morphological features extracted from images are data in their own right. As shown in Section 2.4.3, they can be used to objectively define planktonic morpho-

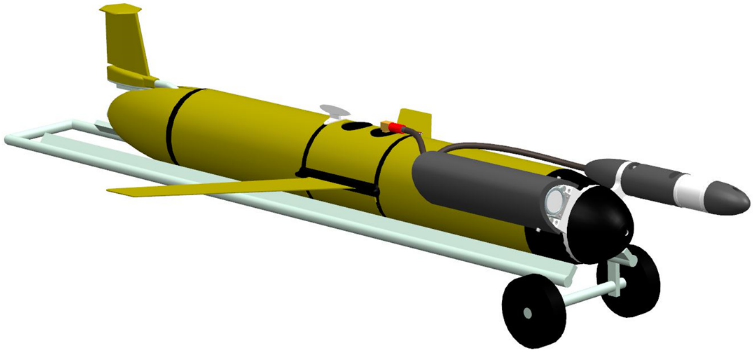


Figure 4.6 3D rendering of UVP6 mounted on a SLOCUM glider sitting on its carrying tray.

types[118], independent of the taxonomic class (Figure 2.23). The morphology of planktonic organisms can reveal both inter-individual variability *within* a taxon (in e.g. size or gut fullness, as translated by their opacity) and common functional traits *across* taxa (e.g. large, transparent organisms are often filter feeders). In addition to the study of plankton diversity in Villefranche, this approach was also recently used to detect a different feeding status of copepods along an ice margin in the Arctic and changes in marine snow composition along the bloom in the same region. All patterns would have been missed by a purely "taxonomic" approach. Therefore, we will study the variability of morphological traits and define morphotypes in the datasets, in addition to taxonomic groups.

T2.3 – Relate taxa and morphs to the environment

The relationships between the potential environmental forcings and the distribution of planktonic taxa and morphotypes will be inferred through multivariate methods that can represent non-linear effects and complex interactions, such as gradient boosted trees[91], which I extensively used in the past. However, any regression method may be confounded by spatial autocorrelation, which makes data points non-independent, especially when points are numerous and densely distributed, like they will be here. We will use distance-based Moran's Eigenvector Maps (dbMEM[55]) as covariates to capture purely spatial effects and be confident that we detect true environmental relationships.

T2.4 – Embed classification of images in autonomous instruments

Finally, to really collect data at scale, over the whole oceans, imaging devices should be installed on autonomous instruments, such as the floats of the Argo network[h]. The miniaturised UVP6, in addition to being usable on gliders, has been engineered to be used on such floats. However, because these floats operate autonomously and have very little communication bandwidth, images need to be classified onboard. To achieve classification at very low power, we extract morphological descriptors based on the central and Hu moment invariants[119] of the objects and

---

[h] http://www.argo.ucsd.edu

classify them, again, through gradient boosted trees. The performance of this approach is currently being tested.

T2.5 – Combine observations of plankton at different scales

With such innovations in place, it would become possible to achieve significant sampling effort across fronts of various dimensions combining (i) floats equipped with UVP6, (ii) gliders adapted for zooplankton (glider with UVP6, ZooGlider), (iii) high frequency sampling devices (e.g. ISIIS) deployed from ships. Floats provide long term observations and can yield numerous examples of local aggregations within submesoscale structures, together with their biogeochemical context. Glider sections are rich, high-resolution snapshots of mesoscale structures that are informative of ecosystem state and can achieve a temporal coverage that cruises cannot. Cruises can inform on faster ecosystem dynamics, through repeated transects across the same water mass, sampled in a Lagrangian way[105], guided by drifters (such as swarms of new miniature floats[120]) and satellite altimetry. The combination of this data is required to understand the influence of submesocale features on zooplankton.

4.4 Infer zooplankton distribution from space

Biological variables can be derived from satellite images

While in situ sampling provides invaluable information, it will always be limited in time and space because of practical constraints. On the other hand, remote sensing through satellites provides us with a quasi-synoptic view of large parts of the oceans (albeit only their surface). The “images” they provide can be interpreted to represent sea surface height (and deduce the dynamics of eddies[121]), sea surface temperature, chlorophyll concentration, and even the coarse composition of phytoplankton communities[122]. To derive the later quantities, the in situ measurements are usually matched to an average satellite pixel and regressed against the multispectral reflectance data (Figure 4.7, top part). This approach is appropriate to predict variables that directly affect the colour of the water (e.g. concentrations of phytoplankton or particulate matter) and for which all the information necessary for the prediction is potentially contained in that average pixel.

This approach was recently extended to zooplankton: the red component of the Visible Infra-red Imaging Radiometer Suite (VIIRS) images was correlated with the presence of swarms of red zooplankton (Calanoid copepods) that accumulated near the surface, off the coast of Norway[123]. The red pigment of these species is associated with their feeding activity and the build-up of lipid reserves . These lipid reserves constitute a large pool of carbon during the highly productive Arctic summer; a pool that is then transferred at depth when the organisms migrate down to hibernate[124], hence contributing to its sequestration.

T3.1 – Quantify zooplankton “redness”…

We will match satellite imagery with in situ estimations of the concentration of Calanoid copepods collected over several past cruises and try to calibrate the satellite signal to quantify these important zooplankton swarms, spatially. Still, optical theory is far from being as developed

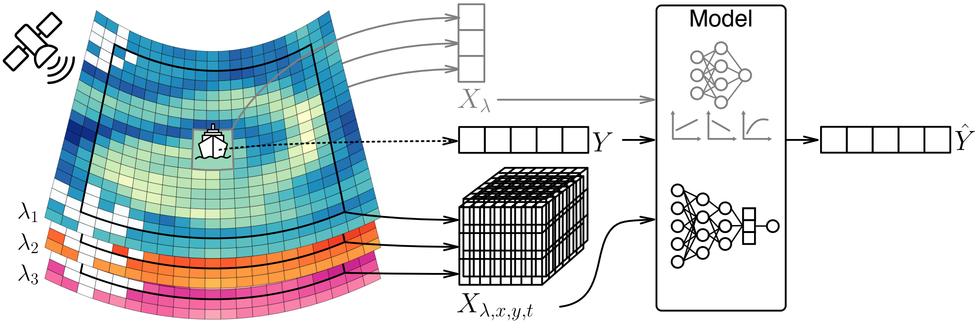


Figure 4.7 The satellite imagery approach. Top (grey): the ground truth data collected by the ship (Y, e.g. phytoplankton functional groups) is matched in space and time with the multispectral satellite data (X_λ, here with three different wavelengths λ), which is averaged over a small region (grey square); Y is correlated with X_λ through models such as Multi-Layer Perceptrons or combination of unitary relationships and can then be predicted ($\hat{Y}$) for any input pixel. Bottom (black; proposed approach): the data from a larger region (black square) and across several time steps is kept as a 4D cube input ($X_{\lambda,x,y,t}$), which is correlated with Y through a CNN.

here as for phytoplankton and little is known regarding how to combine multispectral data to extract relevant "redness" information. Machine learning techniques, such as Random Forests, are particularly adept at selecting relevant variables among a large set of potentially explanatory ones[124]. Such "blind" approaches will provide a useful first step.

…based on individual measurements

However, because pigment concentration depends on the feeding activity of the population, a strong red signal may come from a dense swarm of moderately red individuals or from a looser swarm of intensely coloured individuals. We will use colour images that were collected in the field, together with the samples, to quantify redness at individual level. Extrapolating it to the rest of the population may provide a more accurate measure of in situ redness, to be correlated with the satellite signal. In addition, these individual measurements are interesting in and of themselves, and will be correlated with local environmental variables, to try to understand the environmental conditions that favour intense feeding by these species.

Still, because of the importance of meso and submesoscale structures highlighted previously, it seems relevant to consider the spatio-temporal context of the pixel of interest. Two pixels may contain the same reflectance information but be on the inner vs. outer edge of an eddy; they would be in a very different hydrological context. This context would be particularly relevant to predict new quantities, that depend on the local and temporal distribution of variables that can be remotely sensed (currents, temperature, phytoplankton) but do not have a coloured signature; one such would be the composition of zooplankton communities. To predict a quantity based on its context, I propose to consider the local spatio-temporal context of the in situ measurement as input for the regression model (Figure 4.7). This input would therefore be a 4D cube (wavelength, [x,y] position, time). The convolutions in a CNN can naturally highlight relevant patterns within such a cube and pooling will reduce its dimension to finally correlate it with the response variable (a vector of concentrations of zooplankton taxa).

CNNs could naturally summarise satellite data…

CNNs have already been used with satellite imagery to predict indirectly correlated quantities: time series of Sentinel land images were used to successfully predict household income[125]! However, in those cases, the quantity to predict was representative of the whole area of the image (e.g. average income of a village from an image of the village). Here, we are interested in predicting a *point* measurement, in a given pixel at a given time, but using the information of the neighbouring pixels as context. Therefore, more weight should be given to the central pixel. This can be achieved by either applying gaussian fading to the input data so that the outer pixels have similar values and less is learned from them; or, more elegantly, by modifying the pooling function to give more weight to pixels closer to the centre.

…and can be modified to predict a value from its context

T3.2 – Predict zooplankton communities from space

From EcoTaxa, we will collect existing datasets describing plankton community composition in surface waters (e.g. the ~300 Tara Oceans stations or ~3000 UVP5 casts). Around those, we will extract cubes of satellite data for several variables: ocean colour, temperature, but also derived products, such as maps of finite size Lyapunov exponents[126], that mark oceanic fronts and have proven relevant to predict the distribution of phytoplankton functional types[127]. CMEMS, NASA, and others distribute such products with resolutions down to 4 km; not enough to resolve submesoscale features but sufficient to describe the mesoscale context that creates them (and therefore infer their presence). The spatial size of the extracted cube can originally be fixed but could also be based on the internal Rossby deformation scale computed from sea surface height maps[121]. Its temporal extent would depend on the time scale of ecological processes relevant for the target zooplankton community, likely to be several weeks[128]. Finally, we will use CNNs configured as described above to correlate the composition of the local zooplankton community with its mesoscale context, described by these variables. Finally this should allow to extrapolate plankton composition to every pixel of the world's Ocean.

Chapter 5

Curriculum vitae

Jean-Olivier IRISSON
Maître de Conférences (Associate professor)
Sorbonne Université
Laboratoire d'Océanographie de Villefranche
2 Quai de la Corderie, 06230 Villefranche-sur-Mer
Tel: +33 4 93 76 38 04, Mob: +33 6 21 05 19 90
irisson@normalesup.org
http://www.obs-vlfr.fr/ irisson/

5.1 Curriculum

Maître de Conférences (Associate Professor) at Sorbonne Université	2010 -
Attaché Temporaire d'Éducation et de Recherche (ATER; Assistant Professor) at Université Pierre et Marie Curie (UPMC, now Sorbonne Université)	2009 - 2010
Post Doctoral fellow at RSMAS, University of Miami, USA	2008 - 2009
PhD, Ecology and Oceanography at EPHE and University of Perpignan, France: *Behavioural approach to larval dispersal in marine systems*, magna cum laude	2004 - 2008
"Agrégation" in Life and Earth Sciences, speciality in organisms and population biology, rank 10/2500	2003 - 2004
MSc, Ecology at UPMC and École Normale Supérieure (ENS), France	2001 - 2003
École Normale Supérieure, rue d'Ulm, rank 14/2000	2000 - 2001

5.2 Research

5.2.1 Main projects

2019 - 2023	WWWPIC World Wide Web of Plankton Image Curation (Belmont Forum; 1M€, PI). Setup an international infrastructure for the classification of plankton images assisted by machine learning and the sharing of the resulting datasets.
2019 - 2022	Task 2.3 "Plankton demonstrator" of the BlueCloud project (H2020, EU; 250k€, PI). Use genomics and images jointly to infer new species and functions for unassigned genomic reads from plankton samples; then predict the geographical distribution of these plankton groups, traits, and functions.
2019 - 2021	ARTIFACTZ Application de l'intelligence artificielle à l'identification de traits fonctionnels à partir d'images à haute résolution (Sentinelle-Nord; 40k€, coPI). Use artificial intelligence to extract ecological traits from plankton images.
2019 - 2021	CARDINAL Calanus redness index from artificial intelligence: applications to image analysis (Sentinelle-Nord, The Arctic University of Norway; 154k€, participant). Quantify redness of individual Calanus copepods from images and relate it to redness of surface waters seen from satellites.
2017 - 2018	DL-PIC Deep Learning for Plankton Image Classification (CNRS; 23k€, PI). Use dedicated Convolutional Neural Networks for plankton images classification.
2015 - 2019	Task "Underwater Vision Profiler version 6" of project BRIDGES (H2020, EU; 450k€, participant). Develop an intelligent plankton camera for autonomous instruments; I lead the machine learning developments for onboard classification of images.
2012 - 2016	Biophysical interactions at mesoscale in the ocean (PUF; 240k$, PI). Designed and ran this research, innovation, and education program, with the University of Miami. The program funded three research actions (dealing with plankton and fish larvae), the evolution of two instruments, and directly involved 12 professors and 26 students. http://puf.rsmas.obs-vlfr.fr
2011 - 2016	WP5 "Marine regionalisation planning" of program MERMEX (CNRS; 60k€, PI). I coordinated the data consolidation and valorisation activities of this structuring program; this involved biogeography studies as well as joint research with environmental economists.
2013 - 2015	Task 1.6 "Ecoregionalisation" of project PERSEUS (H2020, EU; 100k€, PI). Use ecological niche modelling and multivariate clustering to define homogeneous regions within the Mediterranean Sea.
2011 - 2014	Biogeographic atlas of the Southern Ocean (SCAR; participant). I led the ecological niche modelling efforts, in particular for krill and fishes.

5.2.2 Peer-reviewed publications

The students and post-docs whom I directly supervised are *underlined and italicised*, those whom I only contributed to supervise are underlined.

2019 *Faillettaz R*, *Voué R*, Crec'hriou R, Garsi L-H, Lecaillon G, Agostini S, Lenfant P, **Irisson J-O**. Spatio-temporal patterns of larval fish settlement in the northwestern Mediterranean Sea. *Marine Ecology Progres Series*, in press

Benedetti F, Ayata S, **Irisson J-O**, Adloff F, Guilhaumon F. Climate change may have minor impact on zooplankton functional diversity in the mediterranean sea. *Diversity and Distributions*, 25(4):568-581

Benedetti F, Jalabert L, Sourisseau M, Beker B, Cailliau C, Desnos C, Elineau A, **Irisson J-O**, Lombard F, Picheral M, Stemmann L, Pouline P. The seasonal and inter-annual fluctuations of plankton abundance and community structure in a North Atlantic Marine Protected Area. *Frontiers in Marine Science*, 6:214

Rossi A, **Irisson J-O**, Levaray M, Pasqualini V, Agostini S. Orientation of mediterranean fish larvae varies with location. *Marine Biology*, 166(8):100

Schröder S, Kiko R, **Irisson J-O**, Koch R. Low-shot learning of plankton categories. GCPR 2018, *Lecture Notes in Computer Science*, vol 11269

2018 Ayata S*, **Irisson J-O***, Berline L, Dutay J, Mayot N, Nieblas A, D'Ortenzio F, Palmiéri J, *Reygondeau G*, Rossi V, Guieu C. Regionalisation of the Mediterranean basin, a MERMEX synthesis. Special issue of *Progress in Oceanography*, 163:7-20. *: co first-author

Faillettaz R, *Durand E*, Paris CB, Koubbi P, **Irisson J-O**. Swimming speeds of mediterranean settlement-stage fish larvae nuance Hjort's aberrant drift hypothesis. *Limnology and Oceanography*, 63(2):509-523

Faillettaz R, Gilletta L, Petit F, Francour P, **Irisson J-O**. First records of dusky grouper *Epinephelus marginatus* settlement-stage larvae in the Ligurian Sea. *Journal of Oceanography, Research and Data*, 10(1)

Faillettaz R, Paris CB, **Irisson J-O**. Larval fish swimming behavior alters dispersal patterns from marine protected areas in the north-western Mediterranean Sea. *Frontiers in Marine Science*, 5:97

Gattuso J, Magnan AK, Bopp L, Cheung WWL, Duarte CM, Hinkel J, Mcleod E, Micheli F, Oschlies A, Williamson P, Billé R, Chalastani V, Gates RD, **Irisson J-O**, Middelburg JJ, Poertner HO, Rau G. Ocean solutions to address climate change and its effects on marine ecosystems. *Frontiers in Marine Science*, 5:337

Luo JY, **Irisson J-O**, Graham B, Guigand C, Sarafraz A, Mader C, Cowen RK. Automatic plankton image analysis using convolutional neural networks. *Limnology and Oceanography: Methods*, 16(12):814-827

2017 *Reygondeau G*, Guieu C, Benedetti F, **Irisson J-O**, Ayata S, Gasparini S, Koubbi P. Biogeochemical regions of the mediterranean sea: an objective multidimensional and multivariate environmental approach. *Progress in Oceanography*, 151:138-148

Chust G, et al. Mare incognitum: a glimpse into future plankton diversity and ecology research. *Frontiers in Marine Science*, 4(8)

Jean M, **Irisson J-O**, Gras G, Bouchand F, Simo D, Duran C, Perronne C, Mulleman D, Bernard L, Dinh A. Diagnostic delay of pyogenic vertebral osteomyelitis and its associated factors. *Scandinavian Journal of Rheumatology*, 46(1):64-68

2016 Ezzat L, Towle E, **Irisson J-O**, Langdon C, Ferrier-Pagès C. The relationship between heterotrophic feeding and inorganic nutrient availability in the scleractinian coral T. reniformis under a short-term temperature increase. *Limnology and Oceanography*, 61(1):89–102

Faillettaz R, Picheral M, Luo JY, Guigand C, Cowen RK, **Irisson J-O**. Imperfect automatic image classification successfully describes plankton distribution patterns. *Methods in Oceanography*, 15-16:60-77

2015 Howes EL, Stemmann L, Assailly C, **Irisson J-O**, Dima M, Bijma J, Gattuso J. Pteropod time series from the north western mediterranean (1967-2003): impacts of pH and climate variability. *Marine Ecology Progress Series*, 531:193-206

Crise A, et al. A MSFD complementary approach for the assessment of pressures, knowledge and data gaps in Southern European Seas: the PERSEUS experience. *Marine Pollution Bulletin*, 95(1):28-39

Sukhovich A, Bonnieux S, Hello Y, **Irisson J-O**, Simons FJ, Nolet G. Seismic monitoring in the oceans by autonomous floats. *Nature Communications*, 6:8027

Irisson J-O, Paris CB, Leis JM, Yerman MN. With a little help from my friends: group orientation by larvae of a coral reef fish. *PLoS ONE*, 10(12): e0144060

Faillettaz R, *Blandin A*, Paris CB, Koubbi P, **Irisson J-O**. Sun-compass orientation in mediterranean fish larvae. *PLoS ONE*, 10(8):e0135213

2014 Cuzin-Roudy J, **Irisson J-O**, Penot F, Kawaguchi S, Vallet C. Chapter 6.9 Southern Ocean Euphausiids. In: *Biogeographic Atlas of the Southern Ocean*, De Broyer C, Koubbi P, Griffiths HJ, Raymond B, Udekem d'Acoz C, et al (eds). Scientific Committee on Antarctic Research, pp 309-320

Duhamel G, Hulley P, Causse R, Koubbi P, Vacchi M, Pruvost P, Vigetta S, **Irisson J-O**, Mormede S, Belchier M, Dettai A, Detrich HW, Gutt J, Jones CD, Kock K, Lopez Abellan LJ, Van de Putte A. Chapter 7 Biogeographic patterns of fish. In: *Biogeographic Atlas of the Southern Ocean*, De Broyer C, Koubbi P, Griffiths HJ, Raymond B, Udekem d'Acoz C, et al (eds). Scientific Committee on Antarctic Research, pp 328-362

Leis J, Paris CB, **Irisson J-O**, Yerman M, Siebeck U. Orientation of fish larvae in situ is consistent among locations, years and methods, but varies with time of day. *Marine Ecology Progress Series* 505:193-208

Luo JY, Grassian B, Tang D, **Irisson J-O**, Greer A, Guigand C, McClatchie S, Cowen R. Environmental drivers of the fine-scale distribution of a gelatinous zooplankton community across a mesoscale front. *Marine Ecology Progress Series* 510:129-149

Mormède S, **Irisson J-O**, Raymond B. Chapter 2.3 Distribution modelling. In: *Biogeographic Atlas of the Southern Ocean*, De Broyer C, Koubbi P, Griffiths HJ, Raymond B, Udekem d'Acoz C, et al (eds). Scientific Committee on Antarctic Research, pp 27-29

Sukhovich A*, **Irisson J-O***, Perrot J, Nolet G. Automatic recognition of T and teleseismic P waves by statistical analysis of their spectra: An application to continuous records of moored hydrophones. *Journal of Geophysical Research: Solid Earth* 119(8):6469-6485. *: co first-author

2013 Giraldo C, Mayzaud P, Tavernier E, **Irisson J-O**, Penot F, Becciu J, Chartier A, Boutoute M, Koubbi P. Lipid components as a measure of nutritional condition in fish larvae (*Pleuragramma antarcticum*) in East Antarctica. *Marine Biology* 160:877-887

Paris CB, Atema J, **Irisson J-O**, Kingsford M, Gerlach G, Guigand C. Reef Odor: A Wake Up Call for Navigation in Reef Fish Larvae. *PLoS ONE* 8(8):e72808

Talec A, Philistin M, Ferey F, Walenta G, **Irisson J-O**, Bernard O, Sciandra A (2013) Effect of gaseous cement industry effluents on four species of microalgae. *Bioresource Technology* 143:353-35

2011 Sukhovich A, **Irisson J-O**, Simons FJ, Ogé A, Hello Y, Deschamps A, Nolet G. Automatic discrimination of underwater acoustic signals generated by teleseismic P-waves: A probabilistic approach. *Geophysical Research Letters*, 38(L18605)

2010 Hubert N, Delrieu-Trottin E, **Irisson J-O**, Meyer C, Planes S. Identifying coral reef fish larvae through DNA barcoding: a test case with the families Acanthuridae and Holocentridae. *Molecular Phylogenetics and Evolution* 55(3):1195-203

Irisson J-O, Paris CB, Guigand C, Planes S. Ontogenetic vertical "migration" in fish larvae: description and consequences for dispersal. *Limnology & Oceanography* 55(2):909-919

Lett C, Ayata S, Huret M, **Irisson J-O**. Biophysical modelling to investigate the effects of climate change on marine populations dispersal and connectivity. Special issue of *Progress in Oceanography* 87:106-112

2009 **Irisson J-O**, Guigand C, Paris CB. Detection and quantification of marine larvae orientation in the pelagic environment. *Limnology & Oceanography: Methods* 7:664–672

Irisson J-O, Leis JM, Paris CB, Browman H. Biological processes: behavior and settlement. In: *Manual of Recommended Practices for Modelling Physical-Biological Interactions in Fish Early-Life History*, North E, Gallego A, Petigas P (eds). ICES Cooperative Research Reports, 295:42-59

Paris CB, **Irisson J-O**, Lacroix G, Fiksen Ø, Leis JM, Mullon C. Application 2: Connectivity. In: *Manual of Recommended Practices for Modelling Physical-Biological Interactions in Fish Early-Life History*, North E, Gallego A, Petigas P (eds). ICES Cooperative Research Reports, 295:63-76

2008 Paris CB, Guigand CM, **Irisson J-O**, Fisher R, D'Alessandro E. Orientation with no frame of reference (OWNFOR): a novel system to observe and quantify orientation in reef fish larvae. In: *Caribbean connectivity: implications for marine protected area management*, Grober-Dunsmore R and Keller DB (eds). Marine Sanctuaries Conservation Series ONMS-08-07:52-62

Irisson J-O, Lecchini D. In situ observation of settlement behaviour in larvae of coral reef fishes at night. *Journal of Fish Biology* 72, 2707–2713

2004 **Irisson J-O**, Levan A, de Lara M, Planes S. Strategies and trajectories of coral reef fish larvae optimizing self-recrutment. *Journal of Theoretical Biology* 227, 205–218

2003 Dercole F, **Irisson J-O**, Rinaldi S. Bifurcation analysis of a prey-predator coevolution model. *SIAM Journal on Applied Mathematics* 63:4, 1378–1391

5.2.3 Conferences

I contributed to 38 conference presentations and 9 posters. I list here only the ones I personally presented or directly supervised.

2020 **Irisson J-O**, *Cailleton C*, Desnos C, Jalabert L, Elineau A, Stemmann L, Ayata S-D (2020) Morphological diversity increases with oligotrophy along a zooplankton time series. In *Ocean Sciences Meeting*. San Diego, California, USA

Panaiotis T, Babin M, Biard T, Carlotti F, Coppola L, Guidi L, Hauss H, Karp-Boss L, Kiko R, Lombard F, McDonnell AMP, Picheral M, Rogge A, Waite AM, **Irisson J-O**, Stemmann L (2020) Typology of plankton communities seen by in situ imaging, from the epi to the mesopelagic layers of the global ocean. In *Ocean Sciences Meeting*. San Diego, California, USA

Vilgrain L, **Irisson J-O**, Ayata S-D, Picheral M, Babin M, Maps F (2020) Morphological traits of zooplankton reveal ecological patterns along ice melt dynamics in the Arctic. In *Ocean Sciences Meeting*. San Diego, California, USA

2019 *Faillettaz R*, Paris CB, *Durand E*, *Blandin A*, Koubbi P, **Irisson J-O**. Larval fish behavior nuances Hjort's aberrant drift hypothesis. In *Larval Fish Conference*. Palma de Mallorca, Spain

Faillettaz R, *Voué R*, Crec'hriou R, Garsi L-H, Lecaillon G, Agostini S, Lenfant P, **Irisson J-O**. Spatio-temporal patterns of larval fish settlement in the Northwestern Mediterranean Sea. Poster in *Larval Fish Conference*. Palma de Mallorca, Spain

Faillettaz R, Paris CB, *Durand E*, **Irisson J-O**. Modelling fish dispersal and settlement in the Mediterranean Sea. In *MODELIFE*. Nice, France

Paris CB, *Faillettaz R*, *Blandin A*, Koubbi P, **Irisson J-O**. Sun-compass orientation in Mediterranean fish larvae. Poster in *Larval Fish Conference*. Palma de Mallorca, Spain

2018 *Cailleton C*, Desnos C, Jalabert L, Elineau A, Stemmann L, Ayata S-D, **Irisson J-O**. Morphological diversity increases with oligotrophy along a zooplankton time series. In *SFEcologie 2018*. Rennes, France

Faillettaz R, Paris CB, **Irisson J-O**. Larval fish swimming behavior alters dispersal patterns from marine protected areas in the North-Western Mediterranean Sea. In *Larval Fish Conference*. Victoria, British Columbia, Canada

Faillettaz R, Picheral M, Luo JY, Guigand CM, Cowen RK, **Irisson J-O**. Imperfect automatic image classification successfully describes plankton distribution patterns. In: *ICES Workshop on Machine Learning in Marine Sciences*. Copenhagen, Denmark

Irisson J-O, Schröder S-M, Picheral M. EcoTaxa: A human-computer interface to classify images along a taxonomy with the help of machine-learning. In: *ICES Workshop on Machine Learning in Marine Sciences*. Copenhagen, Denmark

Irisson J-O and COMPLEx team. Monitoring plankton in the open ocean with quantitative imaging. Invited presentation in *Intergovernmental Conference on Marine Biodiversity of Areas Beyond National Jurisdiction*. United Nations, New-York, USA

2017 **Irisson J-O**, Ocean Observation team of OOV. Dynamic visualisation of ocean observation data. Poster in *RESOMAR*. Bordeaux, France.

2016 *Faillettaz R*, *Durand E*, Paris CB, Koubbi P, **Irisson J-O**. Swimming speeds of Mediterranean settlement-stage fish larvae nuance Hjort's aberrant drift hypothesis. In *SFEcologie 2016*. Marseille, France

2015 *Faillettaz R*, *Blandin A*, *Durand E*, Paris CB, **Irisson J-O**. Behaviour versus oceanic currents during the dispersal of larval fish in the Northwestern Mediterranean sea. In A*quatic Sciences Meeting*. Granada, Spain

Faillettaz R, **Irisson J-O**[*], Koubbi P, Paris CB, *Durand E*, *Blandin A*. Les larves de poissons Méditerranéens, championnes de la course d'orientation. Invited presentation in *DRIVER-SUBLIMO*. Les Embiez, France. [*]: invited presenter

Irisson J-O, *Faillettaz R*, Luo JY, Guigand CM, Cowen RK. Fine-scale distribution of zooplankton over a mesoscale front explored through high frequency imaging. In *Aquatic Sciences Meeting*. Granada, Spain

2014 *Faillettaz R*, Luo JY, Guigand CM, Cowen RK, **Irisson J-O**. Fine-scale distribution of larval fish and zooplankton over a mesoscale front explored through high frequency imaging. In *Ocean Sciences Meeting*. Honolulu, HI, USA

Reygondeau G, Albouy C, Hattab T, Benedetti F, **Irisson J-O**, Ayata S-D, Gasparini S, McKenzie C, Koubbi P.Mediterranean biodiversity (from phytoplankton to top predators) and present threats. In *ICES*. A Coruña, Spain

Reygondeau G, **Irisson J-O**, Albouy C, Ayata S-D, Benedetti F, Guieu C, Gasparini S, Koubbi P (2014b) Ecological geography of the Mediterranean Sea. In **IMBER**. Bergen, Norway

2013 **Irisson J-O**. Autoplot: ready made plots with ggplot2. In *Rencontres R*. Lyon, France

Faillettaz R, **Irisson J-O**. Drivers of temporal variability of larval fish abundance in Villefranche Bay (Northwestern Mediterranean Sea) over a 7-year time series (2006-2012). In *Larval Fish Conference*. Miami, FL, USA.

Reygondeau G, **Irisson J-O**, Albouy C, Ayata S-D, Benedetti F, Guieu C, Gasparini S, Koubbi P. Ecoregionalisation of the Mediterranean Sea. Poster in *CIESM*. Marseille, France

2012 **Irisson J-O**, *Webb A*, Passafiume O, Mousseau L. Detecting hydrologic seasons in a long term monitoring time series. In *Time-series analysis in marine science and applications for industry*. Brest, France

2009 **Irisson J-O**, Paris CB, Cowen RK, Planes S. Ontogenetic vertical "migration" in coral-reef fish larvae communities and its consequences for dispersal. In *Aquatic Sciences Meeting*. Nice, France

2008 **Irisson J-O**, Chérubin LM, Planes S. Consequences of increased mobility and quicker development in warmer waters on the dispersal trajectories of fish larvae. In *Ocean Sciences Meeting*. Orlando, FL, USA. Best student presentation award.

Irisson J-O, Paris CB, De Lara M, Planes S (2008b) The importance of behavior for self-recruitment: a modelling approach. In *International Coral Reef Symposium*. Fort Lauderdale, FL, USA

2006 **Irisson J-O**, Paris CB, Guigand CM, Fisher R. Orientation With No Frame of Reference (OWNFOR): An in situ system to detect and measure orientation in pelagic reef fish larvae. *Ocean Sciences Meeting*. Honolulu, HI, USA

2005 **Irisson J-O**, De Lara M, Planes S. Lagrangian or not lagrangian? In *LAPCOD meeting*. Lerici, Italy

5.3 Personnel supervision

The future situation of the students/post-docs I directly supervised is specified after the title of their work. When the work resulted in published papers, they are mentioned again here.

5.3.1 PhDs and post-docs

Cédric Dubois, PhD (co-supervisor), CNRS-Université Côte d'Azur: *Classification du plancton conjointe en espèces et traits morphologiques et fonctionnels, avec contrainte de relations espèces-traits et de hiérarchie des espèces selon une phylogénie*. Current

Salomé Fabri-Ruiz, Post-doc (supervisor), LOV: *Prediction of the future niche of Ostreopsis cf. ovata in response to climate change*.

Thelma Panaïotis, PhD (main supervisor), Sorbonne Université: *Écologie du plancton à submésoéchelle, apport de l'imagerie et de l'intelligence artificielle*.

Laure Vilgrain, PhD (co-supervisor), Université Laval-Sorbonne Université: *Approche par traits fonctionnels de l'écologie des copépodes arctiques et subarctiques*.

Amelie Rossi, PhD (committee and jury member), Univ Pasquale Paoli: *Approche morphologique et comportementale de jeunes stades de poissons côtiers méditerranéens sur le littoral Nord-Est de la Corse (Méditerranée Nord-Occidentale)*. 2018

- Rossi A, Irisson J-O, Levaray M, Pasqualini V, Agostini S. Orientation of mediterranean fish larvae varies with location. *Marine Biology*, 166(8):100 (2019).

Marc Besson, PhD (committee member), EPHE/Univ Perpignan: *The importance of metamorphosis in the larval recruitment of coral reef fishes facing climate change and pollutions*. 2017

- Besson M, Gache C, Irisson J-O, Berthe C, Laudet V, Lecchini D. Decreased abundance of fish in bleached anemones result from pre- and post-settlement impairments. *In prep*

Leïla Ezzat, PhD (committee member), Univ Pierre et Marie Curie: *Effets de la disponibilité en sels nutritifs sur la réponse physiologique des coraux tropicaux dans le contexte du changement climatique*. 2016

- Ezzat L, Towle E, Irisson J-O, Langdon C, Ferrier-Pagès C. The relationship between heterotrophic feeding and inorganic nutrient availability in the scleractinian coral T. reniformis under a short-term temperature increase. *Limnology and Oceanography*, 61(1):89–102 (2016)

Robin Faillettaz, PhD (main supervisor), Univ Pierre et Marie Curie: *Estimation des capacités comportementales des larves de poissons et leurs implications pour la phase larvaire, Un cas d'étude d'espèces démersales de Méditer-* 2015 *ranée Nord-Occidentale*. Continued as a post-doc at LOG, France, then Univ Miami, USA, is now permanently employed as researcher at IFREMER.

- Faillettaz R, Voué R, Crec'hriou R, Garsi L-H, Lecaillon G, Agostini S, Lenfant P, Irisson J-O. Spatio-temporal patterns of larval fish settlement in the northwestern Mediterranean Sea. *Marine Ecology Progres Series*, in press
- Faillettaz R, Durand E, Paris CB, Koubbi P, Irisson J-O. Swimming speeds of mediterranean settlement-stage fish larvae nuance Hjort's aberrant drift hypothesis. *Limnology and Oceanography*, 63(2):509-523 (2018)
- Faillettaz R, Gilletta L, Petit F, Francour P, Irisson J-O. First records of dusky grouper *Epinephelus marginatus* settlement-stage larvae in the Ligurian Sea. *Journal of Oceanography, Research and Data*, 10(1) (2018)
- Faillettaz R, Paris CB, Irisson J-O. Larval fish swimming behavior alters dispersal patterns from marine protected areas in the northwestern mediterranean sea. *Frontiers in Marine Science*, 5:97 (2018)
- Faillettaz R, Picheral M, Luo JY, Guigand C, Cowen RK, Irisson J-O. Imperfect automatic image classification successfully describes plankton distribution patterns. *Methods in Oceanography*, 15-16:60-77 (2018)
- Faillettaz R, Blandin A, Paris CB, Koubbi P, Irisson J-O. Sun-compass orientation in mediterranean fish larvae. *PLoS ONE*, 10(8): e0135213 (2018)

Jessica Y Luo, PhD (committee and jury member), Univ Miami: *Gelatinous zooplankton in marine communities and ecosystems: fine-scale horizontal and vertical distribution, trophic drivers, and contribution to global carbon cycling*.

- Luo JY, Irisson J-O, Graham B, Guigand C, Sarafraz A, Mader C, Cowen RK. Automatic plankton image analysis using convolutional neural networks. *Limnology and Oceanography: Methods*, 16(12):814-827 (2018)
- Faillettaz R, Picheral M, Luo JY, Guigand C, Cowen RK, Irisson J-O. Imperfect automatic image classification successfully describes plankton distribution patterns. *Methods in Oceanography*, 15-16:60-77 (2016)
- Luo JY, Grassian B, Tang D, Irisson J-O, Greer A, Guigand C, McClatchie S, Cowen R. Environmental drivers of the fine-scale distribution of a gelatinous zooplankton community across a mesoscale front. *Marine Ecology Progress Series* 510:129-149 (2014)

2014 Gabriel Reygondeau, Post-doc (supervisor), LOV: *Regionalisation of the Mediterranean Sea*. Is now employed at University of British Columbia.

- Ayata S*, Irisson J-O*, Berline L, Dutay J, Mayot N, Nieblas A, D'Ortenzio F, Palmiéri J, Reygondeau G, Rossi V, Guieu C. Regionalisation of the mediterranean basin, a MERMEX synthesis. Special issue of *Progress in Oceanography*, 163:7-20 (2018) *: co first-author

- Reygondeau G, Guieu C, Benedetti F, Irisson J-O, Ayata S, Gasparini S, Koubbi P. Biogeochemical regions of the mediterranean sea: an objective multidimensional and multivariate environmental approach. *Progress in Oceanography*, 151:138-148 (2017)

5.3.2 Master students

Thelma Panaïotis, Master 2 Sciences de la Mer, Sorbonne Université: *Typologie globale de la distribution tridimensionnelle des communautés de plancton vues par imagerie in situ*. Continued as a PhD student at LOV, France. 2019

Laure Vilgrain, Master 2 Sciences de la Mer, Sorbonne Université: *Identification de traits morphologiques du zooplancton arctique à partir d'imagerie in situ et description de leurs distributions selon les gradients environnementaux en Baie de Baffin*. Continued as a PhD student at LOV, France.

Caroline Cailleton, Master 2 Sciences de la mer et du littoral, Univ Bretagne Occidentale: *Étude de la diversité morphologique du zooplancton marin le long d'une série temporelle (2009-2017) en Méditerranée*. Is now employed in a biological conservation organisation. 2018

Justine Courboulès, Master 2 Océanographie et Environnement Marin, Univ Pierre et Marie Curie: *Étude comparative des changements à long terme des écosystèmes côtiers et hauturiers en Méditerranée Nord-Occidentale*. Continued as a PhD student at MARBEC, Sète. 2017

Raphaël Voué, DU, Univ Montpellier: *Dynamique spatio-temporelle du recrutement des poissons côtiers en Méditerranée Française*.

- Faillettaz R, Voué R, Crec'hriou R, Garsi L-H, Lecaillon G, Agostini S, Lenfant P, Irisson J-O. Spatio-temporal patterns of larval fish settlement in the northwestern Mediterranean Sea. *Marine Ecology Progres Series*, in press

Mégane Tetaz, Master 1 Biologie et Santé de l'Environnement, Univ Nice: *Influence de la qualité du développement larvaire de poissons méditerranéens sur leur vitesse de nage*. 2015

Agathe Blandin, Master 1 Océanographie et Environnement Marin, Univ Pierre et Marie Curie: *Comportement d'orientation de larves de poissons côtiers en phase de recrutement en mer Méditerranée*. 2014

- Faillettaz R, Blandin A, Paris CB, Koubbi P, Irisson J-O. Sun-compass orientation in mediterranean fish larvae. *PLoS ONE*, 10(8): e0135213 (2018)

Elysanne Durand, Master 1 Océanographie et Environnement Marin, Univ Pierre et Marie Curie: *Vitesse de nage des larves de poissons démersaux méditerranéens et conséquences pour la dispersion*.

- Faillettaz R, Durand E, Paris CB, Koubbi P, Irisson J-O. Swimming speeds of mediterranean settlement-stage fish larvae nuance Hjort's aberrant drift hypothesis. *Limnology and Oceanography*, 63(2):509-523 (2018)

Federica Ferrando, Master, Univ di Messina: *Utilizzo di due differenti metodologie di analisi per lo studio della comunità zooplanctonica in aree idrologicamente diverse del Mediterraneo*. Internship resulted in a published paper.

2013 Mariam Sy, Master 2, Agriculture, Alimentation et Développement Durable, Univ Montpellier: *Indice de services écosystémiques en Méditerranée*. Continued as a PhD student at Univ Montpellier.

2012 Alice Webb, Master 2 Oceanography, IMBRSea: *Hydrological seasonal and inter-annual variability in the Bay of Villefranche sur Mer during 1995-2011 and biochemical consequences*.

2011 Leïla Ezzat, Master, Ecole Polytechnique Fédérale de Lausanne: *Distribution verticale de larves de poissons coralliens à la Barbade*. Continued as a PhD student in Monaco.

2010 Elise Lacoste, Master 2 Océanographie et Environnement Marin, Univ Pierre et Marie Curie: *Couplage entre la distribution des larves de poissons coralliens et de leurs proies potentielles*. Continued as a PhD student in EPHE and Univ Perpignan.

2009 Erwan Delrieu-Trottin, Master 2 Océanographie et Environnement Marin, Univ Pierre et Marie Curie: *Distribution des larves de poissons coralliens: stratégies écologiques et corrélats environnementaux*. Continued as a PhD sutdent in EPHE/Univ Perpignan.

- Hubert N, Delrieu-Trottin E, Irisson J-O, Meyer C, Planes S. Identifying coral reef fish larvae through DNA barcoding: a test case with the families Acanthuridae and Holocentridae. *Molecular Phylogenetics and Evolution* 55(3):1195-203 (2010)

2007 Pierre Torres, Master 1 Écologie Fonctionnelle, Univ Perpignan: *Importance écologique de la distribution verticale des larves de poissons coralliens*.

2006 Aubin Mboumba, Master 1 Environnement Mediterranéen, Univ Perpignan: *Distribution verticale des larves de poissons coralliens autour de la Barbade*.

5.3.3 Undergraduate students

2018 Lauriane Bergeon, L3, Univ Montpellier: *Suivi historique des population de plancton en rade de Villefranche sur mer*. Continued in a Master of Ecology at Univ Montpellier.

2016 Joshua Bac, Intechmer: *Observation de l'hydrologie, du zooplancton et des larves de poissons en rade de Villefranche*.

2015 Julie Lafaye, ESPCI 1ère année: *Suivi du recrutement des larves de poissons côtiers en Méditerranée*.

Agathe Maupetit, VetAgro'Sup: *Étude de la distribution des organismes planctoniques dans le courant Ligure*. Internship contributed to 2014

- Faillettaz R, Picheral M, Luo JY, Guigand C, Cowen RK, Irisson J-O. Imperfect automatic image classification successfully describes plankton distribution patterns. *Methods in Oceanography*, 15-16:60-77 (2018)

Alexis Deru-Denise, INP ENSE3, Grenoble: *Reconnaissance du zooplancton assistée par ordinateur*. 2012

5.4 Teaching

I taught 64h per year as teaching assistant during my PhD and teach over 200h per year as associate professor since 2009, mostly at Masters level. I am in charge of four teaching modules.

Instrumentation and Acquisition of Data in Oceanography (coordinator), Master 2 Sciences de la Mer, Sorbonne Univ: *Scientific cruise planning* (3h), *Databases for oceanography* (6h), *Reading and plotting data with R* (3h), *Plankton imaging* (4.5h), *Supervision of student projects* (6h) 2009-…

Methods for Data Exploitation in Oceanography (coordinator), Master 2 Sciences de la Mer, Sorbonne Univ: *R for data science* (6h), *Mapping and interpolation* with practical (3+6h), *Introduction to multivariate data analysis* with practical (3+3h), *Signal processing* with practical (3+3h), *Supervision of student projects* (6h) [and previously: *Oceanographic "big" data processing* with practical (3+3h)]

Modelling of Marine Ecosystems, Master 2 Sciences de la Mer, Sorbonne Univ: class + practical for all topics, *Lagrangian models* (3+3h), *Advanced population dynamics* (3+3h), *Supervision of student projects* (3h) [and previously: *Optimal behaviour* (3+3h)]

Multivariate Statistics (coordinator), Summer school, Sorbonne Univ: class + practical for all topics, *Introduction to programming with R* (6h), *Data series* (6+6h), *Redundancy analysis* (3+3h), *Tree-based learning* (3+3h) [and previously: *Multivariate regression* (3+3h)]

Agrégation SVTU, Sorbonne Univ: *Behavioural Ecology* (8h), *Sexual selection* (4h), *"Leçons"* (supervision of lectures by students; 5h). 2010-…

Plankton diversity, Summer school, Sorbonne Univ: *Ichthyoplankton diversity*, with practical (3h)

Exploited marine resources (co-coordinator), Master 2 Sciences de la Mer, Sorbonne Univ: *Ichthyoplankton ecology*, with modelling practical (3+3h) 2018-…

Introduction to modelling, Master 1 Sciences de la Mer, Sorbonne Univ: *Introduction to population dynamics* (4h) 2017-2019

2009-2018 Response of plankton to environmental changes, Master 2, Sciences de la Mer, Sorbonne Univ: *Ichthyoplankton's response to its environment* (3h)

2016-2017 Biological processes in the ocean, Master 1 Sciences de la Mer, Sorbonne Univ: *Connectivity of marine populations*, with modelling practical (6h)

2014-2016 Research education, License Biologie, Sorbonne Univ: *Scientific communication* (7h), *Connectivity and marine reserves* (7h)

2012-2016 Coordinator of a student and professor exchange program between the oceanography departments of Univ Pierre et Marie Curie and of the Rosenstiel School of Marine and Atmospheric Sciences or Univ Miami (PUF-funded project).

2015 Graduate courses, RSMAS, Univ Miami: Multivariate Statistics, similar to the one taught at Sorbonne Univ (above) but adapted to the graduate program at RSMAS.

2013 Graduate courses, RSMAS, Univ Miami: *Ichthyoplankton response to physical forcings* (3h)

2009-2015 GIS and spatial processes, Master 2, Sciences de la Mer, Sorbonne Univ: *Introduction to R* (3h), *Species distribution modelling* (3h)

2009-2010 Introduction to modelling for conservation, Master 2, Sciences de la Mer, Sorbonne Univ: *Modelling population dynamics* (8h)

2005-2008 Agrégation SVTU, Univ Montpellier: *The ecology of sociality* (8h)

2004-2008 Biology, License pluridisciplinaire 3, Univ Perpignan: 60h of classes on general biology (from cell processes to environmental sciences); I was coordinator of the biology curriculum, taught most classes and graded all exams

2006-2007 Reef Ecology, EPHE/Univ Perp: *Hydrodynamics, modelling and larval flux* (3h)

2002-2006 Master 1 Ecology, École Normale Supérieure/Univ Pierre Marie Curie: *Bifurcation analysis* (6h), *Interspecific relationships* (3h)

5.5 Community activities

2019-… Leader of team COMPutational PLankton Ecology (COMPLEx; 10 permanent staff, 10 contract staff), Laboratoire d'Océanographie de Villefranche (LOV).

Member of the European Marine Board panel on Big Data.

Chair of the International Council for the Exploration of the Sea (ICES) working group on Machine Learning for Marine Sciences (WGMLEARN).

2013-… Member of the board of the Société Française d'Écologie et d'Évolution (French Society for Ecology and Evolution).

Leader of team Processus dans les Ecosystèmes Pélagiques (PEPS; 19 permanent staff, 15 contract staff), Laboratoire d'Océanographie de Villefranche (LOV). 2016-2018

Panel member of the Commission Nationale de la Flotte Côtière (CNFC). 2015-2018

Coordinator of the ecosystem observation actions at Observatoire Océanologique de Villefranche. 2013-2017

Reviewer for Aquatic Biology, Canadian Journal of Fisheries and Aquatic Sciences, Coral Reefs, Ecology Letters, Ecological Complexity, Estuarine Coastal and Shelf Science, Global Change Biology, ICES Journal of Marine Science, Ichthyological Research, Integrative and Comparative Biology, Interface Focus, Journal of Plankton Research, Journal of Statistical Software, Limnology and Oceanography: Methods, Marine Biology, Nature Communications, PCI Ecology, PeerJ, PLoS ONE, The R Journal. Ongoing

Reviewer or panel member for Agence Nationale de la Recherche (ANR), Belspo, CNRS (INSU, LEFE and GMMC), Fondation pour la Recherche sur la Biodiversité (FRB), Haut Conseil de l'Évaluation de la Recherche et de l'Enseignement Supérieur (HCÉRES), Institut Français de Recherche pour l'Exploitation de la MER (IFREMER), National Geographic, Research Foundation - Flanders (FWO), Sorbonne Université.

Bibliography

[1] Hjort J. Fluctuations in the Great Fisheries of Northern Europe Viewed in the Light of Biological Research. *Rapp P-V Réun Cons Perm Int Explor Mer* **20**, 1–228 (1914).

[2] Cowen RK, Gawarkiewicz G, Pineda J, Thorrold SR, Werner FE. Population Connectivity in Marine Systems: An Overview. *Oceanography* **20**, 14–21 (2007).

[3] Sale PF. Reef Fish Communities: Open Nonequilibrial Systems. In *The Ecology of Fishes on Coral Reefs*, Academic Press, San Diego, California, USA, 564–598 (1991).

[4] Cushing DH. The Natural Regulation of Fish Populations. *Sea fisheries research* 399–412 (1974).

[5] Roberts CM. Connectivity and Management of Caribbean Coral Reefs. *Science* **278**, 1454–1457 (1997).

[6] Leis JM, Sweatman HPA, Reader SE. What the Pelagic Stages of Coral Reef Fishes Are Doing out in Blue Water: Daytime Field Observations of Larval Behavioural Capabilities. *Aust J Mar Fresh Res* **47**, 401–411 (1996).

[7] Leis JM. Are Larvae of Demersal Fishes Plankton or Nekton? *Adv Mar Biol* **51**, 57–141 (2006).

[8] Stobutzki IC, Bellwood DR. An Analysis of the Sustained Swimming Abilities of Pre- and Post-Settlement Coral Reef Fishes. *Journal of Experimental Marine Biology and Ecology* **175**, 275–286 (1994).

[9] Leis JM, Hay AC, Trnski T. In Situ Ontogeny of Behaviour in Pelagic Larvae of Three Temperate, Marine, Demersal Fishes. *Marine Biology* **148**, 655–669 (2005).

[10] Stemmann L, Prieur L, Legendre L, Taupier-Letage I, Picheral M, Guidi L, Gorsky G. Effects of Frontal Processes on Marine Aggregate Dynamics and Fluxes: An Interannual Study in a Permanent Geostrophic Front (NW Mediterranean). *Journal of Marine Systems* **70**, 1–20 (2008).

[11] Paris CB, Guigand CM, Irisson JO, Fisher R, D'Alessandro E. Orientation With No Frame of Reference (OWNFOR): A Novel System to Observe and

Quantify Orientation in Reef Fish Larvae. In *Caribbean Connectivity: Implications for Marine Protected Area Management. Proceedings of a Special Symposium, 9-11 November 2006, 59th Annual Meeting of the Gulf and Caribbean Fisheries Institute, Belize City, Belize*, Grober-Dunsmore R, Keller BD, eds. U.S. Department of Commerce, National Oceanic and Atmospheric Administration, National Marine Sanctuary Program, Silver Spring, MD (2008), Marine Sanctuaries Conservation Series ONMS-08-07, 52–62.

[12] Irisson JO, Guigand C, Paris CB. Detection and Quantification of Marine Larvae Orientation in the Pelagic Environment. *Limnology and Oceanography: Methods* **7**, 664–672 (2009).

[13] Paris CB, Atema J, Irisson JO, Kingsford M, Gerlach G, Guigand CM. Reef Odor: A Wake up Call for Navigation in Reef Fish Larvae. *PLoS ONE* **8**, e72808 (2013).

[14] Leis JM, Paris CB, Irisson JO, Yerman MN, Siebeck UE. Orientation of Fish Larvae in Situ Is Consistent among Locations, Years and Methods, but Varies with Time of Day. *Marine Ecology Progress Series* **505**, 193–208 (2014).

[15] Irisson JO, Paris CB, Leis JM, Yerman MN. With a Little Help from My Friends: Group Orientation by Larvae of a Coral Reef Fish. *PLoS ONE* **10**, e0144060 (2015).

[16] Simons AM. Many Wrongs: The Advantage of Group Navigation. *Trends Ecol Evol* **19**, 453–455 (2004).

[17] Couzin ID, Krause J, Franks NR, Levin SA. Effective Leadership and Decision-Making in Animal Groups on the Move. *Nature* **433**, 513–516 (2005).

[18] Faillettaz R, Blandin A, Paris CB, Koubbi P, Irisson JO. Sun-Compass Orientation in Mediterranean Fish Larvae. *PLoS ONE* **10**, e0135213 (2015).

[19] Batschelet E. *Circular Statistics in Biology*. Academic Press, London (1981).

[20] Miller TJ. Contribution of Individual-Based Coupled Physical–Biological Models to Understanding Recruitment in Marine Fish Populations. *Marine Ecology Progress Series* **347**, 127–138 (2007).

[21] North E, Gallego A, Petitgas P. Manual of Recommended Practices for Modelling Physical-Biological Interactions during Fish Early Life. *ICES Coop Res Rep* **295** (2009).

[22] Irisson JO, Leis JM, Paris CB, Browman H. Biological Processes: Behaviour and Settlement. In *Manual of Recommended Practices for Modelling Physical-Biological Interactions during Fish Early Life*, North E, Gallego A, Petitgas P, eds., ICES Cooperative Research Report, vol. 295, 42–59 (2009).

[23] Faillettaz R, Durand E, Paris CB, Koubbi P, Irisson JO. Swimming Speeds of Mediterranean Settlement-Stage Fish Larvae Nuance Hjort's Aberrant Drift Hypothesis. *Limnology and Oceanography* **63**, 509–523 (2018).

[24] Faillettaz R, Paris CB, Irisson JO. Larval Fish Swimming Behavior Alters Dispersal Patterns from Marine Protected Areas in the North-Western Mediterranean Sea. *Frontiers in Marine Science* **5**, 97 (2018).

[25] MacArthur RH, Pianka ER. On Optimal Use of a Patchy Environment. *The American Naturalist* **100**, 603–609 (1966).

[26] Perry G, Pianka ER. Animal Foraging: Past, Present and Future. *Trends in Ecology and Evolution* **12**, 360–364 (1997).

[27] Houde ED. Comparative Growth, Mortality, and Energetics of Marine Fish Larvae: Temperature and Implied Latitudinal Effects. *Fisheries Bulletin* **87**, 471–495 (1989).

[28] Houde ED. Emerging from Hjort's Shadow. *J Northw Atl Fish Sci* **41**, 53–70 (2008).

[29] Irisson JO, LeVan A, De Lara M, Planes S. Strategies and Trajectories of Coral Reef Fish Larvae Optimizing Self-Recruitment. *Journal of Theoretical Biology* **227**, 205–218 (2004).

[30] Lett C, Ayata SD, Huret M, Irisson JO. Biophysical Modelling to Investigate the Effects of Climate Change on Marine Populations Dispersal and Connectivity. *Progress in Oceanography* **87**, 106–112 (2010).

[31] Munday PL. Climate Change and Coral Reef Connectivity. *Coral Reefs* **28**, 379–395 (2009).

[32] Bellman R. The Theory of Dynamic Programming. *Bulletin of the American Mathematical Society* **60**, 503–515 (1954).

[33] Doty MS, Oguri M. The Island Mass Effect. *ICES Journal of Marine Science: Journal du Conseil* **22**, 33–37 (1956).

[34] Luo JY, Grassian B, Tang D, Irisson JO, Greer AT, Guigand CM, McClatchie S, Cowen RK. Environmental Drivers of the Fine-Scale Distribution of a Gelatinous Zooplankton Community across a Mesoscale Front. *Marine Ecology Progress Series* **510**, 129–149 (2014).

[35] Brun P, Payne MR, Kiørboe T. Trait Biogeography of Marine Copepods–an Analysis across Scales. *Ecology letters* **19**, 1403–1413 (2016).

[36] Fiksen Ø, Giske J. Vertical Distribution and Population Dynamics of Copepods by Dynamic Optimization. *ICES Journal of Marine Science: Journal du Conseil* **52**, 483–503 (1995).

[37] Banse K. On the Vertical Distribution of Zooplankton in the Sea. *Progress in oceanography* **2**, 53–125 (1964).

[38] Wiebe PH, Burt KH, Boyd SH, Morton AW. A Multiple Opening/Closing Net and Environmental Sensing System for Sampling Zooplankton. *Journal of Marine Research* **34**, 313–326 (1976).

[39] Hirata T, Aiken J, Hardman-Mountford N, Smyth TJ, Barlow RG. An Absorption Model to Determine Phytoplankton Size Classes from Satellite Ocean Colour. *Remote Sensing of Environment* **112**, 3153–3159 (2008).

[40] Hirzel AH, Lay GL. Habitat Suitability Modelling and Niche Theory. *Journal of Applied Ecology* **45**, 1372–1381 (2008).

[41] Cuzin-Roudy J, Irisson JO, Penot F, Kawaguchi S, Vallet C. Chapter 6.9 Southern Ocean Euphausiids. In *Biogeographic Atlas of the Southern Ocean*, De Broyer C, Koubbi P, Griffiths HJ, Raymond B, Udekem d'Acoz C, et al, eds., Scientific Committee on Antarctic Research, Cambridge, 309–320 (2014).

[42] Duhamel G, Hulley PA, Causse R, Koubbi P, Vacchi M, Pruvost P, Vigetta S, Irisson JO, Mormede S, Belchier M, Dettai A, Detrich HW, Gutt J, Jones CD, Kock KH, Lopez Abellan LJ, Van de Putte A. Chapter 7 Biogeographic Patterns of Fish. In *Biogeographic Atlas of the Southern Ocean*, De Broyer C, Koubbi P, Griffiths HJ, Raymond B, Udekem d'Acoz C, et al, eds., Scientific Committee on Antarctic Research, 328–362 (2014).

[43] Reygondeau G, Irisson JO, Ayata SD, Gasparini S, Benedetti F, Albouy C, Hattab T, Guieu C, Koubbi P. Definition of the Mediterranean Eco-Regions and Maps of Potential Pressures in These Eco-Regions. Delivrable 1.6, PERSEUS (2014).

[44] Benedetti F, Guilhaumon F, Adloff F, Ayata SD. Investigating Uncertainties in Zooplankton Composition Shifts under Climate Change Scenarios in the Mediterranean Sea. *Ecography* **41**, 345–360 (2018).

[45] Longhurst AR. *Ecological Geography of the Sea*. Elsevier (2010).

[46] Reygondeau G, Guieu C, Benedetti F, Irisson JO, Ayata SD, Gasparini S, Koubbi P. Biogeochemical Regions of the Mediterranean Sea: An Objective Multidimensional and Multivariate Environmental Approach. *Progress in Oceanography* **151**, 138–148 (2017).

[47] Oliver MJ, Glenn S, Kohut JT, Irwin AJ, Schofield OM, Moline MA, Bissett WP. Bioinformatic Approaches for Objective Detection of Water Masses on Continental Shelves. *Journal of Geophysical Research: Oceans* **109** (2004).

[48] Yeung KY, Haynor DR, Ruzzo WL. Validating Clustering for Gene Expression Data. *Bioinformatics* **17**, 309–318 (2001).

[49] Josse J, Pagès J, Husson F. Multiple Imputation in Principal Component Analysis. *Advances in data analysis and classification* **5**, 231–246 (2011).

[50] Cleveland RB, Cleveland WS, McRae JE, Terpenning I. STL: A Seasonal-Trend Decomposition Procedure Based on Loess. *Journal of Official Statistics* **6**, 3–73 (1990).

[51] Steltzer H, Post E. Seasons and Life Cycles. *Science* **324**, 886–887 (2009).

[52] Giorgi F. Climate Change Hot-Spots. *Geophysical research letters* **33** (2006).

[53] Grosjean P, Picheral M, Warembourg C, Gorsky G. Enumeration, Measurement, and Identification of Net Zooplankton Samples Using the ZOOSCAN Digital Imaging System. *ICES Journal of Marine Science: Journal du Conseil* **61**, 518–525 (2004).

[54] Villéger S, Mason NW, Mouillot D. New Multidimensional Functional Diversity Indices for a Multifaceted Framework in Functional Ecology. *Ecology* **89**, 2290–2301 (2008).

[55] Legendre P, Legendre L. *Numerical Ecology*. Elsevier (2012).

[56] Rosenzweig ML, Abramsky Z. How Are Diversity and Productivity Related? In *Species Diversity in Ecological Communities*, University of Chicago Press, 52–65 (1993).

[57] McIntyre A. *Life in the World's Oceans: Diversity, Distribution, and Abundance*. John Wiley & Sons (2010).

[58] Lecaillon G. The "CARE"(Collect by Artificial Reef Eco-Friendly) System as a Method of Producing Farmed Marine Animals for the Aquarium Market: An Alternative Solution to Collection in the Wild. *SPC Live Reef Fish Information Bulletin* **12**, 17–20 (2004).

[59] Faillettaz R, Voué R, Crec'hriou R, Garsi L, Lecaillon G, Agostini S, Lenfant P, Irisson J. Spatio-Temporal Patterns of Larval Fish Settlement in the Northwestern Mediterranean Sea. *Mar. Ecol. Prog. Ser.* **Advanced View**, 1–21 (2020).

[60] Victor BC. Recruitment and Population Dynamics of a Coral Reef Fish. *Science* **219**, 419–420 (1983).

[61] Koenker R. *Quantile Regression*, vol. 38. Cambridge University Press (2005).

[62] Chen K, Ying Z, Zhang H, Zhao L. Analysis of Least Absolute Deviation. *Biometrika* **95**, 107–122 (2008).

[63] Linton O, Whang YJ. The Quantilogram: With an Application to Evaluating Directional Predictability. *Journal of Econometrics* **141**, 250–282 (2007).

[64] Giraldo C, Mayzaud P, Tavernier E, Irisson JO, Penot F, Becciu J, Chartier A, Boutoute M, Koubbi P. Lipid Components as a Measure of Nutritional Condition in Fish Larvae (\textitPleuragramma Antarcticum) in East Antarctica. *Marine Biology* **160**, 877–887 (2013).

[65] Talec A, Philistin M, Ferey F, Walenta G, Irisson JO, Bernard O, Sciandra A. Effect of Gaseous Cement Industry Effluents on Four Species of Microalgae. *Bioresource Technology* **143**, 353–359 (2013).

[66] Howes EL, Stemmann L, Assailly C, Irisson JO, Dima M, Bijma J, Gattuso JP. Pteropod Time Series from the North Western Mediterranean (1967-2003): Impacts of pH and Climate Variability. *Marine Ecology Progress Series* **531**, 193–206 (2015).

[67] Ezzat L, Towle E, Irisson JO, Langdon C, Ferrier-Pagès C. The Relationship between Heterotrophic Feeding and Inorganic Nutrient Availability in the Scleractinian Coral T. Reniformis under a Short-Term Temperature Increase. *Limnology and Oceanography* **61**, 89–102 (2016).

[68] Jean M, Irisson JO, Gras G, Bouchand F, Simo D, Duran C, Perronne C, Mulleman D, Bernard L, Dinh A. Diagnostic Delay of Pyogenic Vertebral Osteomyelitis and Its Associated Factors. *Scandinavian Journal of Rheumatology* **46**, 64–68 (2017).

[69] Labrousse S, Vacquié-Garcia J, Heerah K, Guinet C, Sallée JB, Authier M, Picard B, Roquet F, Bailleul F, Hindell M. Winter Use of Sea Ice and Ocean Water Mass Habitat by Southern Elephant Seals: The Length and Breadth of the Mystery. *Progress in Oceanography* **137**, 52–68 (2015).

[70] Labrousse S, Sallée JB, Fraser AD, Massom RA, Reid P, Sumner M, Guinet C, Harcourt R, Mcmahon C, Bailleul F. Under the Sea Ice: Exploring the Relationship between Sea Ice and the Foraging Behaviour of Southern Elephant Seals in East Antarctica. *Progress in oceanography* **156**, 17–40 (2017).

[71] Labrousse S, Sallée JB, Fraser AD, Massom RA, Reid P, Hobbs W, Guinet C, Harcourt R, McMahon C, Authier M. Variability in Sea Ice Cover and Climate Elicit Sex Specific Responses in an Antarctic Predator. *Scientific reports* **7**, 43236 (2017).

[72] Biard T, Stemmann L, Picheral M, Mayot N, Vandromme P, Hauss H, Gorsky G, Guidi L, Kiko R, Not F. *In Situ* Imaging Reveals the Biomass of Giant Protists in the Global Ocean. *Nature* **532**, 504–507 (2016).

[73] Dumollard R, Minc N, Salez G, Aicha SB, Bekkouche F, Hebras C, Besnardeau L, McDougall A. The Invariant Cleavage Pattern Displayed by Ascidian Embryos Depends on Spindle Positioning along the Cell's Longest Axis in the Apical Plane and Relies on Asynchronous Cell Divisions. *eLife* (2017).

[74] Gattuso JP, Magnan AK, Bopp L, Cheung WWL, Duarte CM, Hinkel J, Mcleod E, Micheli F, Oschlies A, Williamson P, Billé R, Chalastani V, Gates RD, Irisson JO, Middelburg JJ, Poertner HO, Rau G. Ocean Solutions to Address Climate Change and Its Effects on Marine Ecosystems. *Frontiers in Marine Science* **5**, 337 (2018).

[75] Sukhovich A, Irisson JO, Simons FJ, Ogé A, Hello Y, Deschamps A, Nolet G. Automatic Discrimination of Underwater Acoustic Signals Generated by Teleseismic P-Waves: A Probabilistic Approach. *Geophysical Research Letters* **38** (2011).

[76] Sukhovich A, Bonnieux S, Hello Y, Irisson JO, Simons FJ, Nolet G. Seismic Monitoring in the Oceans by Autonomous Floats. *Nature Communications* **6**, 8027 (2015).

[77] Wickham H. *Ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York Inc (2009).

[78] De Broyer C, Koubbi P, Griffiths H, Grant SA. *Biogeographic Atlas of the Southern Ocean*. Scientific Committee on Antarctic Research Cambridge (2014).

[79] Beaugrand G, Lenoir S, Ibañez F, Manté C. A New Model to Assess the Probability of Occurrence of a Species, Based on Presence-Only Data. *Marine Ecology Progress Series* **424**, 175–190 (2011).

[80] Behrenfeld MJ, Randerson JT, McClain CR, Feldman GC, Los SO, Tucker CJ, Falkowski PG, Field CB, Frouin R, Esaias WE. Biospheric Primary Production during an ENSO Transition. *Science* **291**, 2594–2597 (2001).

[81] Volk T, Hoffert MI. Ocean Carbon Pumps: Analysis of Relative Strengths and Efficiencies in Ocean-Driven Atmospheric CO2 Changes. *The carbon cycle and atmospheric CO2: natural variations Archean to present* **32**, 99–110 (1985).

[82] Frederiksen M, Edwards M, Richardson AJ, Halliday NC, Wanless S. From Plankton to Top Predators: Bottom-up Control of a Marine Food Web across Four Trophic Levels. *Journal of Animal Ecology* **75**, 1259–1268 (2006).

[83] Thiel M, Macaya EC, Acuna E, Arntz WE, Bastias H, Brokordt K, Camus PA, Castilla JC, Castro LR, Cortes M. The Humboldt Current System of Northern and Central Chile: Oceanographic Processes, Ecological Interactions and Socioeconomic Feedback. *Oceanography and marine biology: an annual review* (2007).

[84] Hays GC, Richardson AJ, Robinson C. Climate Change and Marine Plankton. *Trends in ecology & evolution* **20**, 337–344 (2005).

[85] Beaugrand G. Monitoring Pelagic Ecosystems Using Plankton Indicators. *ICES Journal of Marine Science* **62**, 333–338 (2005).

[86] De Vargas C, Audic S, Henry N, Decelle J, Mahé F, Logares R, Lara E, Berney C, Le Bescot N, Probert I. Eukaryotic Plankton Diversity in the Sunlit Ocean. *Science* **348**, 1261605 (2015).

[87] Guidi L, Chaffron S, Bittner L, Eveillard D, Larhlimi A, Roux S, Darzi Y, Audic S, Berline L, Brum JR, Coelho LP, Espinoza JCI, Malviya S, Sunagawa S, Dimier C, Kandels-Lewis S, Picheral M, Poulain J, Searson S, Tara Oceans Consortium Coordinators, Stemmann L, Not F, Hingamp P, Speich S, Follows M, Karp-Boss L, Boss E, Ogata H, Pesant S, Weissenbach J, Wincker P, Acinas SG, Bork P, de Vargas C, Iudicone D, Sullivan MB, Raes J, Karsenti E, Bowler C, Gorsky G. Plankton Networks Driving Carbon Export in the Oligotrophic Ocean. *Nature* **532**, 465–470 (2016).

[88] Russakovsky O, Deng J, Su H, Krause J, Satheesh S, Ma S, Huang Z, Karpathy A, Khosla A, Bernstein M. Imagenet Large Scale Visual Recognition Challenge. *International journal of computer vision* **115**, 211–252 (2015).

[89] Krizhevsky A, Sutskever I, Hinton GE. Imagenet Classification with Deep Convolutional Neural Networks. In *Advances in Neural Information Processing Systems* (2012), 1097–1105.

[90] LeCun Y, Bengio Y, Hinton G. Deep Learning. *Nature* **521**, 436 (2015).

[91] Elith J, Graham CH, Anderson RP, Dudík M, Ferrier S, Guisan A, Hijmans RJ, Huettmann F, Leathwick JR, Lehmann A, Li J, Lohmann LG, Loiselle BA, Manion G, Moritz C, Nakamura M, Nakazawa Y, Overton JM, Peterson AT, Phillips SJ, Richardson K, Scachetti-Pereira R, Schapire RE, Soberón J, Williams S, Wisz MS, Zimmermann NE. Novel Methods Improve Prediction of Species' Distributions from Occurrence Data. *Ecography* **29**, 129–151 (2006).

[92] Uhlmann D, Schlimpeet O, Uhlmann W. Automated Phytoplankton Analysis by a Pattern Recognition Method. *Internationale Revue der gesamten Hydrobiologie und Hydrographie* **63**, 575–583 (1978).

[93] Graham B. Spatially-Sparse Convolutional Neural Networks. *arXiv:1409.6070 [cs]* (2014).

[94] Graham B. Fractional Max-Pooling. *arXiv:1412.6071 [cs]* (2014).

[95] Luo JY, Irisson JO, Graham B, Guigand C, Sarafraz A, Mader C, Cowen RK. Automatic Plankton Image Analysis Using Convolutional Neural Networks. *Limnology and Oceanography: Methods* **16**, 814–827 (2018).

[96] Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z. Rethinking the Inception Architecture for Computer Vision. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2016), 2818–2826.

[97] Malde K, Kim H. Beyond Image Classification: Zooplankton Identification with Deep Vector Space Embeddings. *arXiv:1909.11380 [cs]* (2019).

[98] Elineau A, Desnos C, Jalabert L, Olivier M, Romagnan JB, Brandao M, Lombard F, Llopis N, Courboulès J, Caray-Counil L, Serranito B, Irisson JO, Picheral M, Gorsky G, Stemmann L. ZooScanNet: Plankton Images Captured with the ZooScan (2018).

[99] Bakkum P, Skadron K. Accelerating SQL Database Operations on a GPU with CUDA. In *Proceedings of the 3rd Workshop on General-Purpose Computation on Graphics Processing Units*. ACM (2010), 94–103.

[100] Zhu X, Bain M. B-CNN: Branch Convolutional Neural Network for Hierarchical Classification. *arXiv preprint arXiv:1709.09890* (2017).

[101] Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation. *arXiv:1505.04597 [cs]* (2015).

[102] Badrinarayanan V, Kendall A, Cipolla R. SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **39**, 2481–2495 (2017).

[103] He K, Gkioxari G, Dollár P, Girshick R. Mask R-CNN. *arXiv:1703.06870 [cs]* (2018).

[104] Ibarbalz FM, Henry N, Brandão MC, Martini S, Busseni G, Byrne H, Coelho LP, Endo H, Gasol JM, Gregory AC. Global Trends in Marine Plankton Diversity across Kingdoms of Life. *Cell* **179**, 1084–1097 (2019).

[105] Lévy M, Franks PJS, Smith KS. The Role of Submesoscale Currents in Structuring Marine Ecosystems. *Nat Commun* **9**, 1–16 (2018).

[106] Mahadevan A. The Impact of Submesoscale Physics on Primary Productivity of Plankton. *Annu. Rev. Mar. Sci.* **8**, 161–184 (2016).

[107] Ohman MD, Powell JR, Picheral M, Jensen DW. Mesozooplankton and Particulate Matter Responses to a Deep-Water Frontal System in the Southern California Current System. *J Plankton Res* **34**, 815–827 (2012).

[108] Greer AT, Cowen RK, Guigand CM, Hare JA. Fine-Scale Planktonic Habitat Partitioning at a Shelf-Slope Front Revealed by a High-Resolution Imaging System. *Journal of Marine Systems* **142**, 111–125 (2015).

[109] Schmid MS, Cowen RK, Robinson K, Luo JY, Briseño-Avena C, Sponaugle S. Prey and Predator Overlap at the Edge of a Mesoscale Eddy: Fine-Scale, in-Situ Distributions to Inform Our Understanding of Oceanographic Processes. *Scientific Reports* **10**, 1–16 (2020).

[110] Greer AT, Cowen RK, Guigand CM, McManus MA, Sevadjian JC, Timmerman AHV. Relationships between Phytoplankton Thin Layers and the Fine-Scale Vertical Distributions of Two Trophic Levels of Zooplankton. *J Plankton Res* **35**, 939–956 (2013).

[111] Bertrand A, Grados D, Colas F, Bertrand S, Capet X, Chaigneau A, Vargas G, Mousseigne A, Fablet R. Broad Impacts of Fine-Scale Dynamics on Seascape Structure from Zooplankton to Seabirds. *Nat Commun* **5**, 1–9 (2014).

[112] Snyder S, Franks PJS, Talley LD, Xu Y, Kohin S. Crossing the Line: Tunas Actively Exploit Submesoscale Fronts to Enhance Foraging Success. *Limnology and Oceanography Letters* **2**, 187–194 (2017).

[113] Lombard F, Boss E, Waite AM, Vogt M, Uitz J, Stemmann L, Sosik HM, Schulz J, Romagnan JB, Picheral M, Pearlman J, Ohman MD, Niehoff B, Möller KO, Miloslavich P, Lara-Lpez A, Kudela R, Lopes RM, Kiko R, Karp-Boss L, Jaffe JS, Iversen MH, Irisson JO, Fennel K, Hauss H, Guidi L, Gorsky G, Giering SLC, Gaube P, Gallager S, Dubelaar G, Cowen RK, Carlotti F, Briseño-Avena C, Berline L, Benoit-Bird K, Bax N, Batten S, Ayata SD, Artigas LF, Appeltans W. Globally Consistent Quantitative Observations of Planktonic Ecosystems. *Frontiers in Marine Science* **6** (2019).

[114] Cowen RK, Guigand CM. In Situ Ichthyoplankton Imaging System (ISIIS): System Design and Preliminary Results. *Limnology and Oceanography: Methods* **6**, 126–132 (2008).

[115] Ohman MD, Davis RE, Sherman JT, Grindley KR, Whitmore BM, Nickels CF, Ellen JS. Zooglider: An Autonomous Vehicle for Optical and Acoustic Sensing of Zooplankton. *Limnology and Oceanography: Methods* **17**, 69–86 (2019).

[116] Picheral M, Guidi L, Stemmann L, Karl DM, Iddaoud G, Gorsky G. The Underwater Vision Profiler 5: An Advanced Instrument for High Spatial Resolution Studies of Particle Size Spectra and Zooplankton. *Limnology and Oceanography: Methods* **8**, 462–473 (2010).

[117] Faillettaz R, Picheral M, Luo JY, Guigand C, Cowen RK, Irisson JO. Imperfect Automatic Image Classification Successfully Describes Plankton Distribution Patterns. *Methods in Oceanography* **15-16**, 60–77 (2016).

[118] Cailleton C, Desnos C, Jalabert L, Elineau A, Stemmann L, Ayata SD, Irisson JO. Morphological Diversity Increases with Oligotrophy along a Zooplankton Time Series. In *SFEcologie 2018*. Rennes (2018).

[119] Hu MK. Visual Pattern Recognition by Moment Invariants. *IRE transactions on information theory* **8**, 179–187 (1962).

[120] Jaffe JS, Franks PJS, Roberts PLD, Mirza D, Schurgers C, Kastner R, Boch A. A Swarm of Autonomous Miniature Underwater Robot Drifters for Exploring Submesoscale Ocean Dynamics. *Nat Commun* **8**, 1–8 (2017).

[121] Fu LL, Chelton DB, Le Traon PY, Morrow R. Eddy Dynamics from Satellite Altimetry. *Oceanography* **23**, 14–25 (2010).

[122] Alvain S, Moulin C, Dandonneau Y, Loisel H. Seasonal Distribution and Succession of Dominant Phytoplankton Groups in the Global Ocean: A Satellite View. *Global Biogeochemical Cycles* **22** (2008).

[123] Basedow SL, McKee D, Lefering I, Gislason A, Daase M, Trudnowska E, Egeland ES, Choquet M, Falk-Petersen S. Remote Sensing of Zooplankton Swarms. *Scientific reports* **9**, 686 (2019).

[124] Jónasdóttir SH, Visser AW, Richardson K, Heath MR. Seasonal Copepod Lipid Pump Promotes Carbon Sequestration in the Deep North Atlantic. *Proceedings of the National Academy of Sciences* **112**, 12122–12126 (2015).

[125] Jean N, Burke M, Xie M, Davis WM, Lobell DB, Ermon S. Combining Satellite Imagery and Machine Learning to Predict Poverty. *Science* **353**, 790–794 (2016).

[126] Aurell E, Boffetta G, Crisanti A, Paladin G, Vulpiani A. Predictability in the Large: An Extension of the Concept of Lyapunov Exponent. *Journal of Physics A: Mathematical and General* **30**, 1 (1997).

[127] d'Ovidio F, De Monte S, Alvain S, Dandonneau Y, Lévy M. Fluid Dynamical Niches of Phytoplankton Types. *Proceedings of the National Academy of Sciences* **107**, 18366–18370 (2010).

[128] Huntley ME, Gonzalez A, Zhu Y, Zhou M, Irigoien X. Zooplankton Dynamics in a Mesoscale Eddy-Jet System off California. *Marine Ecology Progress Series* **201**, 165–178 (2000).

Document typeset by LaTeX — 13th March 2020

Résumé

Ma carrière scientifique a commencé par l'étude du comportement des larves de poissons. J'ai prouvé qu'elles nagent à des vitesses non-négligeables, s'orientent, notamment en fonction de l'azimut solaire, et s'orientent mieux en groupe que seules. Quand ces capacités comportementales sont introduites dans des modèles Lagrangiens de dispersion, elles influencent les patrons spatiaux de recrutement, réduisant souvent les distances de dispersion.

Je me suis également intéressé à la distribution tridimensionnelle de ces larves et du plancton associé. Ces études ont montré que, dans les structures à mésoéchelle, une très haute résolution spatio-temporelle est nécessaire pour parvenir à détecter des patrons clairs dans la distribution des organismes ou les associations entre taxa.

À plus large échelle, je me suis ensuite intéressé à la biogéographie d'espèces pélagiques. Dans un atlas de l'océan Austral, la modélisation de la niche écologique d'espèces de poissons et de krill a souligné la forte influence des fronts circum-antarctiques. Les travaux de régionalisation de la Méditerranée ont montré que des communautés spécifiques d'organismes pélagiques exploitent la zone de gradient entre les bassins Est et Ouest. La comparaisons entre diverses régionalisations de cette région a permis d'identifier des zones, de stabilité ou de fort gradient, cohérentes quelles que soit les caractéristiques des masses d'eau considérées.

Enfin, j'ai contribué à l'effort d'observation à long terme mené à Villefranche-sur-Mer, notamment en mettant en place un système de centralisation et visualisation des données récoltées. L'exploitation de ces données a confirmé que la Méditerranée est une zone où le changement climatique est prononcé, que l'oligotrophisation associée résulte en une diversification des organismes zooplanctoniques et que le recrutement des poissons côtiers est cohérent spatialement mais extrêmement sporadique temporellement.

Le point commun entre ces activités apparemment disparates est l'utilisation d'outils numériques pour répondre à des questions écologiques. Mon projet de recherche est également focalisé sur des approches "computationnelles", telles que l'apprentissage machine, qui combine de nombreux calculs unitaires pour reconnaitre des structures dans de grandes masses de données. Ces données sont de maintenant disponibles pour l'étude du plancton, notamment grâce à de nouveaux instruments d'imagerie. Je commencerai par poursuivre mes efforts actuels vers l'automatisation de l'identification d'organismes planctoniques sur des images. En utilisant ces nouveaux outils instrumentaux et logiciels, il deviendra possible de décrire la distribution spatio-temporelle du plancton à la même, haute, résolution que son environnement biogéochimique. Cela permettra d'étudier leurs intéractions dans les structures à submésoéchelle, qui sont primordiales pour la dynamique de ces organismes à plus grande échelle. Enfin, à cette grande échelle, les satellites permettent d'avoir une vision synoptique de la couche de surface des océans. Je propose d'utiliser les corrélations spatio-temporelles locales dans données qu'ils récoltent pour décrire, et prédire à l'échelle globale, la structure des communautés de zooplancton.