

# Computation of functional diversity from long-term monitoring time series

Der Fakultät für Mathematik und Naturwissenschaften  
der Carl von Ossietzky Universität Oldenburg  
zur Erlangung des Grades und Titels  
Doktor der Naturwissenschaften (Dr. rer. nat.)  
angenommene Dissertation

von

**Pedro Manuel Carrasco De La Cruz**

geboren am 27.02.1988 in Lambayeque, Peru

**Gutachter:** Prof. Dr. Thilo Gross  
**Gutachter:** Prof. Dr. Helmut Hillebrand  
**Tag der Disputation:** 28.11.2024

Sea-purple harmony  
While Kogi hid secrets into seashells  
And even the ocean laughed  
Beneath that celestial canopy  
Cuz it started out so nice - *Sixto Rodriguez*.

... el mundo por suerte es un sinfin de incomprensiones.  
Por eso hacemos libros,  
por la ilusion siempre fallida  
de alguna vez entender algo - *Martin Caparros, Ñamerica*.

# Acknowledgements

Throughout the past four years I received the help of many people. People who made my time as a PhD student enriching and beautiful. Sure it was a challenge. Everything which demands growing and improvement from us, is a challenge. That is why, I would like to deeply thank the people that in one way or another, either scientifically or emotionally, provided me with the necessary support to complete this work. My more sincere gratitude to:

Thilo Gross, my PhD supervisor, because  
your scientific support, friendly encouragement, and above all  
your trust on me, made this work possible.

Josie Antonucci, Jana Massing, and Alex Ryabov, my colleagues and scientific collaborators  
because  
your knowledge, expertise, and support in all those little,  
but laborious tasks made all the difference.

Jana, Ole, Laura, Rana, and Melanie; my friends and colleagues, because  
we took this ship together. You made the trip so much fun.

the HIFMB scientific and administrative staff, because  
a workplace is a second home,  
and only nice people make a workplace special.

mis queridos Mama, Papa y hermano, porque  
creyeron en mi, aun cuando yo dudaba. Son mi corazon.  
En especial a ti Papa, contigo aprendi que lo unico importante es ser una buena persona.

my Helgolander friends, Dieu, Antonella, Hadeer, Akoko, Adreeja,  
Jeffrey, Andrea, Manfred and Sharloth, because  
distance and time means little when we are together,  
we are the power of ten, remember?

Leonie, my partner and best friend, because  
your support, love and all the laughs we shared, also made this possible,  
everything is better when we are together.

the people who recommended me a book, a podcast, a movie, or any other source of ideas  
that sparkled a bit of creativity on me  
whoever helped me solve a problem  
everyone who just smiled, or made me smile during a difficult time, because  
why not to thank to you too?

# Abstract

Biodiversity is a concept which has become transversal in modern society. Although its core idea is linked to ecology, its influence goes beyond natural sciences, reaching disciplines as distinct as economics, politics, law, and conservation biology. Perhaps it was due to the realization of its accelerated loss, or the awareness of our responsibility on it, that increasing international efforts have been advocated to ‘improve the status of biodiversity’. Yet unclear or incomplete reports about biodiversity changes and trends remain, a direct consequence of the limited availability of long-term biodiversity indicators. Furthermore, the focus of what biodiversity itself means and how it should be measured in ecosystems, has shifted from just considering the amount of species on ecosystems, towards considering the influence that those species and their interactions have on the functionality of ecosystems. An approach known as functional biodiversity.

Ideally, we should measure functional biodiversity directly in ecosystems, seeking to establish well-designed programmes to measure its change in key ecosystems or species groups. But such programmes have high operation costs, their logistics are complicated, and the expertise needed to maintain them is probably not available where is needed, thus efforts will remain local or regional, or inoperative. Currently there is an effort to use existing information, often collected for other purposes, to gain insights on how the state of biodiversity is changing. Ecological long-term datasets of species, which gathered data on biomass, abundance or just presence-absence, represent a unique source for assessing biodiversity, given the appropriate tools and methodology. Diffusion maps offer an appropriate approach.

Diffusion maps, a method from networks science, can coarse-grain the ecological communities in terms of their structuring traits, without dismissing the species interactions. Therefore, high-dimensional ecological datasets are turned to a lower, more relevant dimensional space, in which species can be located and assessed according to their similarity or dissimilarity. Ultimately, the functional diversity can be computed from the pairwise distances in this lower dimensional space, for each sample in the dataset. The result is a functional diversity time series of the monitored ecosystem, and the ecological community. However, the application of this approach to ecology is quite recent, having yet multiple knowledge gaps to be filled.

In this thesis, I work towards expanding the applications of diffusion maps, as well as on seeking solutions for the limitations imposed by data availability. Firstly, I apply the diffusion map approach to a long-term abundance dataset of demersal fish in the North Sea obtained from the ICES coordinated, International Bottom Trawl Survey (IBTS). To date, diffusion maps were mainly applied to unicellular organisms. Therefore, this is one of the first application of this approach to biological communities of more complex organisms. Subsequently, I seek to address a limitation that diffusion maps have for its application to biodiversity assessment, the scarcity of long-term datasets. I propose an approach to aggregate datasets from multiple origins, which allows not just to obtain a larger, more robust dataset, but also to expand spatially the assessment of functional diversity. Lastly, I seek to scale-up the assessment of functional diversity from a local to a regional scale by means of dataset aggregation. The scenario is the North Sea and the protagonists are phytoplankton abundance datasets from different monitoring programmes conducted and maintained by research institutions located at its shorelines.

# Zusammenfassung

Biodiversität ist ein Konzept, welches in der modernen Gesellschaft eine übergreifende Bedeutung erlangt hat. Obwohl der Kerngedanke der Biodiversität auf der Ökologie basiert, geht ihr Einfluss über die Naturwissenschaften hinaus und erstreckt sich auf unterschiedliche Disziplinen wie Wirtschaft, Politik, Recht und Naturschutzbiologie. Die Erkenntnis des beschleunigten Biodiversitätsverlusts, sowie das wachsende Verantwortungsbewusstsein, führten möglicherweise dazu, dass verstärkte internationale Bemühungen zur ‘Verbesserung des Zustands der biologischen Vielfalt’ befürwortet wurden. Dennoch bestehen weiterhin unklare oder unvollständige Berichte zu Veränderungen und Trends in der Biodiversität, was aus begrenzter Verfügbarkeit von langfristigen Biodiversitätsindikatoren resultiert. Zudem hat sich der Fokus, was Biodiversität bedeutet und wie es in einem Ökosystem gemessen werden soll, verändert: von der bloßen Betrachtung der Artenvielfalt hin zum Einfluss der Arten und deren Interaktion untereinander auf die Funktionsweise des Ökosystems. Ein Ansatz, der als funktionelle Biodiversität bekannt ist.

Idealerweise sollten wir funktionale Biodiversität direkt im Ökosystem messen und versuchen, gut konzipierte Programme zu entwickeln, die Veränderungen in wichtigen Ökosystemen oder Artengruppen messen können. Allerdings sind die Kosten für solche Programme oftmals sehr hoch, sie verfügen über eine komplizierte Logistik und die Expertise, die nötig ist, um solche Programme zu warten, ist oft nicht dort verfügbar, wo sie benötigt wird. Daher bleiben Bemühungen entweder nur lokal oder regional, oder wirkungslos. Derzeit gibt es Bemühungen, bestehende Informationen, die häufig für andere Zwecke gesammelt werden, zu nutzen, um Erkenntnisse zu dem aktuellen Zustand der Biodiversität zu gewinnen. Ökologische Langzeitdatensätze von Arten, die Daten zu Biomasse, Abundanz oder bloßem Fehlen zusammentragen, stellen mit den richtigen Instrumenten und Methoden eine gute Möglichkeit zur Bewertung der Biodiversität dar. Sogenannte ‘diffusion maps’ bieten einen geeigneten Ansatz.

Diffusion Maps sind eine Methode aus der Netzwerkwissenschaft, welche ökologische Gemeinschaften grob hinsichtlich ihrer strukturellen Eigenschaften unterscheiden können, ohne die Arteninteraktion zu vernachlässigen. Dabei werden hochdimensionale ökologische Datensätze einem niedrigeren, relevanteren Raum zugeordnet, in dem Arten lokalisiert und anhand ihrer Ähnlichkeit oder Unähnlichkeit bewertet werden können. Letztendlich kann die funktionale Diversität aus den paarweisen Abständen im niedrigdimensionalen Raum für jede Stichprobe im Datensatz berechnet werden. Das Ergebnis ist eine Zeitreihe der funktionalen Diversität des überwachten Ökosystems und der ökologischen Gemeinschaft. Allerdings ist die Anwendung dieses Ansatzes auf die Ökologie noch recht neu und weist noch zahlreiche Wissenslücken auf, die geschlossen werden müssen.

In dieser Promotionsarbeit arbeite ich daran, sowohl die Anwendungsmöglichkeiten von Diffusionskarten zu erweitern als auch Lösungen für die Einschränkungen durch limitierte Datenverfügbarkeit zu finden. Zunächst wende ich den Ansatz der Diffusionskarten auf einen Datensatz an, der die Abundanz von Grundfischen in der Nordsee dokumentiert. Dieser Datensatz stammt aus der vom ICES koordinierten International Bottom Trawl Survey (IBTS). Bisher wurden Diffusionskarten hauptsächlich auf einzellige Organismen angewendet. Daher ist dies eine der ersten Anwendungen von dieser Methode auf biologische Gemeinschaften komplexerer Organismen. Anschließend möchte ich auf eine Einschränkung eingehen, die Diffusionskarten für ihre Anwendung bei der Bewertung der biologischen Vielfalt haben: die limitierte Verfügbarkeit von Langzeitdatensätzen. Ich schlage einen Ansatz vor, der Datensätzen aus mehreren

Quellen vereint, um nicht nur einen größeren und robusteren Datensatz zu erhalten, sondern auch die Bewertung der funktionalen Biodiversität räumlich zu erweitern. Schließlich versuche ich, die Bewertung der funktionalen Vielfalt mittels Datensatzaggregation von einer lokalen auf eine regionale Ebene anzuwenden. Das Untersuchungsgebiet ist die Nordsee und die Protagonisten sind Datensätze zur Abundanz von Phytoplankton aus verschiedenen Monitoringprogrammen, die von Forschungseinrichtungen entlang der Nordseeküste durchgeführt und gepflegt werden.

# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>Concepts and Methodology</b>	<b>5</b>
2.1	Complexity in Ecosystems . . . . .	6
2.2	Biodiversity: functional or not? . . . . .	8
2.2.1	The Biodiversity-Ecosystem function relation . . . . .	9
2.2.2	How can functional diversity be measured? . . . . .	10
2.3	Networks . . . . .	13
2.3.1	Network representation . . . . .	14
2.4	Diffusion maps . . . . .	15
2.4.1	Diffusion maps process method . . . . .	17
2.5	Calculation of functional diversity . . . . .	21
<b>3</b>	<b>Diffusion mapping a long-term time series</b>	<b>23</b>
3.1	Dynamics of fish communities . . . . .	24
3.2	Reconstructing the functional trait space . . . . .	24
3.2.1	Long-term data and area of study . . . . .	25
3.2.2	Diffusion mapping the functional trait space . . . . .	25
3.2.3	Identifying functional traits . . . . .	27
3.2.4	Important functional traits in the North Sea fish dataset . . . . .	29
3.3	Functional Diversity . . . . .	42
3.4	Discussion . . . . .	46
<b>4</b>	<b>Aggregation of time series</b>	<b>50</b>



4.1	Issues of ecological time series . . . . .	51
4.2	Failure of simple aggregation . . . . .	52
4.3	Successful aggregation of phytoplankton datasets . . . . .	54
4.4	Functional diversity status of southern North Sea and Wadden Sea . . . . .	57
4.5	Summary . . . . .	61
<b>5</b>	<b>Scaling-up the aggregation of time series</b>	<b>63</b>
5.1	Phytoplankton monitoring programs in the North Sea . . . . .	64
5.2	Reconstructing the functional i-trait space . . . . .	65
5.3	Functional diversity of North Sea and Wadden Sea phytoplankton community .	68
5.4	Discussion . . . . .	70
<b>6</b>	<b>Conclusions</b>	<b>76</b>

# List of Figures

2.1	Levels of ecological organization represented by Russian Nesting Dolls . . . . .	7
2.2	Example of undirected network . . . . .	13
2.3	Example of a data manifold . . . . .	17
2.4	Paths along the low-dimensional geometric structure of the dataset . . . . .	20
3.1	Ranking of diffusion maps variables in four scenarios . . . . .	26
3.2	Fish growth chart representing the size at maturity . . . . .	28
3.3	Diffusion map of the two most relevant i-traits for the dataset, color coded according to weighted average depth of sampling for each fish species . . . . .	29
3.4	Bathymetry and topography of the North Sea . . . . .	31
3.5	Map of the North Sea two-part system, displaying its key characteristics . . . . .	32
3.6	Diffusion maps which highlight six examples of fish species separated in adults (A) and juveniles (J) . . . . .	33
3.7	Diffusion map of the two most relevant i-traits for the dataset, color coded according to weighted average year of higher abundance for each fish species . . . . .	34
3.8	Diffusion maps which highlight the species with highest abundance during two periods, 1980-1999 and 2010-2021 . . . . .	35
3.9	Box-plot abundance time series of ‘Boreal fish’ which disappeared from areas of the northern and central North Sea . . . . .	36
3.10	Box-plot abundance time series of ‘Boreal fish’ with decreasing abundance trends in areas of the northern and central North Sea . . . . .	37
3.11	Box-plot abundance time series of ‘Boreal fish’ with decreasing abundance trends in areas of the southern North Sea, the Shagerrak, and the Kattegat . . . . .	38
3.12	Box-plot abundance time series of ‘Lusitanian fish’ with increasing abundance trends in areas of the northern and central North Sea . . . . .	39
3.13	Box-plot abundance time series of ‘Lusitanian fish’ with increasing abundance trends in areas of the southern North Sea, the Shagerrak, and the Kattegat . . . . .	40
3.14	Box-plot abundance time series of ‘Lusitanian fish’ which appeared in areas of the northern North Sea and the Irish Sea . . . . .	41

3.15	North Sea functional diversity variation according to area and season . . . . .	42
3.16	North Sea functional diversity variation according to area and year of sampling .	43
3.17	Box-plot abundance time series of key species reported in Area 1 and Area 4 of North Sea . . . . .	44
3.18	Box-plot abundance time series of key species reported in Area 6 and Area 9 of North Sea . . . . .	45
3.19	North Sea Map displaying the sampling hauls colored according to yearly percentage variation of functional diversity . . . . .	46
4.1	Wadden Sea and Southern North Sea Map including the phytoplankton monitoring stations from NLWKN and Rijkswaterstaat . . . . .	53
4.2	Diffusion map of the reconstructed i-traits, and raking of variables for the German and Dutch datasets . . . . .	54
4.3	Diffusion map of the two most relevant i-traits from the aggregated monitoring dataset using the simple aggregation method . . . . .	55
4.4	Schematic of the proposed method for aggregating monitoring datasets . . . . .	55
4.5	Diffusion map of the two most relevant i-traits from the aggregated monitoring dataset using the proposed aggregation method . . . . .	56
4.6	Diffusion map of the two most relevant i-traits color coded according to salinity and DIN for each phytoplankton species . . . . .	57
4.7	Wadden Sea functional diversity variation according to station and season . . . .	58
4.8	Southern North Sea functional diversity variation according to station and season	59
4.9	Phytoplankton functional diversity on Wadden Sea monitoring stations . . . . .	60
4.10	Phytoplankton functional diversity in southern North Sea monitoring stations off the Dutch sand barrier islands . . . . .	61
5.1	Diffusion map of the two most relevant i-traits for the aggregated CEFAS and MSS dataset, and the aggregated NLWKN and Helgoland Roads dataset . . . .	66
5.2	Diffusion map of the two most relevant i-traits for four scenarios of datasets aggregation . . . . .	67
5.3	Diffusion map of the two most relevant i-traits for the aggregated NLWKN, Helgoland Roads, Flanders, CEFAS, Norway, MSS, and Rijkswaterstaat datasets	68
5.4	Phytoplankton functional diversity variation on the CEFAS and the Scottish MSS monitoring stations . . . . .	69
5.5	Phytoplankton functional diversity variation on the Wadden Sea monitoring stations . . . . .	71
5.6	Phytoplankton functional diversity variation on the Southern North Sea monitoring stations . . . . .	72

5.7 North Sea Map displaying the monitoring stations colored according to yearly percentage variation of functional diversity . . . . . 73

# List of Tables

5.1 Table with the ranking of diffusion maps first variable . . . . . 66

# Chapter 1

## Introduction

Beginnings are shadowy. So it is with the term biodiversity. Although it has, apparently, always been there, it had no name. Nothing does really. But strangely, it was not until one component of it, the only one with the capacity to name things, provided it with one that awareness of its existence emerged over disciplines as distinct as ecology, economics, politics or conservation biology [194, 211, 407, 430]. Perhaps it was the fact that naming it, also brought the realization that it was disappearing. Slowly, yes, but with an accelerating rate [184, 323]. However, no matter the amount of questions raised about it or the solutions proposed to stop its loss, naming biodiversity is never enough to understand it.

The term biodiversity is a blend of the phrase biological diversity [255]. Although the term counts with several historical origins from ecology, genetics and evolutionary biology [161], it was first mainstreamed by Edward O. Wilson in his book *The Diversity of Life* [431]. In his book Wilson draws attention to species loss and in particular to such loss caused by human activities. No a new message, certainly, but with the use of biodiversity as a synthesis term, the debate became tangible. Despite the birth and popularization of biodiversity was certainly concerned with species and the amount of them in ecosystems, the term biodiversity meant more.

Biodiversity refers to the fact that heterogeneity at different ecological levels is a fundamental property of natural systems [161]. Even E. O. Wilson, although largely concerned with species numbers, recognized that an important theme of his work is the idea that diversity cannot be captured by species numbers alone. Species are important, but it should not be assumed that we can measure biodiversity just by counting species numbers. For the functioning of biological systems depends on the kinds and combinations of organisms present [272]. In few words, biodiversity is not only the variety of organisms on our planet, but also the interdependence of all these living beings, including humans, and the influence that such interactions have on the functionality of ecosystems [391].

This new approach towards diversity is known as functional diversity. Although, there is not a standard definition of functional diversity [383], most of the proposed definitions can be summarized into two categories: the first one sees the organism as a unit, emphasizing on the quantity and properties of the organism, i.e., diversity of functional groups [401]; and the second regards the trait as a unit, and it emphasizes the range and distribution of traits, i.e., functional traits diversity [178]. Under this new paradigm ecologists acknowledge first, that biota plays an essential role in ecosystem processes [59, 410], and second, that the biodiversity of an ecosystem

plays a significant role in such processes [46]. However, accepting that biodiversity is important, and even that there is more to biodiversity than species number, opens two important questions: how do we measure functional biodiversity in an ecosystem? And, would this measure capture the effect of biodiversity in the functionality of such ecosystem?

Researchers have attempted to calculate the functional diversity in different ways [364]. An initial common method for calculating functional diversity is expressed by functional group richness [401]. It has been more effective though, to use species traits for calculating functional diversity. One of the simplest functional diversity index, uses the sum of the Euclidean distance between any two species within the assemblage based on how distinct the species traits are [421]. This index, however, depends strongly on species richness [282]. Schmera and colleagues suggested that functional diversity should be calculated by the distance matrix divided by the number of functional units [365]. Increasing knowledge of functional diversity has led the indices to include species abundance weight (such as the community weighted mean CWM) [145], functional divergence [264], functional regularity [282], multiple traits [50], intraspecific variation [364], and many more indices [224] which were developed over the past two decades.

The relationship between functional diversity and ecosystem functioning can be explained via two mechanisms. The first mechanism is ‘diversity hypothesis’ [399]. Tilman found that organisms and their functional traits in the community had an effect on ecosystem functioning by complementarity of resources use. Therefore, a highly diverse community displays less resource utilization overlap than that of a low diversity community, having also a higher proportion of resources available for consumption. Consequently, ecosystem functioning would increase [105, 401]. The other mechanism is ‘mass ratio hypothesis’ [156]. Grime found that the contribution of species to ecosystem functioning is mainly dominated by the dominant species traits, and low abundance species is relatively insensitive [145, 226]. Some traits make particular species utilize a greater proportion of resources than other species in a community, thus these species contribute disproportionately to ecosystem functioning [61]. While the diversity hypothesis reflects the trait difference between species which could maximize resource utilization strategies, the mass ratio hypothesis describes the weight of community traits. These two mechanisms are not contradictory, they are two forms of functional diversity and they show that both functional diversity and dominant species identity have significant impact on ecosystem functioning [249, 383].

A good method to measure functional diversity in an ecosystem, which considers the points previously touched, is the Rao’s quadratic entropy [342]. With this approach we can compute functional diversity from pairwise functional distances between species [349]. To obtain those distances, we need first to identify traits of the species under consideration and then, compute functional diversity from distances in trait space [295, 341, 421]. Nevertheless, a bottleneck typically occurs when the researcher must quantify the trait space of all organisms considered. Moreover, the decision of which traits are relevant functional characteristics is made based on the researcher’s experience and it is dependent on the group of species and functions under consideration. Some traits may be difficult to measure, and their values may be dependent on environmental conditions [395], hence, being context dependent. All these constraints mean that trait-based functional diversity quantification remains constrained to comparatively small groups of well-studied organisms and suffers from limited data availability [68, 359].

Nevertheless the difficulties to measure traits, a more feasible option would be to infer those traits from data [285]. The former are o-traits, physiological characteristics directly identified from observation, the latter are i-traits, approximations to the fundamental niche axis which

species occupy in an ecosystem. Trait reconstruction builds on the possibility to identify those traits which make species fit to their environment by monitoring where they are distributed over time, thus, obtaining those particular traits which make species coexist in an ecosystem [114, 285]. Questions remain however on, what approach to use? And what kind of data is adequate for such purpose?

One extraordinary source of information on species distribution is such provided by the multiple ecological monitoring programmes. Over long periods of time, in some cases spanning few decades, these programmes have focused their attention towards more accessible and easier to quantify characteristics of species in ecosystems, such as the species identity, their biomass, and/or their abundance [32, 297]. The result are long-term datasets that have captured hundreds of species over large spatial scales. These datasets are, in fact, a fingerprint of the history of ecological communities and the ecosystems they inhabit [432].

Precisely using long-term monitoring data, Ryabov and colleagues [359] have developed an approach to infer *i*-traits. Using diffusion maps [79, 80], a nonlinear dimensionality reduction method, Ryabov's approach is able to reconstruct the *i*-trait space of a phytoplankton community by using only the biomass of phytoplankton species in samples. Diffusion map is a method from networks science. Its feasibility for ecological research is based on the fact that it considers not just isolated species in an ecosystem but also the interactions between them [127, 267]. At its core this method establishes a network of species in which only reliable comparisons are considered. High-dimensionality of ecological datasets is then turn to a lower, more relevant dimensional space, in which species can be located and evaluated according to their similarity or dissimilarity [79, 98]. Those species located closer to others in the network, either share similar functional attributes or tend to respond similarly to environmental factors. Consequently, the functional diversity can be computed from the pairwise distances in the *i*-trait space, for each sample in the dataset.

In spite of the feasibility of the method, its application in ecology is quite recent, thus multiple knowledge gaps must be filled. In one hand, diffusion maps have so far been applied to monitoring datasets of unicellular organisms, e.g., marine bacteria and phytoplankton communities [68, 127, 267, 359]. Biologically, this organisms have short time-spans, hence, fast reproduction and seasonal succession [57, 138]. Ecologically, due to limited mobility and size, their spatial distribution tends to be homogeneous at short scales, but highly influenced by seasonality, or inter-annual oceanographic changes [347]. Species with a longer time-span, varying from few years to decades, might have a single species occupying different habitats according to age-specific living requirements, mobility and competitive advantages [75, 259, 389]. A longer timespan also means they are subject to environmental disturbances which other organisms, e.g., bacteria or phytoplankton, are not [113, 141]. Expanding the application of diffusion maps to more complex organisms will allow us to asses the fitness of this method when dealing with species whose biology and ecology provide an extra layer of complexity.

On the other hand, most of the available long-term ecological datasets are gathered by monitoring programmes which, due to logistic or budget limitations, have focused mainly on monitoring stations [297]. Hence, the spatial resolution covered is usually low and mainly restricted to coastal areas [397]. Furthermore, these datasets are often cross-sectionally wide (e.g., census many interacting species) but short in the time dimension [76, 188]. This lack of appropriate time series data would limit the application of diffusion maps, and provide poor reconstruction of the *i*-trait space [21]. A feasible solution to address this issue is to aggregate those individual efforts (e.g. Helgoland, Marsdiep, Norderney, Rijkswaterstaat monitoring pro-



gramme) into a single time series which overcomes the shortcomings of a low spatial resolution, and a short time dimension [68, 266]. In this thesis, I extend the diffusion maps approach, testing its feasibility when assessing functional diversity in marine communities, seeking to fill exiting knowledge gaps, and overcoming the limitations imposed by the availability of data itself.

I start in Chapt. 2, first introducing important concepts and tools that I use in this work. After I give a brief introduction to the idea of complexity in ecosystems, I provide an overview of the paradigm change towards a functional view of biodiversity, the relation of biodiversity and ecosystem functioning, and the approaches to measure functional biodiversity in ecosystems. Thereafter, I explain the methodological approaches and main analytical tools which are used in this research. I introduce briefly some concepts of networks, to then focus on diffusion maps. In particular, I show how diffusion maps can overcome pervasive issues, such as the curse of dimensionality, and identify major explanatory variables in high-dimensional datasets. Lastly, I detail the process to calculate functional diversity from the reconstructed i-trait space.

In Chapt. 3, I apply the diffusion maps method to a long-term dataset of North Sea demersal fish abundances obtained from the ICES coordinated, International Bottom Trawl Survey (IBTS). The diffusion maps reveal the community variation is well-explained by two i-trait axis, which have structured the fish community in both the spatial and temporal scales. Using these newly identified variables, I could measure the functional diversity of each sampling haul, thus converting the long-term dataset into a functional diversity time series.

In Chapt. 4, I develop an approach to aggregate individually gathered monitoring datasets. Aggregation of datasets cannot be done directly, as any artifact from the data gathering produces clusters in data, delivering poor ecological insights. Instead, a more careful approach is proposed, one in which the particularities present in the data, which are the products of differences in sampling methodologies or personnel expertise in taxonomic identification, are not neglected. Therefore, we avoid the shortcoming of short datasets, and we expand the spatial area in which we can assess functional diversity.

In Chapt. 5 I use the previously developed dataset aggregation method to generate a dataset of phytoplankton abundance for the North Sea basin. To do this, I developed a computer programme which can standardize the process of data harmonization, data aggregation, diffusion mapping, and functional diversity measuring. Using this programme, I was able to evaluate the performance of diffusion maps with increasing addition of data. Furthermore, I could measure the functional diversity of the monitoring stations in the North Sea basin.

Finally, in Chapt. 6 I conclude this work by summarizing the results and expressing a brief outlook for future research. In particular, I envision the expansion of diffusion maps to heterogeneous groups of species and larger spatial areas. Equally relevant is to seek new sources of data (e.g. eDNA), and other tools, such as machine learning, which enhance the performance of this approach. An ultimate goal is the generation of an online engine which, powered by diffusion maps, can estimate functional diversity in samples, for regions that were already adequately assessed.

## Chapter 2

# Concepts and Methodology

Evelyn Hutchinson in his book '*An Introduction to Population Ecology*' [192], explores the evolution of the term niche as an ecological concept. Starting with Gause's idea on competitive exclusion [148], R. H. Johnson's conjecture about the species' assigned place in nature [203], or Joseph Grinnell's wondering whether the fauna of an ecosystem could ever be full [157]; the niche was always considered natural to any species, but affected only by competition with other species [148]. Hutchinson, however, highlighted C. S. Elton's book, *Animal Ecology* [123], as a significant change of direction for the ecological concept of the niche. For Elton, the niche is affected by "*all manner of external factors acting on an animal... so, the niche should be characterized as the place of an animal in the biotic environment, in relation to food and enemies*" [123]. Although preserving the idea of the niche as a species' given role in nature, Elton included the role of environmental factors on the shaping of any species niche [192].

Hutchinson's definition builds on Elton's idea, and considers that the niche, a geometrical shape with 'n-dimensions,' is influenced by a non-linear combination of species traits and environmental features [191]. Ecologists have suggested that to understand ecological communities we must first understand the factors which structure their constituent species' niche space [186]. Additionally, understanding how the niche space varies in different ecological communities could shed light into another relevant, but often elusive, concept in ecology, the biological diversity [39].

Simply known as biodiversity [255], the concept of biodiversity has evolved from a mere focus on species numbers, i.e., taxonomic richness [206], towards one in which more attention is given to the influence of species assemblages on the functioning of ecosystems [179, 391]. In other words, on the functionality of biodiversity for ecosystems. A functionality that is driven by particular species traits or environmental features, which make species fit to their environment and have a relevant role powering multiple ecosystem processes.[30] The main challenge in functional diversity is, consequently, to identify those composite traits that map ecological roles, or the properties that summarize organisms' functional capabilities [83, 178]. A powerful analysis method for meeting this challenge is offered by diffusion maps [79, 80]. This method defines a new coordinate system, a niche space for the species in the ecological community, where the axes, or variables, are non-linear composites of the most important factors structuring such community.

In this chapter the objective is to introduce important concepts and the methodology applied to develop this research. I begin in Sec. 2.1 with an introduction to the concept of complexity

applied to ecosystems. Then, I provide a short overview about the concept of biodiversity in the context of complexity, its role for ecosystems functioning, and the paradigm shift towards a functional view of biodiversity (Sec. 2.2). In Sec. 2.3, I will provide an introduction to the key ideas of Networks borrowed for this research. In Sec. 2.4 there is a detailed introduction to the non-linear dimensionality reduction method, diffusion maps, which is the key methodology driving this research. I close in Sec. 2.5, with a explanation on the methodology to measure functional diversity from the reconstructed trait space obtained via the diffusion maps.

## 2.1 Complexity in Ecosystems

Ecologist have long dedicated to study species and ecological communities. From pioneering works like those of Charles Darwin in the Galapagos Islands [92], Alfred Wallace in the Malay Archipelago [422], Charles W. Thomson and John Murray with the HMS Challenger Expedition [284], towards more recent examples such as the extensive surveys of Simon Ferrier in Australia [130] or Jorge Cortes in Costa Rica [85], the goal has been to find, describe and taxonomically classify new species. The focus has been mostly centered on the species [142].

Nevertheless, descriptive ecology is not a futile exercise [149]. Description is a first step towards understanding a system, stated Ricard Sole and Jordi Bascompte in 2006. A variety of authors however, acknowledge that description alone is not enough. Instead, the development of a theoretical framework is necessary. One that allows the generation of further insights and increases our capacity for forecasting and prediction [382].

Ecologist are increasingly facing the challenge of predicting the consequences of human-induced changes in the biosphere [382]. The consequences of biodiversity decline [67, 437], the reduction of ecosystems functionality [128, 287], the impact of harvesting activities over trophic webs [141, 344], or the vulnerability of socio-ecological systems to climate change [243, 283], are just a few of the many issues for which the formulation of general laws, such as the ones in physical sciences, are still unreachable [251].

But why is that ecology has not reached the formulation of any important law applicable to all natural systems or biological communities? Does something intrinsic exist in biological systems which prevents us from that? [227, 241, 355]. When comparing biological systems, i.e., a phytoplankton cell or a coral reef, to physical systems, i.e., the solar system or a simple pendulum, there are three fundamental differences which are worth paying attention. Biological systems posses features which provide them with functionality [187]. In other words, components of the system posses functions which resemble little to none the individual parts which they are composed of [381]. Second, this components can reproduce at multiple levels [428]. Third, such cycles of reproduction lead to the evolution of biological systems [54], eventually leading to the whole change of the system.

From an ecological point of view, this systemic view is summarized in the emergence of the term ecosystem [392]. Initially coined as a reaction to the excessive focus of naturalist on individual species, it proposed a study of the biotic as well as the abiotic world as a whole [3]. Currently, ecosystems are largely considered a complex environment, affected no just by the underlying variation of geophysical conditions, such as regional climate or nutrients availability, but also by the outcomes of self-organization, and the properties arisen from the interactions of their components [235].

This interaction between the internal organization of an ecosystem and the influence of external factors, makes it worth to wonder, how does this happens? How is that a group of apparently unrelated individuals formed an intricate network of interactions. More importantly, why an ecosystem is complex? Complexity refers to the impossibility to analytically decompose, reduce, or simplify a system without losing important characteristics [176]. In complex systems, the relevance attached to the interactions between their components, would compromise our understanding of the system's behavior if we analyze the system based solely on its components, ignoring these relationships by breaking such interactions [88].

If we consider true that any ecosystem is a complex environment, then the level of complexity that it exhibits can be understood as the dynamic interactions of its components. However, ecosystems can be understood at very different nested scales [382] (Fig. 2.1). At the basics of them all, single species. But beyond this level, interactions with other species must be considered. Phenomena such as population cycles [144, 223], the appearance of parasites and pests [166], or the organization of community assemblages and the influence of spatial variability on their distribution [424], cannot be understood by single-pointing a specific level of organization. Especially because this nested structure does not represent a hierarchical one, but rather a bidirectional influence between component and organizational levels. Therefore, understanding how complexity shapes ecosystems requires a frame of reference, which focuses on both components and interactions.

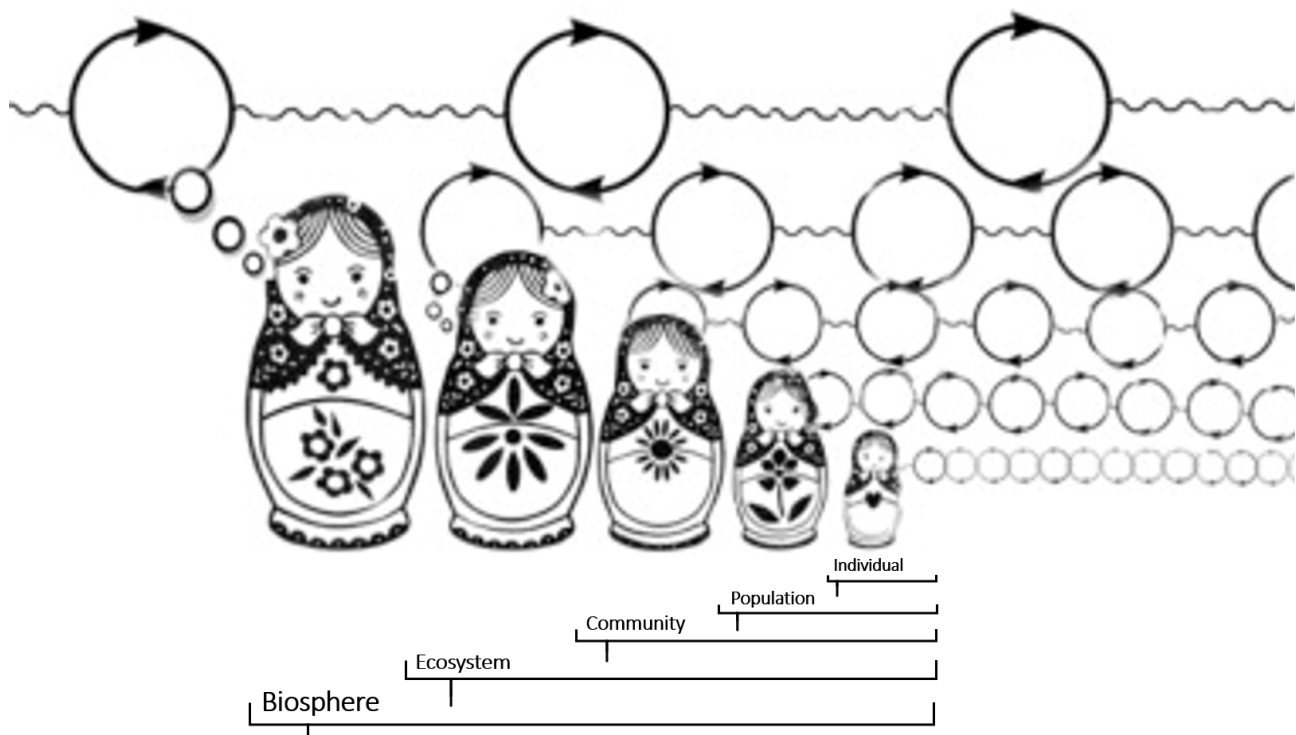


Figure 2.1: Levels of ecological organization represented by a set of Russian Nesting Dolls with varying degrees of feedback. As depicted, the largest doll in the set has the most complex flow of feedback and organization. But a similar process is occurring at every local level, with each inner doll responding to its own stream of feedback, self-regulation at the relevant scale in time and space, together accomplishing unified coherence across the whole. Reproduced from Kauffman 2015 [212]; open access under the Creative Commons Attribution License - CC BY 4.0.

Ricard Sole and Jordi Bascompte in their book *Self-Organization in Complex Ecosystems* considered that '*...ecosystems are the result of historical processes. The building of an ecosystem involves, on short time scales, path-dependent processes defining ecological succession. On larger time scales, species themselves change and co-evolutionary dynamics arise* [382].' This statement summarizes important characteristics of ecosystems, e.g., scale dependency, influence of physical constraints, dynamic interactions of species, lack of a clear top-down control,

absence of stability [104, 260]. Such properties lead the authors to catalogue ecosystems as a particular kind of far-from-equilibrium system, a complex adaptive system (CAS).

Back in 1995 the ecologist Simon A. Levin had already defined ecosystems as ‘*prototypical examples of complex adaptive systems... [235].*’ The concept of CAS refers to a particular type of open system in which their composing elements have a high interactivity between themselves and with the environment, thus, they can learn and modify the way they interact. These allow the system to self-organize and to adapt to changes in the surrounding environment [88].

John Holland attributed to any CAS four properties: aggregation, non-linearity, diversity, and flows [185]. But it was again Simon Levin, expanding the findings of Arthur and colleagues [14] from economics to ecology, who recognized that ecological CAS posses five intrinsic characteristic [235, 236], none totally independent to each other:

1. Localized interactions: the components of any ecosystem are constraint to local interactions, which are responsible for the emergence of features such as spatial patterns [46, 398], or the coexistence of competing species [373, 404].
2. Absence of well-defined top-down control: Top-down control in ecosystems is usually counterbalance by bottom-up forces [41].
3. Heterogeneity in network organization: Exist a network structure in ecosystems which pervades their behaviour and response to perturbations [163, 353, 441]. But it also influences the possibilities for adaptation of species. Heterogeneity also refers to the ‘diversity’ among components of the ecosystem [236].
4. Adaptation: The capacity to change as a response to external and internal influences. Such adaptation response occur both at single species level [261, 411] as well as at higher hierarchical levels, e.g., foodwebs [273].
5. Evolvability: This characteristic is what, ultimately, allows new features to emerge from the rearrangement of components and structures whitin ecosystems. What originally was might not be anymore given a considerable amount of time [277, 322].

Heterogeneity is one of he most recognizable features in any CAS [185, 235], but in an ecological context, it is commonly known as diversity. Diversity in ecosystems remains central in the ecological debate [100, 352]. Explored from a variety of disciplines and using multiple tools, the concept of diversity in ecosystems remains elusive and debatable. In the following sections the attention will be focused on grasping diversity in the context of complexity, addressing relevant questions such as, what tools can be used to understand diversity from a complexity approach? Can we measure the diversity of an ecological CAS, such as a marine ecosystems? How this diversity is distributed spatially? And, is diversity stable, or temporarily dynamic?

## 2.2 Biodiversity: functional or not?

The diversity of biological organisms, better known as biodiversity, is a powerful concept in ecology. It has been used in a variety of contexts and with slightly different variations, but

when it was first introduced it served as an indicator of the number of species which were present in a particular habitat [252]; and as a linking term for genetic diversity and ecological diversity, i.e., the number of species in a community of organisms [292]. Although neither of those publications offered a formal definition of biodiversity, it was clear that for the authors it was indisputably linked to taxonomic diversity [167].

Taxonomic diversity is the usual place to begin when thinking about biodiversity [83, 236]. This paradigm is based in two main assumptions: (1) all species are equal and their relative abundance determines their relative importance for the ecosystem, and (2) all the individuals of a particular species are equal in spite their length and weight [83, 257]. Consequently, a variety of methodologies were developed to measure the influence that the number of species, the dominance of certain species, the proportion of species, or the proportion of species change, have in structuring the distribution and patterns of diversity in ecosystems [177, 199, 206, 320, 371, 376].

Nevertheless, focusing only on taxonomic diversity meant that the distribution of species was largely a function of abiotic (physical and chemical conditions) and biotic factors (interactions among species such as competition, facilitation, disease, and predation). In other words, under this paradigm biodiversity is a passive consequence of extrinsic factors (i.e., climate, geology, and chance events), which set regional patterns in distribution and abundance, while biotic factors (i.e., number, type, and arrangement of interspecific interactions) secondarily modify regional patterns [78, 214, 333]. Such approach to biodiversity prevailed for many decades [286, 416]. But in the decade of 1990's the first traces of a new paradigm started to emerge [287, 367].

Ecologist started to abandon the idea of diversity limited to taxonomy and equilibrium notions, in favor of a view in which physical and chemical conditions of the environment are increasingly recognized as driven by ecosystem function (e.g., nutrient cycling and energy flow) [167, 286]. Shahid Naeem summarized this emerging view: "*...the existence of life alters the environment and the diversity of life determines the manner in which life alters the environment, much as if diversity were a catalyst to life's biogeochemical activities*" [286]. Under this new paradigm two issues take relevance. First, that biota plays an essential role in ecosystem processes [59, 410]. Second, that biodiversity plays a significant role in such processes [46].

### 2.2.1 The Biodiversity-Ecosystem function relation

The influence of life on ecosystem process has been reported in the literature decades before the new paradigm started to emerge. A classic example is the role of the sea otter (*Enhydra lutris*), in structuring the littoral and sub-littoral community of the Aleutian Archipelago [126]. Overexploited for most of the 18th century, sea otters disappeared in many islands of this archipelago, surviving only in remote, unpopulated bays. When James Estes and John Palmisano studied the continuing reduction of kelp forest and seagrass beds of this area, they found a direct relation between the overpopulation of sea urchins and the lack of sea otters. Sea otters control herbivorous invertebrate populations by predation, so after their removal, the increasing sea urchin grazing resulted in the destruction of macrophyte associations in the coastal ecosystem of the Aleutian Archipelago. This is a seminal example of what Robert Paine defined as a 'keystone species' [305], a species highly relevant in the configuration of the food web topology, thus, with capacity to affect ecosystem functionality in spite of their relative low

biomass [332].

The influence of diversity in ecosystems processes, however, is usually not that simple [250, 401]. In a variety of ecosystems, there are groups of species, rather than single species, performing critical tasks for the maintenance of ecosystem's functionality, e.g., honey bees, wasps, beetles, bats or flying foxes undertaking pollination [15, 306]; or highly specialized bacteria and cyanobacteria carrying the nitrogen-fixation function in ecosystems [372]. In either of these cases, the loss of a few pollinator species or nitrogen-fixing bacteria, would not cause a total collapse of the function [132, 356].

Paul and Anne Ehrlich have a great metaphor which points towards this: "*...ecosystems are like well-made airplanes, with redundant subsystems and features that permit them to continue functioning after absorbing a certain amount of abuse. A dozen rivets, or a dozen species, might be missed. On the other hand, a thirteenth rivet popped from a wing flap, or the extinction of a key species could lead to a serious accident*" [119]. These renowned ecologists highlighted two important points with this metaphor. First, that ecosystems might have 'alternatives' to replace those missing components, thus, avoiding the collapse of system functions. Second, that cumulative effect of species loss might have a higher impact, even if the loss of individual species does not.

The two previous examples show the importance that species diversity have for ecosystem processes, either as top-down regulators or as bottom-up providers. This is why a sole focus on taxonomic diversity would miss the importance of biodiversity below, as well as above the species level [236]. Moreover, we must remember that, in any ecosystem, species are present in populations rather than as isolated organisms. Population ecologists have argued that decimated species populations, with just a handful of individuals cannot fulfill the same role in the ecosystems, so whether a species is present or not provides limited information to assess diversity [146].

Nonetheless, it is yet unclear how diversity contributes to ecosystem functionality [198]. Initial experimental research supported a positive relationship of biodiversity to ecosystem functionality and stability, thus having a higher variety of species worked as a sort of insurance which prevented declines in ecosystems functioning caused by environmental fluctuations [276, 287, 439]. Empirical research has reported examples of ecosystems in which the relation biodiversity-ecosystem function appeared more complicated than initially thought [383]. Thus, it is now relevant to focus attention on how scientist assess functionality in ecosystems and how functional, in fact, is biodiversity.

## 2.2.2 How can functional diversity be measured?

Due to the growing interest for understanding the relation between diversity and ecosystem functioning [47, 159, 440], the definition and limits of what functional diversity represents has varied over the past 30 years. Whereas for David Tilman functional diversity represents, '*the value and range of those species and organismal traits that influence ecosystem functioning*' [400]; for Diaz and Cabido it was related to '*the number, type and distribution of functions performed by organisms within an ecosystem*' [105]. Petchey and Gaston talked about functional diversity of communities as '*the diversity of traits present in a community weighted by their abundances*' [317]. More recently Zhang and colleagues argued that functional diversity is

*'based on the various traits of species and their adaptations and responses to environmental change'* [442].

In spite the multiple definitions, most researchers have agreed on the direct link between functionality and traits [317], because the diversity of traits can better reflect species resource acquisition and niche complementarity [265, 328]. Therefore, functional diversity is related to the diversity of functional traits which are present in a biological community with respect to their presence and abundance [83, 356, 418]. In simple words, measuring functional diversity is about measuring functional trait diversity [317]. Consequently, quantifying functional diversity in ecosystems is, in theory feasible, but in practice it is not simple.

Several methods and indices have been described and discussed on how to calculate functional diversity [73, 317, 326, 349, 364]. However, most of them do not estimate all the components of biodiversity that influence ecosystem functioning in a satisfactory way, thus rendering the interpretation of results difficult and controversial [349]. While a focus on species diversity only demands to count individuals from different species (i.e., sort them into several categories), to focus on trait-functional diversity, demands extensive knowledge about the species and the identification of those traits which influence functionality [242]. Although, intuitively we might expect that different species possess different traits, in ecosystems some species are very similar to each other, i.e., functional redundancy; while some are very different, i.e., functional distinctiveness [363].

In order to understand this, we can think again about the two cases described in the previous section. In ecosystems exist a variety of manners in which species support functionality. For example, for the coastal area of the Aleutian Archipelago, only one species held control on invertebrate grazers population [126]. For a grassland, pollinators hold major relevance, but probably there will be more than one species undertaking this function [331]. For many ecologists this was the first approach toward functional diversity. Because in ecosystems, function seems to be supported by a group of 'key-species,' rather than by a single species. Simon Levin called this a 'keystone group' [236]. Currently, that idea diverged to what is commonly known as a 'functional group' [218, 343, 400].

A functional group is defined as a set of species exhibiting similar responses to environmental conditions and having similar effects on the dominant ecosystem processes [132]. Although it offered more insights to the initial taxonomic approach, this idea soon reached serious limitations, because it demanded exhaustive knowledge about what characteristics of the species could be considered functional or not, it does not distinguish possible ontogenic changes on species, it assumes that inter-specific differences are higher than intra-specific differences among species, and lastly, because it considers absolute functional equivalence within a group [34, 317]. Since the constitution of groups anticipates the assessment, the determination of the functional groups quantity as well as the placing of species within groups is attain to high subjectivity and the researcher level of expertise [83]

On the contrary, a direct approach, one that can assess functional diversity based on specific functional traits measured for each species is more desirable and promises a finer resolution [52, 317]. On this regard, Owen Petchey and Kevin Gaston argued that an index would successfully measure functional diversity if, ideally, meet each of the following conditions [317]:

1. Appropriate functional information (traits) about organisms to be included in the measure, and irrelevant information to be excluded (what functional traits should be in-



- cluded?) [233].
2. Traits to be weighted according to their relative functional importance [316, 354].
  3. The statistical measure of trait diversity to have desirable mathematical characteristics [50, 349]. For example, discontinuous vs. continuous measures of diversity.
  4. The measure to be able to explain and predict variation in ecosystem level processes.

Three of the conditions are related to the second most relevant concept for this research, ‘the functional trait.’ A trait is a well-defined, measurable property of organisms, usually measured at the individual level and used comparatively across species. A functional trait is one that strongly influences organismal performance [275, 419]. Functional traits can be morphological traits that represent adaptations to different diets or habitats, physiological traits (e.g., temperature tolerance), reproductive traits (e.g., number of eggs and egg diameter), or behavioral traits (e.g., migratory behavior or parental care) [52, 233].

But regardless the great enthusiasm that trait-based functional diversity generated among ecologist in the decade of the 2000’s [275], serious limitations soon emerged. The obvious one is related to the availability of traits themselves. For instance, it is easier to categorize fish species by their general diet than to obtain measurements on their size, gape width, stomach length, etc. [87]. Furthermore, ecosystems are composed of dozen if not hundreds of species, thus, having a complete list of functional traits is an almost impossible task.

When functional traits are available, researchers interested in applying functional diversity indices need to make various methodological decisions such as, how many and which traits to use, how to weight them, how to combine traits that are measured at different scales and how to quantify the species’ relative abundances in a community [233]. This again circles around to the initial problem of the functional groups approach. As subjectivity is unavoidable in science, the ideal is to minimize its influence, therefore, a functional diversity index should reduce to the minimum the steps in which subjective decision can bias the analysis [359].

If the points addressed by Owen Petchey and Kevin Gaston [317] advocated for an ideal index, following them in sequential order would demand an extensive knowledge about the ecosystem and their biological communities. This limits the application of any index to just a handful of well-researched environments. On the contrary, in this thesis I argue that an inverse sequential approach to this conditions would provide the necessary basic raw material to evaluate functional diversity in any ecosystem which has been adequately monitored:

1. Use a measure which is able to predict variations in the presence and abundance of species.
2. Create a product with mathematical characteristics, which allows to discern those species according to their functional redundancy (functionally similar species), or to their functional distinctiveness (functionally dissimilar species).
3. Based on the organization and distinctiveness of our community, we will discern those likely traits which made them organized in such particular manner, therefore, inversely reconstruct those traits which can be functional.
4. After traits have been weighted according to their relative functional importance, a functional diversity assessment can be perform using an adequate index.

This approach requires to reconstruct the functional trait space (i.e., coordinates which corresponds to a selected trait), in which points will represent species of the community [364]. This is a novel approach to ecological research, one which seeks to turn around the usual ecological postulate [215], using a measure of how species varied in the ecosystem first, obtaining those functional traits which had likely influenced such organization afterwards. We will return to a detail explanation of the postulate and the method used to calculate functional diversity from long-term monitoring datasets in the Sec. 2.5. It is now time to explain in detail the methods used in this research.

## 2.3 Networks

Ecosystems are the union of hundreds, if not thousands, of species interacting dynamically with each other and with the physical environment surrounding them [153]. Although these interactions have a key role for ecology and evolutionary biology, it was only recently that the increasing availability of computational tools to store and retrieve biological data and multi-disciplinary approaches to science, have facilitated a wider application of networks theory by biologists to address multiple open questions in ecology [336].

Network thinking is by no means new to ecology. Great ecologists such as Lindenman and Odum pioneered the use of networks as a way to represent and describe food webs [239, 298]. Recent decades have seen the application of networks theory to diverse areas of ecology such as spatial ecology and metapopulations [163, 408], epidemiology [12, 308], or evolutionary dynamics [238, 360]. This is because, from a complexity perspective, the patterns of connections in a given biological system can be represented as a network [288].

A network is a collection of units which interact with each other as a system [336]. In the most simple case, a network can be represented by a set of uniform nodes connected by undirected edges (Fig. 2.2). The nodes represent units at most levels of the biological hierarchy, from genes and proteins to neurons and organs and limbs, and from individuals in a population to species in a community. Edges, or links, represent some kind of interaction between nodes, including transcriptional control, biochemical interaction, energy flow and species interactions [288, 336].

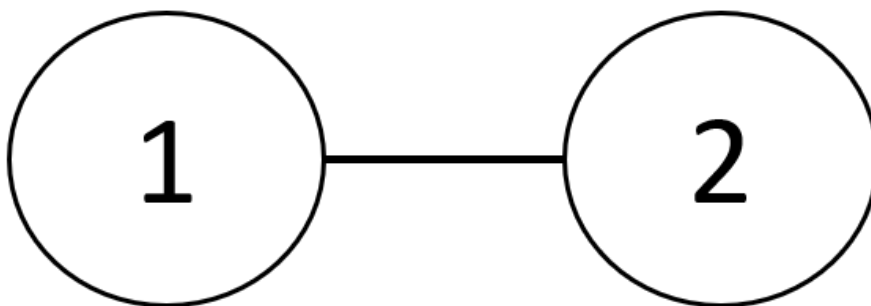


Figure 2.2: Example network of two nodes connected by one undirected link.

Representation and description of networks has been the domain of a branch of discrete mathematics known as graph theory [44, 45]. With a large history, whose beginning dates

back to 1736 when Leonhard Euler published the solution to the Königsberg bridge problem [42], graph theory has provided answers to a series of practical questions such as: what is the maximum flow per unit time from source to sink in a network of pipes, how to color the regions of a map using the minimum number of colors so that neighboring regions receive different colors, or how to fill  $n$  jobs by  $n$  people with maximum total utility. The end of the 1990s decade nonetheless, saw the raise of a new movement of interest and research which focused on ‘complex networks’ [271, 288].

The rise of the internet fueled the interest in complex networks, and two seminal papers disseminated its application to biological systems, that by Watts and Strogatz on small-world networks in 1998 [425], and that by Barabási and Albert on scale-free networks which appeared one year later [19]. Biological systems are perfect examples of complex networks, a network whose structure is irregular and dynamic [271, 336]. Applications of this framework on real networks has raised continuously covering novel topics and problems in network structure and their dynamical behaviour [120, 406]. In spite network representation of biological systems will necessarily involve condensation of information, this approach provides a useful tool to tackle full communities interactions and dynamics, rather than interactions of pairs of species [22].

Networks has raised as a science on its own, with proper terminology, an evolving theory and methodologies which are applied to other scientific disciplines [390]. Extensive reviews, such as [42, 289, 390], or specialized books, e.g., [31, 49, 110] offer an in-depth exploration of networks theory and applications. In the next subsection, I will focus on introducing a formal mathematical description of network representation in the context of this research, touching upon the most relevant theoretical background borrowed from the networks theory.

### 2.3.1 Network representation

The arrangement of nodes and interactions within a network is known as topology, and the best way to represent such topology is by using an adjacency matrix [42, 288]. The adjacency matrix is a square matrix with dimensions corresponding to the size of the network. For every link in the network, e.g. from node  $i$  to node  $j$  the adjacency matrix has a nonzero entry at position  $A_{i,j}$ , whereas all other entries are set to zero, indicating the absence of a link. For an unweighted network, the entries are 1 (link present) or 0 (link absent), whereas for a weighted network, the adjacency entries represent the weights of the corresponding links. The adjacency matrix for the example network in Fig. 2.2 is

$$\mathbf{A} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \quad (2.1)$$

Closely related to the adjacency matrix is the laplacian matrix [74], which can be derived as follows:

$$\mathbf{L} = \mathbf{D} - \mathbf{A} \quad (2.2)$$

In the Eq. 2.2,  $\mathbf{D}$  represents the degree matrix of the network. For the case of undirected networks, the number of links that connects to the respective node.  $\mathbf{D}$  contains the sum over each row of  $\mathbf{A}$  on its diagonal, as

$$D_{i,i} = \sum_j A_{i,j} \quad (2.3)$$

So replacing Eq. 2.2 with matrix notation we obtain

$$\mathbf{L} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} - \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad (2.4)$$

$$\mathbf{L} = \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} \quad (2.5)$$

The Laplacian matrix captures the amount of divergence from a point, i.e., it measures whether the neighboring points of a point are larger than it or smaller than it or are on a line, thus, showing how much linear the neighborhood of a point is [150]. In a single phrase the Laplacian is a measure of relation of neighbor points [28].

Matrices such as the Adjacency or Laplacian are useful in ecology because they allow reduction of multi-factor systems, such as a biological community, into linear systems which can be mathematically handled [443]. This is performed by means of what in spectral theory is known as eigenvalues and eigenvectors of a matrix [361]. To understand what eigenvectors and eigenvalues represent, let's introduce a simple example. Imagine we have a vector  $\boldsymbol{\nu}$  of coordinates [1,2], multiply by a matrix  $\mathbf{A}$  as

$$\begin{pmatrix} 5 \\ 10 \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \end{pmatrix} \cdot \begin{pmatrix} 1 & 2 \\ 8 & 1 \end{pmatrix} \quad (2.6)$$

The product of  $\boldsymbol{\nu}\mathbf{A}$ , could also be represented as

$$\boldsymbol{\nu}\mathbf{A} = 5 \begin{pmatrix} 1 \\ 2 \end{pmatrix} \quad (2.7)$$

In this case, we call 5, an eigenvalue ( $\lambda$ ) and  $\begin{pmatrix} 1 \\ 2 \end{pmatrix}$ , an eigenvector ( $\boldsymbol{\nu}$ ) [27]. The eigenvectors of a matrix are those vectors for which multiplication by a matrix results in a vector in the same direction or opposite direction to the original, as

$$\mathbf{A}\boldsymbol{\nu} = \lambda\boldsymbol{\nu} \quad (2.8)$$

The eigenmodes are also useful because they allow to track how a system evolves step-by-step [13]. Therefore, the matrices used to represent and resolve networks will be the cornerstone for understanding and developing further our methodology in the next section.

## 2.4 Diffusion maps

David Donoho called this the ‘century of data.’ He claimed that: ” *A combination of blind faith and serious purpose makes our society invest massively in the collection and processing of data of all kinds, on scales unimaginable until recently*” [108]. For Donoho at the verged of the new

century, our society faced three major intellectual trends: that data will be more than ever ubiquitous; data analysis will be extremely relevant to make sense of data; and that more than ever there will be a pervasiveness of high-dimensionality [108].

The dimensionality of a dataset, refers to the number of variables measured per sample which can easily amount to thousands [98]. Consider the following example, a sampling campaign in a bay to collect information about its fish community. Each time we do a hauling we catch tens or even hundreds of species, the dimensionality of the data space equals the number of different variables that are recorded in the dataset, e.g. the number of detected species, the weight of each fish, the temperature or salinity of the water at the moment it was catch, hence several hundreds, if not thousands.

High dimensionality in datasets makes it challenging to understand what drives the variation in such system. Let's imagine that in our previous example one fish species is enough to describe the community and its variation. For instance, we would need to check in our dataset those samples with the highest and the lowest abundance of that particular species. If in our bay two rather than one species matter, the number of extreme points would double. The dataspace has now four corners, the lowest value of species 1, the highest value of species 1, the lowest value of species 2 and the highest value of species 2. Therefore, each variable that is added, doubles the number of corners of the dataspace. Let's say we find 1,000 different species, each sample would then map to a point in a 1,000 dimensional space. This space would have  $2^{1000}$  corners, which is greater than the estimated number of species on Earth of around 1-6 billion ( $\sim 2^{230} - 2^{233}$ ) [225]. This exponential increase is the first consequence of what is usually referred as the 'curse of dimensionality' [29, 417, 438].

To overcome this issue it is important to realize that, although datapoints lie in a coordinate space, they only cover a part of such space. In our fish sampling we would not expect to find all possible combination of fish species of the bay, but rather species appearing together may do so because they use similar resources, profit from certain interactions or benefit from the same environmental conditions [373, 426]. Due to such interactions and dependencies our datapoints might approximate a curved surface, or some other comparatively low-dimensional object within the data (Fig. 2.3). These underlying structures are summarized under the term manifold [253] and the goal of locating such structures is known as dimensionality reduction [69, 384]. It would therefore be possible to characterize the data and the relationship between individual datapoints using fewer dimensions, if we were able to measure distances on the manifold itself [98].

PCA (Principal Component Analysis)[204] is the most traditional tool used for dimensionality reduction. PCA projects data on a lower-dimensional space, choosing axes that keep a maximum of the data initial variance. Unfortunately, PCA is a linear tool, thus, non-linear relations between the components of the initial data may get lost [417]. In fact many statistical frameworks (e.g., principal components analysis, generalized linear models, multivariate autoregressive models), which reduce linearly high-dimensional data, assume that causal factors do not interact with each other and have independent or additive effects on a response variable [440]. Such assumptions might cause the loss of relevant information, specially when dealing with data in biological fields, where the relevant units (e.g., species or other variables) may not behave according to fundamental equations [99] and where datasets are often cross-sectionally wide (e.g., census many interacting species) but short in the time dimension [76, 188].

The challenge, then, is to determine the lower-dimensional data structure that encapsulates

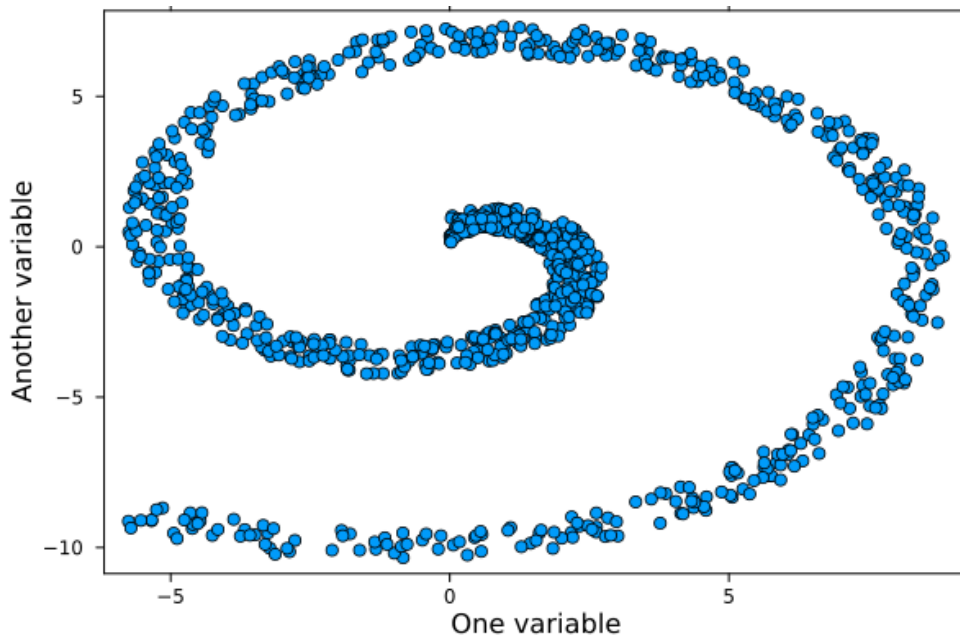


Figure 2.3: Example of a data manifold. The datapoints can be described in terms of ‘One variable,’ and ‘Another variable.’ However, since the datapoints cluster around the spiral structure, dataset could also be explained in terms of this new variable, that follows the main dimension of the data manifold.

the data, achieving dimensionality reduction at the time the important relationships between datapoints are preserved. A methodology offering a solution is diffusion maps [79, 80]. Diffusion maps, a non-linear technique, achieve dimensionality reduction by re-organising data according to parameters of their underlying geometry. This method has proven more reliable because it allows to establish a network of species in which only reliable comparisons are considered. Thus, species located closer to other in the network either share similar functional attributes, or tend to respond similarly to environmental factors. The establishment of such network recreates the original high dimensional space in a manifold which will then allow us to measure the diffusion space between points to determine how dissimilar, or not, the points of our network are [79].

The next section will provide a detail explanation of how diffusion map works; but also of how by applying diffusion maps it is possible to approximate the inverse approach to determining functional diversity of ecosystems by using long-term ecological datasets.

### 2.4.1 Diffusion maps process method

Diffusion map was first introduced by Ronald Coifman and colleagues [80], here we use a variation of the method by Edmund Barter and Thilo Gross [21], which consist of the following six steps:

1. Standardize the data.
2. Compute distances between all datapoints.
3. Construct a similarity matrix.
4. Threshold the similarity matrix.
5. Define a Laplacian matrix.

## 6. Eigendecomposition of the Laplacian provides new variables

We examine each step of the diffusion map method by revisiting our hypothetical bay fish community. Our starting point is the data matrix  $\mathbf{A}$  with the dimensions  $M \times N$ , where  $M=1,000$  is the number of datapoints, i.e., the fish sampled, and  $N=200$  is the number of sampling hauling performed in the bay.

First step is to standardize the data. Standardization is a common procedure to the analysis of complex ecological datasets. Its objective is either bring data recorded in distinct units to a common scale or eliminate size differences in datapoints [230, 293]. Here we bring all values in the range  $[0, 1]$ , i.e., a mean equal to zero and a standard deviation equal to 1 [62], according to

$$\hat{A}_{m,n} = \frac{A_{m,n} - \mu_n}{\sigma_n} \quad (2.9)$$

with

$$\mu_n = \frac{\sum_m A_{m,n}}{M} \quad (2.10)$$

$$\sigma_n = \sqrt{\frac{\sum_m (A_{m,n} - \mu_n)^2}{M}} \quad (2.11)$$

being the mean ( $\mu_n$ ) and standard deviation ( $\sigma_n$ ) of the  $n$ th column of  $\mathbf{A}$ .

In the step 2 the standardized datapoints will be used to compute the Euclidean distances between all datapoints. Thereby we obtain an  $M \times M$  distance matrix  $\mathbf{D}$ , where

$$D_{i,j} = \sqrt{\sum_n (\hat{A}_{i,n} - \hat{A}_{j,n})^2} \quad (2.12)$$

is the Euclidean distance between the datapoints  $i$  and  $j$  in the data space. This includes also many long-distance comparisons.

Now we define similarities of two datapoints as the inverse of the Euclidean distance of the respective datapoint pair. As a result, we convert the distance matrix into a similarity matrix  $\mathbf{C}$  (step 3), where

$$C_{i,j} = \frac{1}{D_{i,j}} \quad (2.13)$$

is the similarity of the datapoint pair  $i, j$ . The diagonal elements of the matrix  $\mathbf{C}$ , those comparing each datapoint to itself, are set to zero. The selection of the distance and similarity metric depends on the dataset at hand. Hence, a variation might be necessary in this step

depending on the selected index of similarity [21, 359]. For instance, in some research the Spearman correlation coefficient [385] has been applied as similarity measure, e.g., for phytoplankton abundance dataset [359]; in others the hamming distance was used as distance measure to compare bacterial genomes using gene presence-absence data [127, 267], whereas the most common measure of similarity, the Pearson Correlation Coefficient [310], might be of limited use due to being mostly appropriate for linear correlations [311].

Here it is worth pausing to explain the concept of similarity and how it operates for diffusion maps. The idea of distance as a measure of similarity is not new for ecology [380]. Originally applied to measure taxonomic similarity [11, 172], it is used now in a broader sense, implying similarity as a measure of correlation between datapoints due to not just intrinsic but also extrinsic factors. For the case of high-dimensional space, distances are often large due to datapoints being sparsely distributed. As a consequence we are left with a mixed of small similarity distances indicating high correlation, but also large distances, which provide very little information on the nature of the discrepancy [98]. Translating this to the bay fish community, we could have that ten fish species appear together in most of the samples, whereas the rest of fish species appear in lower frequency or in specific locations. In one hand, when comparing those ten species, a high similarity score is likely, which could be assessed against factors common to the environment or others which might influence such shared distribution. On the other hand, when comparison are done to the less frequent species, a lower degree of similarity might result, but the reasons for that can be hardly attributed to any particular factor.

To eliminate all long-distance comparisons, diffusion maps threshold the similarity matrix to local comparisons only, thus, defining the area within which we trust our local similarity measure to be accurate [21, 79, 98]. This is done in step 4, where we introduce a kernel threshold number ( $\alpha$ ). This kernel will modify the degree of connectivity of our network. By tweaking the kernel scale, we choose the size of the neighborhood, based on prior knowledge of the structure and density of the data. For intricate, non-linear, lower-dimensional structures, a small neighborhood is chosen. For sparse data, a larger neighborhood is more appropriate. Previous research studies showed that the value of 10 is often a good threshold, providing the most robust results [127, 147, 267, 359]. Hence, for this research the threshold kernel was set to 10, i.e., only the top-10 highest similarity entries were kept for each datapoint and all other entries in the matrix were set to zero. Therefore, an entry  $C_{i,j}$  is kept if it is among the top-10 highest similarity scores for either datapoint  $i$  or for datapoint  $j$ , or both.

For step 5, having now built a network of trusted comparisons, we will quantify the distance between dissimilar datapoints as the distance on this network. A way to approach this is to consider the shortest path distance [330]. However, due to the dependence of this distance on the presence and absence of single links, it is very susceptible to noise. Diffusion maps use the concept of diffusion distance [80, 98], which robustly quantifies the distance between datapoints. Along the geometric structure, points are dense and therefore highly connected. Pathways form along short, high probability jumps. On the other hand, paths that do not follow this structure include one or more long, low probability jumps, which lowers the path's overall probability, as for example in Fig. 2.4, where the red path becomes a viable alternative to the green path as the number of steps increases when we follow the red manifold.

The diffusion distances can be computed from the eigenvectors of the Laplacian matrix. Accordingly, from the thresholded similarity matrix we compute the corresponding  $M \times M$  row-normalized Laplacian matrix, defined by



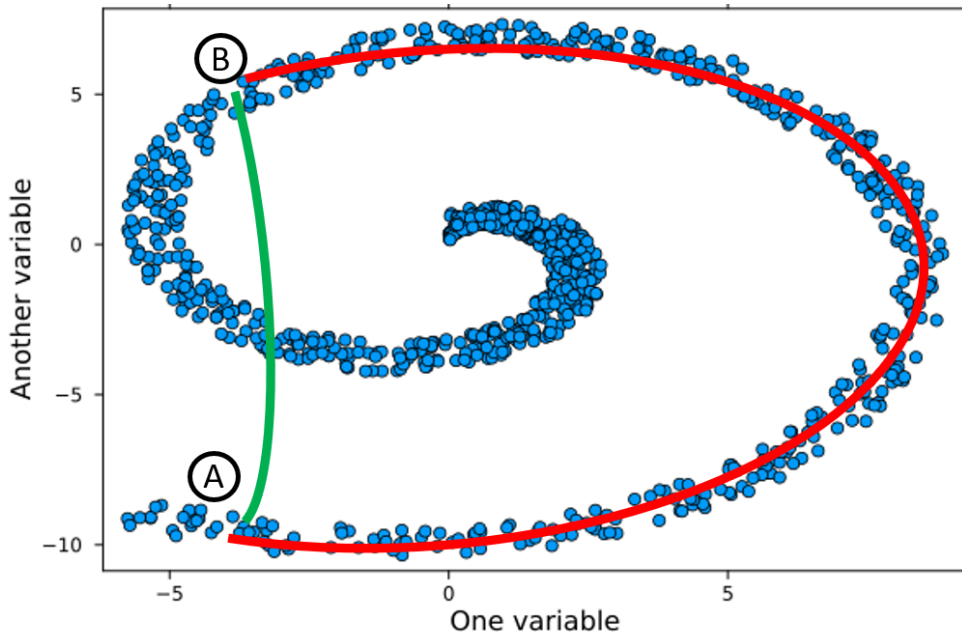


Figure 2.4: Paths along the low-dimensional geometric structure of the dataset. The red path becomes a viable alternative to the green path as the number of steps increases when we follow the red manifold. The circles A and B are dataset points in the manifold

$$L_{ij} = \begin{cases} 1 & \text{for } i=j \\ -\frac{c_{ij}}{\sum_k c_{kj}} & \text{otherwise} \end{cases} \quad (2.14)$$

Laplacian matrices, introduced in Sec. 2.3.1, are positive semi-definite matrices closely related to many natural processes such as different types of diffusion processes, heat conduction, or the spreading of vibrations [325]. They hold some interesting, useful dynamical properties [361]:

1. At least one of its eigenvalues is zero.
2. All the other eigenvalues are either zero or positive.
3. The number of its zero eigenvalues corresponds to the number of connected components in the network.
4. If the network is connected, the dominant eigenvector is a homogeneity vector.
5. The smallest non-zero eigenvalue is called the spectral gap of the network, which determines how quickly a diffusion process takes place on the network.

The properties 1, 2 and 3 will help us to keep the connectivity of our network. If we obtain more than a single zero eigenvalue, the number of components of the network is more than one [131], which means that the network has become disconnected. When this happens we have to repeat our analysis applying a higher threshold, in step 4. The properties 4 and 5 point out towards our next and final step 6.

The diffusion map preserve a dataset's intrinsic geometry, and since the mapping measures distances on a lower-dimensional structure, we expect to find that fewer coordinates are needed to represent datapoints in the new space. The question becomes which dimensions to neglect, in

order to preserve diffusion distances (and therefore geometry) optimally [98]. Eigenvectors have a dimensionality equal to the number of datapoints, thus each eigenvector assigns one value to every datapoint, with an associate eigenvalue. This eigenvalue indicates the importance of each dimension, having the smallest non-zero eigenvalue gathering the directions of the largest variation, or the main dimensions of the data manifold [253].

The eigenvectors represent new variables that are non-linear combinations of the original variables and can be interpreted as coordinates in trait space [21, 80]. Hereby, we have moved from a high-dimensional space which is difficult to make sense of, to a space in which we can confidently compare datapoints by measuring the Euclidean distances in this new trait space [359].

We can conclude that diffusion map is a dimensionality reduction method, which is able to find non-linear structures embedded in complex high-dimensional systems, such a biological communities [127, 359]. This method's only parameter is the selection of the threshold kernel, thus, it reduces to a minimum subjectivity and avoids over-parametrisation, which is often an issue in data analysis and ecological modeling.

## 2.5 Calculation of functional diversity

Diffusion maps have had successful applications into distinct research in ecology. Using this method researchers were able to identify metabolic strategies in marine bacterial communities [127, 267], reconstruct functional traits from marine phytoplankton time series [359], even enhancing trait reconstruction by merging phytoplankton datasets from different origins [68], and to interpret the  $\beta$ -diversity between geographically distant samples [147]. This highlights diffusion maps as a tool that yields new explanatory variables that represent composite functional traits of the studied organisms. Such new variables are derived solely from ecological monitoring data.

Returning to our postulate of an inverse approach to the Owen Petchey and Kevin Gaston ideal index [316] (SubSect. 2.2.2), the application of diffusion map to a long-term time series have allowed the reconstruction of a functional trait space. Therefore, it is possible now to determine the distinctiveness of species based on the approach developed by Ryabov and colleagues [359]. Consequently, for each species pair,  $i$  and  $j$ , functional distinctiveness is defined as  $d_{ij}$ , which is the euclidean distance in the reconstructed trait space, where the species traits are now given by the eigenvector elements corresponding to the species, re-scaled by the respective eigenvalue, as in

$$d_{ij} = \sqrt{\sum_k \left( \frac{v_{k,i} - v_{k,j}}{\lambda_k} \right)^2} \quad (2.15)$$

where  $v_{k,i}$  is the eigenvector of species  $i$  present in sample  $k$ , whereas  $v_{k,j}$  is the eigenvector of species  $j$  present in sample  $k$ ,  $\lambda_k$  is the eigenvalue corresponding to sample  $k$ .

The pairwise functional distances are then used to calculate the functional diversity of each sample by means of the Rao's quadratic entropy [342]. The functional diversity (FD) for sample

$k$  is computed as

$$FD_k = \sum_{i=1}^{n-1} \sum_{j=i+1}^n d_{ij} p_k^{(i)} p_k^{(j)} \quad (2.16)$$

where  $d_{ij}$  is the pair-wise distance between species  $i$  and  $j$ ,  $p_k^{(i)}$  and  $p_k^{(j)}$  are the relative abundance of species  $i$  and  $j$  in sample  $k$  and the summation indices  $i, j$  run over all  $n$  species in the system [50, 350].

Obtaining a functional diversity score for each sample of our dataset will allow functional diversity variation assessment, both in the temporal and the spatial scale, of any ecosystem that has been moderately monitored. Hence, all the information which remains store in ecological monitoring datasets can be unravel and re-interpreted by means of diffusion maps. Nevertheless, there still remain multiple knowledge gaps, precisely on determining how far this methodology is useful to understand functional biodiversity (Chapter 3), the limitations imposed by data availability, extension and standardization for the accuracy of results (Chapter 4), as well as the applicability to different group of organisms, or to wider and longer spatial and temporal scales (Chapter 3 and 5). In the following chapters we will address this knowledge gaps, approaching functional diversity from a complexity perspective by applying the methods of networks introduced here.

## Chapter 3

# Diffusion mapping a long-term time series: The North Sea Fish Community

Since 2010, when the parties of the United Nations Convention on Biological Diversity (CBD) adopted the Aichi Targets for 2020 [407], increasing efforts have been advocate to ‘improve the status of biodiversity’. But in spite of global agreement, and the urgency in the matter, pressures on biodiversity have increased. Having unclear or incomplete reports about biodiversity changes and trends [72, 179, 402] is a direct consequence of the limited availability of long-term biodiversity indicators [5]. Although good biodiversity-indicator programmes exist, they are biased towards developed countries in temperate regions [190, 262], or are limited to short-time periods, thus making it difficult to monitor biodiversity change across wide spatial scales and longtime frames [208, 274].

While there is an urgent need to initiate well-designed programmes to measure changes in biodiversity [5, 240], there is currently an effort to use existing information, often collected for other purposes, to gain insights on how the state of biodiversity is changing [247, 335, 359]. Ecological long-term datasets of species, which gathered data on biomass, abundance or just presence-absence, represent a unique source for assessing biodiversity, given the appropriate tools and methodology [269, 279].

This chapter focuses on presenting a framework based on diffusion maps to coarse-grain a taxonomically-rich fish community in terms of their most relevant functional traits. Our approach reduces the high-dimensionality of the time series, reconstructing the trait-space and creating a network of trusted connections. Applying this framework to a long-term fish abundance time series from the North Sea, I am able to reconstruct the dynamics of the functional diversity of the individual samples in the dataset.

We start with a general introduction to the dynamics of fish communities of the North Sea (Sec. 3.1). In Sec. 3.2, the diffusion map approach is applied to the North Sea fish community time series. The method identifies important functional i-traits associated with community variation and delineates the network structure of the community. In Sec. 3.3, I estimate the functional diversity of the North Sea fish community over time and across the North Sea basin. Finally, I discuss the results in Sec. 3.4.

### 3.1 Dynamics of fish communities

In an ideal community scenario, an homogeneous species spatial distribution would spread the risks of extinction over various habitats when species encounter environmental changes [35]. However, species naturally distribute heterogeneously and adjust their spatial distribution in response to population dynamics and environmental variability [63, 296, 340]. For instance, heterogeneous distribution of fish species is ruled by biotic or abiotic factors which, according to different degrees of relevance, shape what is known as the ‘species niche’ [111].

The niche as defined by Hutchinson, is a multidimensional hypervolume in which a species can maintain a viable population [191]. The niche concept is enhanced by M. Tokeshi’s ideas of the niche changing over ecological times, and niches of different species overlapping partially, but not completely in ecosystems [403]. The niche is, therefore, a proxy of those factors structuring a particular biological community, and the niche space the key to understand the functionality of species in ecosystems [337].

Diffusion maps offer a way to reconstruct the niche space of a community, by assessing ecological similarities of co-occurring species [79, 80]. In other words, by assessing those species which occur together we can get a hint on the factor, or traits, which summarize a species’ niche in a given community [234]. Nevertheless, for the case of fish, their biology provides a layer of complexity. Fish species have a longer timespan, varying from few years to decades, opening the possibility for a single species to occupy different habitats according to age-specific living requirements, mobility and competitive advantages [75, 259, 389]. Moreover, some species are massively abundant, which leads to a high intraspecific competition [209]. A longer timespan also means they are subject to environmental disturbances, which other organisms, e.g., bacteria or phytoplankton, are not. Therefore, oceanographic cycles, or other man-made induced changes will also influence the survival rates of fish species or the habitats they can occupy. [113, 141].

Additionally, assessment of functional diversity in fish communities also needs careful attention. It has been shown that for fish communities the scale in which we perform the analysis might influence the relevance of some factors over others. In one hand, at short spatial scales, biotic factors, also known as ‘limiting similarity’ or interspecific competition [254], prevents species with similar ecological niches from co-occurring [17, 77, 101, 312]. On the other hand, at larger spatial scales a strong abiotic control, known as ‘environmental filtering,’ results in co-existing species that are likely to share similar ecological traits that enable them to survive in such environment [213, 419]. Therefore, whether the resulting functional diversity is high [281] or low [281, 312], careful consideration of scale is vital to interpret the results.

### 3.2 Reconstructing the functional trait space

The diffusion map analysis was performed on scientific bottom-trawl survey data available from the International Council for the Exploration of the Sea (ICES) online database [195], specifically on one of the longest time series for North Sea demersal fish species abundance, the ICES coordinated International Bottom Trawl Survey (IBTS). This monitoring program carries annual campaigns since February 1965 [169]. Herring (*Clupea harengus*) was, initially, the only targeted species, having the survey restricted to the southern and central North Sea.

Since 1974 although, the entire North Sea has been included and all demersal fish species are monitored [77]. Therefore, we selected the IBTS surveys carried during the period 1980-2021. Additionally, two younger ICES coordinated surveys were included to the analysis in order to increase the spatial coverage and data availability. The Scottish West Coast Bottom Trawl Survey (SWC-IBTS), with data available from 1985 to 2018 [18]; and The Irish Groundfish Survey (IGFS), with available surveys for the period 2003-2017 [387]. These surveys were also recovered from the ICES DATRAS portal [195]. Our analysis covers the entire North Sea basin, the eastern English Channel, the north and west Scottish Coast, as well as the Irish north coast and the Celtic Sea.

### 3.2.1 Long-term data and area of study

For each survey haul (i.e., sampling unit), the species were identified, the number of individuals were counted, and their length was measured, representing an indirect measure of abundance in number of individuals at size caught per species and haul. Hauls last approximately 30 min and the data was converted into abundance per hour fishing. Only taxonomic groups for which the species names were specified were kept, and scientific names were checked with the World Register of Marine Species [7] as well as with FishBase [137]. All invertebrates and strictly pelagic species were excluded, limiting the analysis to demersal and other bottom related fish species. To standardize for haul duration, only hauls with a minimum duration of 20 min were retained. No restriction was done for depth of hauling. The final dataset included in total 36,467 hauls and 229 species.

To assess the contribution of abiotic and biotic variables to the variation of the reconstructed  $i$ -trait space, we recovered supplementary data. Accordingly, we used some oceanographic parameters that were gathered at the moment of the sampling haul, such as depth of sampling, hence data was available in the DATRAS portal [195]. Furthermore, we used fish ecological traits, gathered by Esther Beukhof and colleagues [36], which are available to retrieve from the PANGEA open source data repository [129]. This data is based on fish taxa observed during international scientific bottom-trawl surveys regularly conducted in the North Sea, Northeast and Northwest Atlantic, and the Northeast Pacific. The majority of trait values were verified via FishBase [137], or supplemented with values from primary literature.

### 3.2.2 Diffusion mapping the functional trait space

We will follow the algorithm introduced in Sec. 2.4.1 for diffusion mapping the fish abundance dataset, using the co-occurrence of species as similarity measure. The starting point for the analysis is the data table of fish species (rows) and haul sampling identifier (columns). When a species is present in a particular sample the entry will be the abundance recorded in the dataset, if the fish species is absent in the sample the table entry is 0. Now, we establish a notion of similarity between our species. Due to the massive disparities between recorded abundances of some fish species, we use the Spearman correlation [385]. The similarity values are now standardized in the range  $[0, 1]$ , i.e., its mean is zero and its standard deviation is 1 [62]. The resulting similarity scores are gathered in a similarity matrix, which is thresholded to keep only the top-10 highest similarity entries for each datapoint. From the thresholded similarity matrix we calculate the row-normalized Laplacian matrix. Finally, we compute the eigenvectors

and eigenvalues of the Laplacian matrix. All the diffusion maps steps were implemented and executed in the Julia programming language [38].

After the diffusion map analysis, the dataset dimensionality has been reduced to fewer, new variables that represent the functional trait space of the fish community. The eigenvectors assign an entry to each fish species for each new variable. Each one of these new variables corresponds to an eigenvalue which encloses information on the importance of the identified variable (Sect. 2.3.1). The most important eigenvector, hereafter called variable 1, corresponds to the smallest non-zero eigenvalue. The eigenvector corresponding to the second-smallest non-zero eigenvalue is the second most important, hereafter called variable 2, and so on [21, 266].

Each new variable represents a possible functional response of the fish community to abiotic or biotic forcing. However, we must make sure the diffusion map analysis captures the main features of compositional change across the community. A fast way to determine if our analysis has identified the main dimensions of variation in the data is to plot the inverse of the ranked non-zero eigenvalues [147]. The inverse eigenvalues should be larger for meaningful dimensions of variation and relatively small for non-meaningful dimensions of variation. In Fig. 3.1, there is a substantial gap between the first non-zero eigenvalue and the remaining eigenvalues, hence, most of the data variation was gathered and can be explained by it.

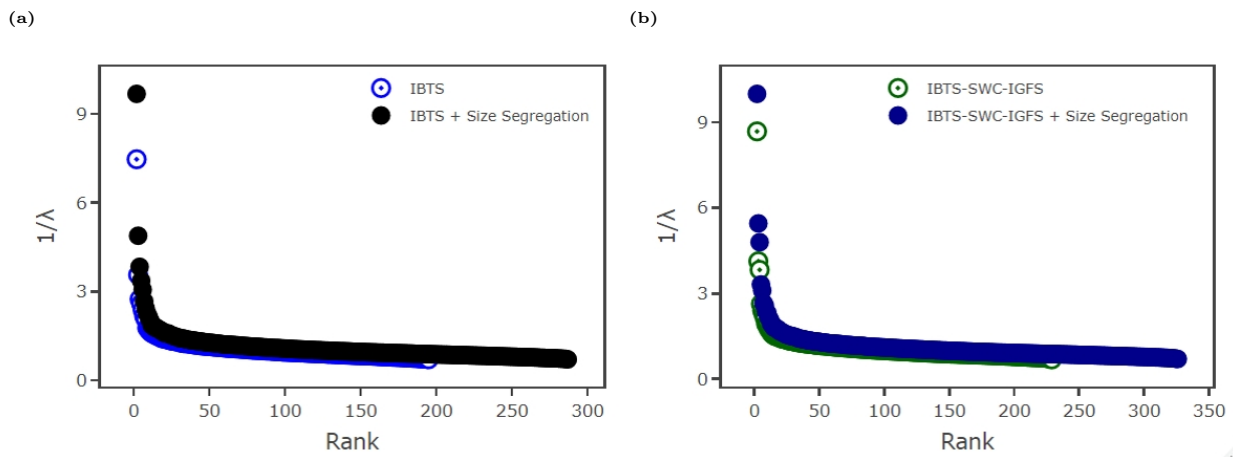


Figure 3.1: Ranking of diffusion maps variables in four scenarios. The ranking score is calculated with the inverse eigenvalue spectrum ( $1/\lambda$ ), across all datapoints of (a) the IBTS, and the IBTS with the adult/juvenile segregation; (b) the aggregation of IBTS, SWC, IGFS, and the aggregation of IBTS, SWC, IGFS with the adult/juvenile segregation. For all scenarios the first non-zero eigenvalue explains most of the variation. However, with adult/juvenile segregation the eigenvalue becomes larger and the second eigenvalue becomes more relevant.

Moreover, in Fig. 3.1 we can also observe two panels that represent different amounts of datapoints used in the analysis. Diffusion maps is a data-hungry method, it delivers best outcomes when more datapoints are used [21, 127]. This can be clearly seen when comparing Fig. 3.1(a), where the analysis was only performed with data from the International Bottom Trawl Survey (IBTS); and Fig. 3.1(b), where the analysis was performed with the aggregation of data from IBTS, the Scottish West Coast Bottom Trawl Survey (SWC-IBTS) and The Irish Groundfish Survey (IGFS). In Fig. 3.1 (b), the relative magnitude of the first inverse non-zero eigenvalue becomes larger, therefore, the magnitude of variation contained in the corresponding dimension is also larger [21, 147].

We also must consider that we are dealing with more complex organisms than in previous ecological applications of diffusion maps [127, 267, 359]. For unicellular organisms, their short

life timespan, which ranges from hours to a few days, in addition to a limited spatial motility, reduces significantly the ecological considerations for the diffusion map analysis. After a parent phytoplankton cell reproduces, there is a high certainty that daughter-cells will occupy the same geographical location [347]. Moreover, different generations will not overlap in time, therefore, reducing significantly the intraspecific competition [86]. For fish species, biology and ecology are more complicated, as explained in Sect. 3.1. Thus, key ecological features which influence the spatial or time distribution of fish species, must be included in the analysis when it is possible.

A feature which might be relevant, is the sometimes dissimilar spatial distribution of adults and juveniles fish species [84, 154, 183]. It has been reported in the scientific literature that most fish species provide little to none parental care, having instead spawning events with million of larvae, which are then indistinctly transported by currents to the open sea, or by tides to bays or shallow waters [9, 248, 339]. In other cases, fish travel to specific spawning locations, quite different and distant from where the adult population live (e.g., Atlantic Salmon (*Salmo salar*) [2, 151], or the European Eel (*Anguilla anguilla*) [366, 435]). There are also cases in which not much movement is performed, thus, the juvenile and adult population tend to have certain degree of overlap [205, 294] Therefore, separating juveniles from adults in our diffusion map analysis will be relevant so that the diffusion distance successfully captures the relationship between environmental distance and compositional dissimilarity in the fish community [147].

A proxy way to differentiate juveniles from adults is by using the concept of size at maturity, which is defined as the length at which 50 % of a population become sexually mature for the first time ( $L_m$ ) [405]. Since the IBTS has gathered information of the length of each collected individual, we will use the  $L_m$  to separate a fish species in a juvenile and an adult version of itself (Fig. 3.2), so all the individuals whose size is smaller than  $L_m$  will be tagged juvenile (J), whereas individuals with a size equal or larger than their reported  $L_m$  will be tagged as adults (A). The values of  $L_m$  were recovered for 130 fish species, available at the dataset by Esther Beukhof and colleagues [36] as well as from FishBase [137].

In panels (a) and (b) of Fig. 3.1 we can appreciate the effectiveness of including this feature. The magnitude of the variable 1 is larger in both cases, when the analysis is performed with size differentiation species data from IBTS only (Fig. 3.1(a)), and when the analysis is performed with size segregated species data from IBTS, SWC-IBTS and IGFS (Fig. 3.1(b)). Moreover, the variable 2 also becomes more relevant (Fig. 3.1(b)), therefore, the two largest non-zero inverse eigenvalues should identify the major axes of variation in the data. The eigenvectors associated with these two eigenvalues provide a new coordinate system for the data, i.e. the diffusion map [79, 147].

### 3.2.3 Identifying functional traits

Once the i-trait space has been reconstructed, and the most relevant variables that explain the variation in the community were identified, we must attribute a meaning to those variables. Diffusion maps do not provide a biological interpretation of the eigenvectors. This might be seem as a disadvantage, as long as there is no intuitive sense on how the eigenvector emerges from the original dataset. But there are coherent structures in the long-term dataset which act as explanatory variables, independent of human interpretation. Those variables can be uncover and interpreted by analyzing additional data [21, 127, 359].



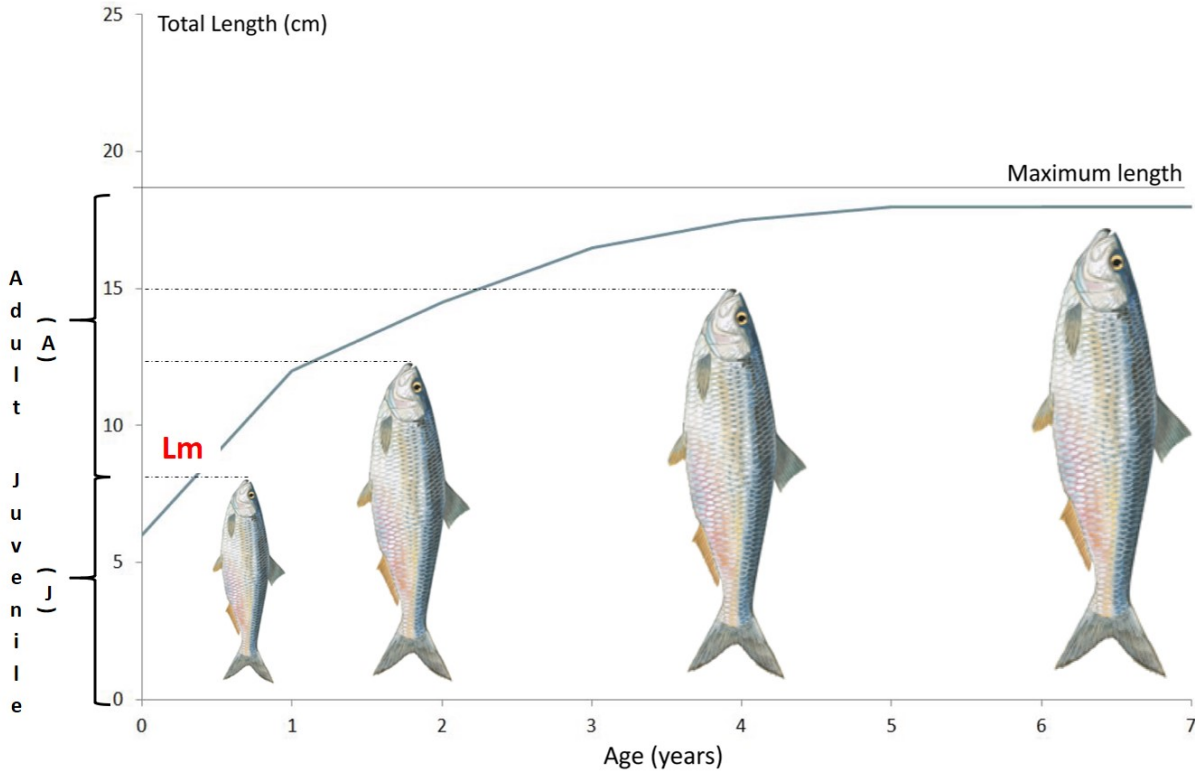


Figure 3.2: Anchovy (*Engraulis encrasicolus*) growth chart representing the association between fish length to age. The chart shows the hypothetical  $L_m$  attributed when fish reaches 8.0 cm or 1 year of age. Every individual whose size is lower than 8.0 cm is assumed a juvenile, whereas an individual whose size is equal or larger than 8.0 cm is considered an adult. Modified from OSPAR Portal; open access under the Creative Commons Attribution License - CC BY 4.0.

An initial approach is to verify the extreme values in the respective variable and check for biological or ecological characteristics of those species [127, 267]. These species might provide a working hypothesis to be tested with additional data, or via scientific literature [21]. However, it is usual that for ecological datasets, information on environmental parameters were also gathered at the moment of sampling (e.g., day of year, sea surface temperature, total  $NO^{-3}$  concentration, total  $PO^{-4}$  concentration, salinity, dissolved inorganic nitrogen (DIN), dissolved inorganic phosphorus (DIP), suspended particles, depth of capture). When this data is available we can calculate the mean environmental conditions at which each species was observed. This is done by computing a weighted average of each environmental parameter, where the biomass of the species under consideration is used as the statistical weight of the sample.

$$\hat{E}^{(r,i)} = \frac{\sum_{j=1}^n a_j^{(i)} E_j^{(r)}}{\sum_{j=1}^n a_j^{(i)}} \quad (3.1)$$

where  $a_j^{(i)}$  is the relative abundance of species  $i$  in sample,  $E_j^{(r)}$  is the environmental factor  $r$  in this sample, and  $n$  is the number of samples. In this way we obtain the species-specific environmental value for each phytoplankton species, then used to colorize the diffusion maps results.

### 3.2.4 Important functional traits in the North Sea fish dataset

Further analysis is performed only in the diffusion map that provides the best possible reconstructed trait space, as explained in Sec. 3.2.2. This diffusion map corresponds to the analysis performed on the merged fish dataset from IBTS, SWC-IBTS and IGFS, with a juvenile/adult differentiation for fish species with an available size at maturity value ( $L_m$ ). The environmental data gathered during sampling was either incomplete for the entire period of monitoring, or was not available for most of listed parameters in previous Sec. 3.2.3. Only the depth of hauling was consistently reported, as well as sampling date. Fish traits [36], both categorical and quantitative, were also used for assessing the variables of the diffusion map [36].

#### First variable: Depth

The first i-trait identified by the diffusion map, or EV1 in Fig. 3.3, aligns datapoints along a manifold which represents the depth in which the fish species was more abundant. In the positive extreme of EV1, we find fish more abundant at shallow areas, not deeper than 50 meters, e.g., *Gasterosteus aculeatus*, *Syngnathus rostellatus*, *Platichthys flesus*, juveniles of *Clupea harengus*. Whereas in the negative extreme we find fish that reached abyssal areas, up to depths of around 400-500 meters, e.g., *Polymetme corythaeola*, *Notacanthus bonaparte*, *Synphobranchus kaupii*, *Mora moro*.

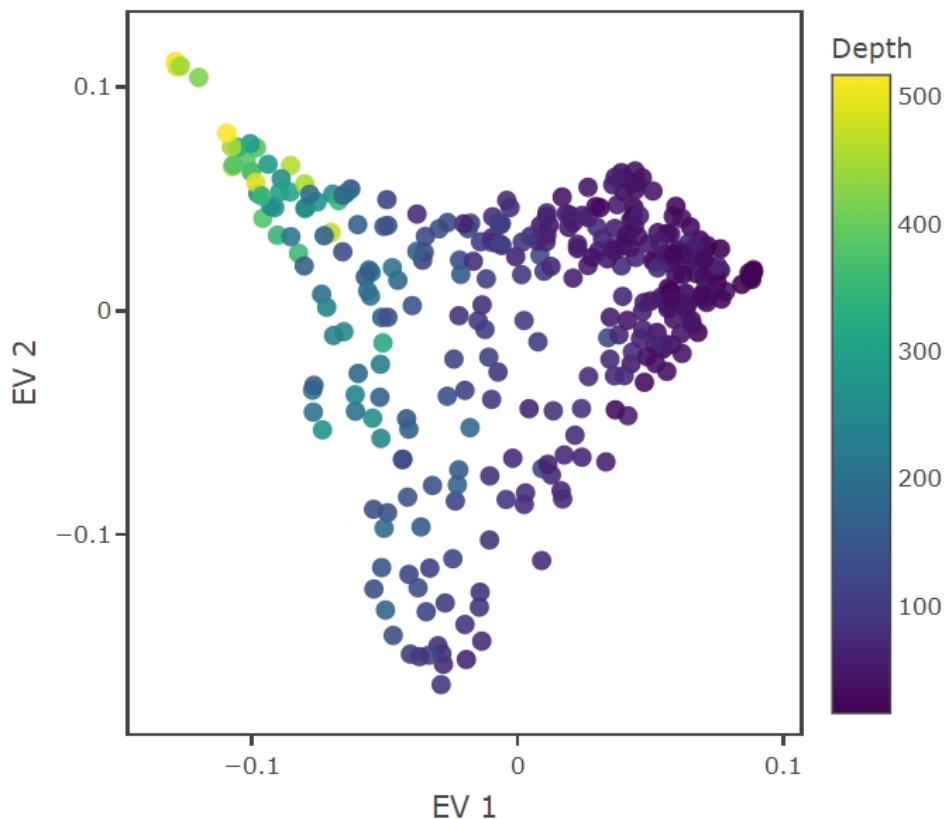


Figure 3.3: Diffusion map displays the reconstructed i-traits that explained best the variation of dataset. Color coded is the environmental conditions under which the species were observed with high relative abundance. The EV1 aligns well with depth of sampling. The manifold alignment might represent the response of fish species to the depth gradient of the North Sea basin and a sign of adaptation to the oceanographic conditions that are characteristic of different depths, i.e., salinity, water temperature, food availability, light availability.

That depth plays a key role in structuring the North Sea fish community is not a surprise [36, 90, 101]. Diverse fish communities, at different latitudes and habitats, have strong responses to changes in depth. Fish communities usually get replaced partially or totally for others. For instance, species in coastal shallow waters will resemble little to none those inhabiting the twilight zone or the abyssal planes [155, 334]. However, what might result counter-intuitive is that the reconstruction of the functional trait-space shows depth as the most important trait in the fish community, when depth itself is not a functional trait which can be attribute to any fish species. Therefore, we must examine closer, using further knowledge about the North Sea oceanographic conditions, so that we can interpret the diffusion map axis in an ecologically meaningful manner [21].

The North Sea, including the Skagerak, has a surface area of 575,000 km<sup>2</sup> [219]. A massive extension which nevertheless, is an exceptionally well-studied basin [301]. An example of what in oceanography is known as a shelf sea. A coastal system at the interface of land and sea, driven by weather and climate patterns that either promote inflow of oceanic waters from the ocean, or dominance of terrestrial runoff [124, 228]. This binary influence makes the North Sea a two-part system [124, 193]. In one hand, the southern portion is a shallow sea, with depths bellow 50 meters (Fig. 3.4), a well-mixed water column and high nutrient concentrations, hence, high seasonal phytoplankton production (Fig. 3.5). On the other hand, the central and northern section have average depths of 100 meters (Fig. 3.4), a seasonally stratified water column but nutrient-limited surface mix-layer, hence, with low primary production [124, 193] (Fig. 3.5).

The oceanographic differences between the north-central and southern North Sea ecosystem will influence the possibilities of fish species to thrive in either part of the system [36, 64, 346]. In this sense, depth becomes a summary variable of the oceanographic conditions that are affecting fish. What the first variable of diffusion map is capturing, is a complex nonlinear combination of environmental parameters which co-vary with depth (e.g. temperature, seasonality, salinity and primary productivity), and that are a direct derivation of the North Sea two-part system [116, 124, 193]. From a functional trait perspective, the diffusion map first axis could represent the physiological adaptation of fish to water temperature and salinity [122, 256, 321], the fish larval survival rate [189, 209], the adaptability of fish to seasonal extremes [16, 117], diet preferences [58, 139], or even behavioral responses to oceanographic conditions [229].

Applications of diffusion map to ecology have not used, so far, any extra ecological consideration. In part because they deal with unicellular organisms [127, 267, 359] with a simpler life cycle, or because relevant ecological data was not available. For diffusion mapping a fish community dataset, the incorporation of size at maturity ( $L_m$ ) is an asset which, nonetheless, is not excepted from certain assumptions. It is worth noticing that neither all fish species show adults/juveniles habitat segregation nor it is possible to know the value of  $L_m$  for all fish species displaying such behaviour. Moreover, size at maturity ( $L_m$ ) has been recently considered a non-fixed parameter, i.e., a parameter whose value could vary temporally, or between different fish populations [309, 368]. This observation is most relevant for species holding commercial value, whose populations have experienced significant size reduction [357], or have been subject to habitat distribution reduction [107]. In spite of these considerations, the differentiation of fish species between juvenile and adults, improved the explanatory power of variable 1 (see Sect. 3.2.2).

The diffusion map captured correctly differences in preferred depth for juveniles or adults. For instance in Fig. 3.6(a) and Fig. 3.6(b), it can be noticed how adults of herring (*Clupea harengus*) and whiting (*Merlangius merlangus*) were more abundant on depths below 80 meters,



Figure 3.4: Bathymetry and topography of the North Sea displaying some important regions. Red dashed line represents a country Exclusive Economic Zone (EEZ). Prominent regions are the Norwegian trench with depths below 200 meters, the Dogger Bank, a sand bank located in the central basin and the southern shallow sea with depth below 50 meters. The International Bottom Trawl Survey (IBTS) is carried out over the entire basin. Reproduced from Martinez-Gordon and colleagues [263]; open access under the Creative Commons Attribution License - CC BY 4.0.

whereas juveniles were so on depths between 40 and 60 meters. This coincides with reported herring spawning grounds located on the eastern English coast up to the Strait of Dover, while alevines are transported by residual currents towards nursery areas in southern North Sea, particularly the German bay [33, 84, 196]. The whiting showed a similar behaviour, however with a smaller depth range between juveniles and adults (approximately 25 meters). This could be explained by Whiting population dynamics. Two main spawning grounds have been reported, one in the south and western North Sea and the other in the northern North Sea [154, 182], while juveniles and old adults are predominantly found in offshore waters [23, 171]. Thus, while recently settled whiting occur extensively throughout coastal waters, juvenile whiting tend to migrate off-shore, returning to shallower waters to spawn [71, 154].

In Fig. 3.6(c) and Fig. 3.6(d) there are two examples of milder spatial separation. In one hand, the juveniles of the european river lamprey (*Lampetra fluviatilis*), showed preference for depths between 20-30 meters, whereas few adults were found up to 70 meters depth. This coincide with the extensive literature about the lamprey anadromous life cycle [106, 268, 377], and the reports of metamorphosed ammocoetes in southern North Sea and Skagerak shallow

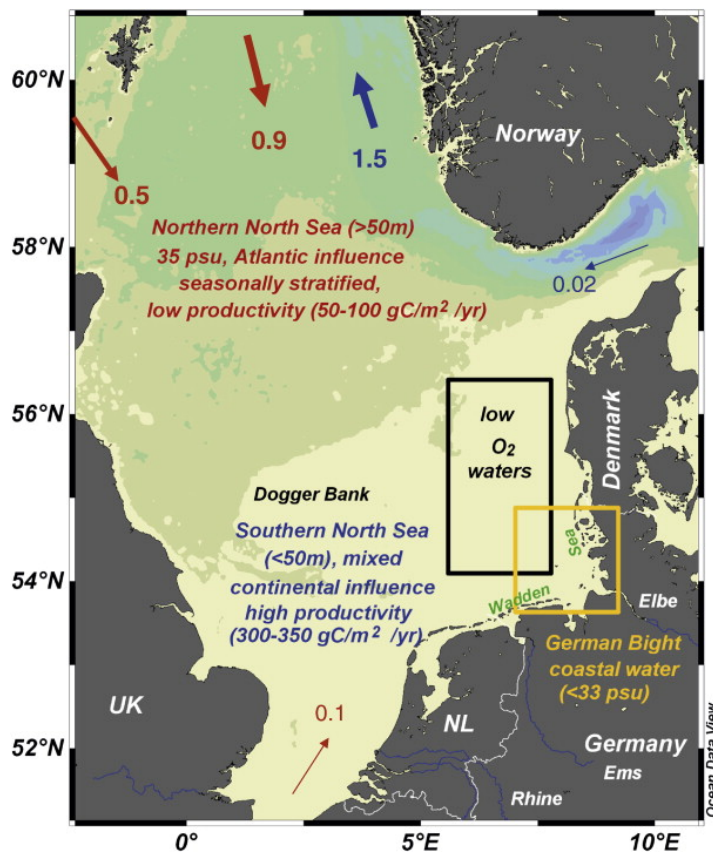


Figure 3.5: North Sea two-part system. The southern area is shallow, and highly influenced by continental runoff; thus, the nutrient rich waters attain high seasonal productivity. The northern and central areas, are highly influenced by Atlantic water inflow. Average depth is 100 meters, the water column gets seasonally stratified, and the surface mixed-layer is nutrients-limited, hence, with low productivity. Arrows with numbers denote average water mass transports (Sv). Reproduced from Emeis and colleagues [124]; open access under the Creative Commons Attribution License - CC BY 4.0.

waters and river estuaries [4, 121, 377]. The long rough dab (*Hippoglossoides platessoides*), an inhabitant of deeper waters of the northern North Sea, has minor reports on spatial segregation or spatial movements towards spanning grounds [294, 423]. The habitat segregation strategy results advantageous in some cases, because it reduces intra-specific competition, or cannibalism [89, 183, 318]. But the high levels of energy necessary for massive spawning, or large migrations between feeding and spawning grounds, leaves little or none energy to adult fish to provide parental care, thus, natural mortality of fish is extremely high during the first years of life [270, 374]. Other fish species did not evolve such behavior, hence having adults and juveniles occupying indistinctly the same habitat, e.g. the dab (*Limanda limanda*) and the flounder (*Platichthys flesus*) [43, 65] (Fig. 3.6(e) and Fig. 3.6(f)).

## Second variable: Year of maximum abundance

The second i-trait identified by the diffusion maps, or EV2 as shown in Fig. 3.7, aligns datapoints along a manifold which represents the year in which the fish species was present in higher abundance. The manifold shows asymmetry between its positive and negative extreme. While the manifold is wider in relation to the negative side of EV1, it is narrower in relation to the positive side of EV1 (Fig. 3.7). The gap between the inverse eigenvalue ( $1/\lambda$ ) of EV1 and EV2 (Fig. 3.1) suggested that EV2 gathers less of the variation in the fish community. However, this might also be a response of the system itself rather than a less satisfactory result. When plotting separately the species which were more abundant during the period 1980-1999 (Fig. 3.8(a)) and those more abundant between 2010-2021 (Fig. 3.8(b)), we get a hint of what this diffusion map variable might represent.

A first impression would suggest that at the deeper northern North Sea, i.e., datapoints at

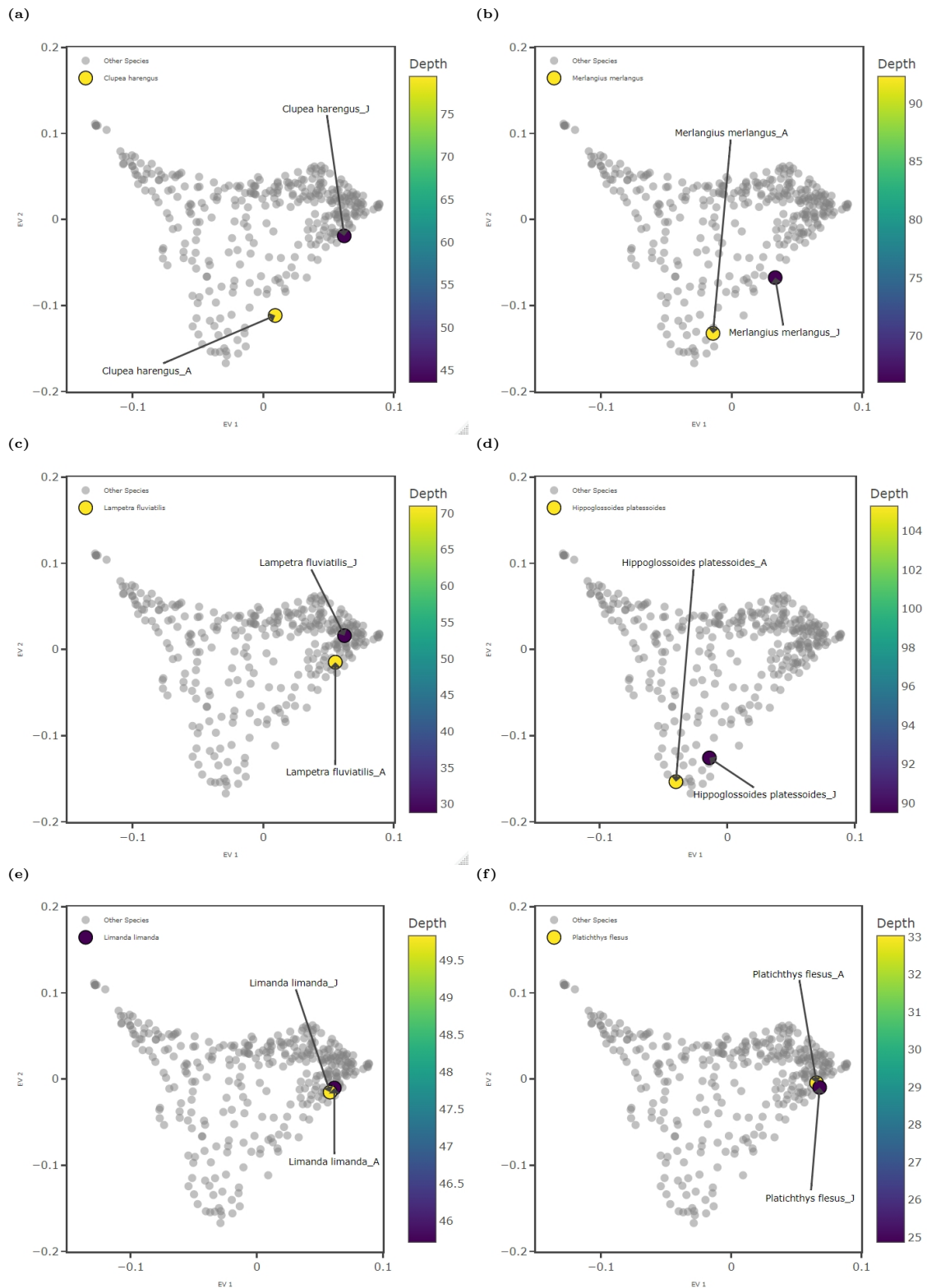


Figure 3.6: Diffusion map of the reconstructed i-trait space, color-coded according to depth of higher abundance. Panels highlight some examples of fish species separated in adults (A) and juveniles (J) according to their size at maturity ( $L_m$ ). Panels (a) and (b) show fish with marked spatial segregation (*Clupea harengus* and *Merlangius merlangus*); panels (c) and (d) fish with milder spatial segregation (*Lampetra fluviatilis* and *Hippoglossoides platessoides*); and panels (e) and (f) fish which did not show any spatial segregation (*Limanda limanda* and *Platichthys flesus*).

EV1 negative side, species reaching abundance peak in the period 1980-1990 are very distant from those reaching their abundance peak at the period 2010-2021. On the contrary, in the shallower southern North Sea and Shagerrak-Kattegat, i.e., datapoints at EV1 positive side, the species dominant for each period are quiet similar, thus, they locate close in the man-

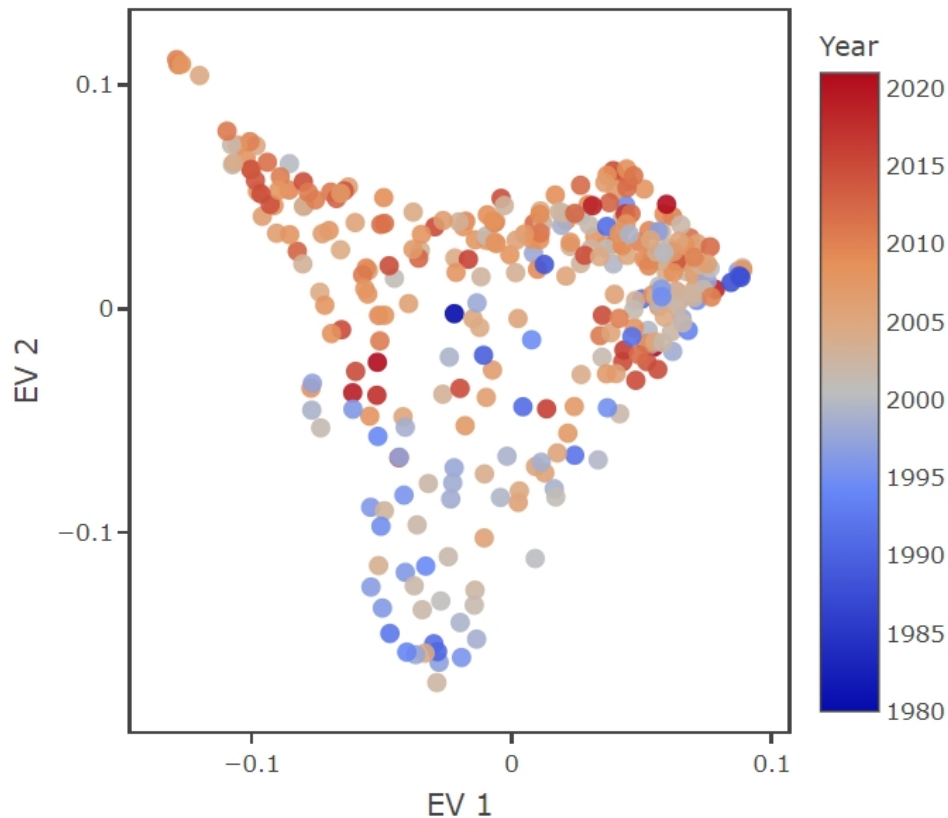


Figure 3.7: Diffusion map displays the reconstructed i-traits that explained best the variation of dataset. Color coded is the factor under which the species were observed with high relative abundance. The EV2 aligns with year in which the species abundance was highest. The manifold alignment might represent the response of fish traits to temporal variations in environmental trends, population dynamic oscillations or the outcome of human exploitation in the North Sea basin.

ifold (Fig. 3.8). This might be a result of distinct responses of fish species inhabiting the North Sea two-part system [124, 193]. For instance, the shallower southern North Sea displaying a dominance shift between species which tend to coexist, either due to distinct dynamics in their population abundances or due to favorable environmental conditions. In the deeper North-Central North Sea however, such coexistence fades away, giving room to a likely shift in dominance by species turnover. To discern this, at least partially, we can take a closer look to those species which dominate in each period and region of the North Sea.

In one hand, those species that were more abundant in the period 1980-1999 have either almost disappeared, or suffered severe decreasing abundance trends in the forward decades (Fig. 3.9, Fig. 3.10 and Fig. 3.11). In the northern and central North Sea (Areas 1, 2, 3 or 7), the former includes adults of *Hippoglossoides platessoides*, *Anarhichas lupus*, *Sebastes viviparus*, *Pollachius pollachius*, *Molva molva*, and juveniles of *Anarhichas lupus* (Fig. 3.9). The latter includes juveniles of *Amblyraja radiata*, *Gadus morhua*, *Melanogrammus aeglefinus*, *Merlangius merlangus* and adults of *Gadus morhua*, *Microstomus kitt* (Fig. 3.10). In Southern North Sea (Areas 4, 5, or 6), the Shagerrak (Area 8) or the Kattegat (Area 9), fish species experienced mostly decreasing abundance trends in subsequent decades, e.g., juveniles and adults of *Gadus morhua*; adults of *Myoxocephalus scorpius* and *Enchelyopus cimbrius*; *Lycodes vahlii* and *Lycenchelys sarsii* (Fig. 3.11).

On the other hand, fish species that were more abundant in the period of 2010-2021 have experienced diverse trends in the North Sea. For instance, in the northern and central North Sea (Areas 1, 2, 3 or 7), adults and juveniles of *Merluccius merluccius* and *Pleuronectes platessa*;

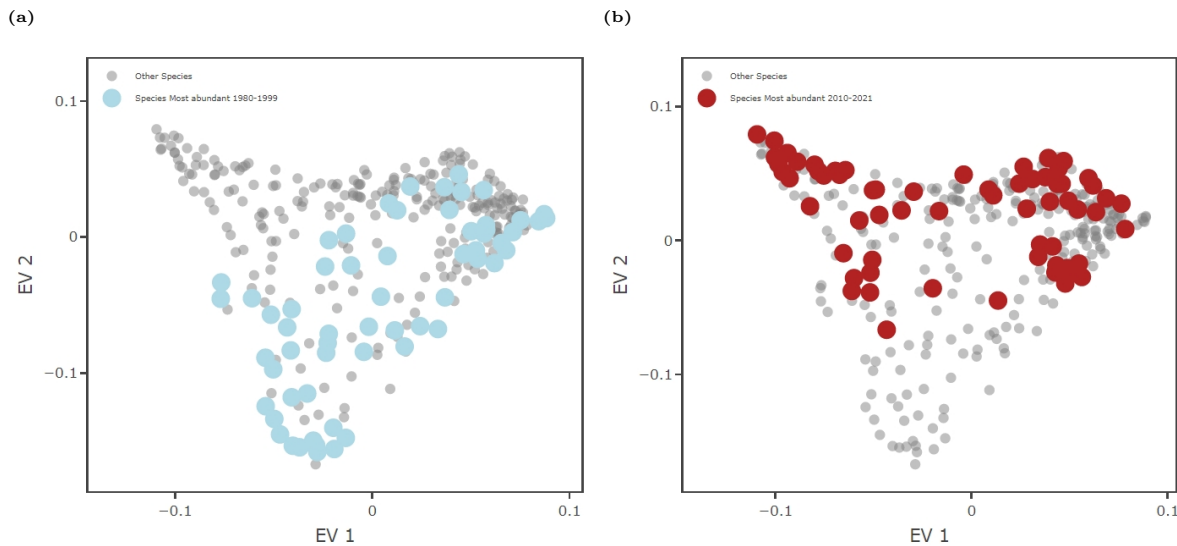


Figure 3.8: Diffusion map displays the reconstructed i-traits that explained best the variation of dataset. Color coded is the decade in which the fish had highest abundance peak. Panel (a) displays species which reached highest abundance in the period 1980-1999. Panel (b) displays species which reached highest abundance in the period 2010-2021. The asymmetry between the positive and negative side of EV2 is related to the different dynamics of fish inhabiting the deeper northern and the shallower southern North Sea.

juveniles of *Microstomus kitt* and adults of *Scyliorhinus canicula*, have experienced a dramatic increasing abundance trend, especially after mid-2000s (Fig. 3.12). A moderate increasing, or stationary, trend in abundance could be seen in species at the southern North Sea (Areas 4, 5, or 6), the Shagerrak (Area 8) or the Kattegat (Area 9), e.g., juveniles and adults of *Pleuronectes platessa*; juveniles of *Merlangius merlangus*, *Solea solea*, *Microstomus kitt* and *Eutrigla gurnardus* (Fig. 3.13). Few other species were only recorded from mid 2000s in Area 1 or Area 3, and some areas of the Irish Sea, reaching peaks of abundance during the 2010s decade, e.g., *Phycis blennoides*; adults of *Capros aper*; adults and juveniles of *Micromesistius poutassou* (Fig. 3.14).

The geographical position of the North Sea, makes its fish community highly connected to the North Atlantic fauna through the entrance between Scotland and Norway, with the Baltic Sea fauna through Shagerrak and Kattegat, and with the Celtic Sea/ Bay of Viscay fauna through the English Channel [118]. Yan Jimming, based on the work of Wheeler [429], considered that the North Sea fish fauna was composed of three main groups [202]. The ‘Boreal fish’, with distribution centered north of the English Channel; the ‘Lusitanian fish’, with distribution centered south of the English Channel; and the ‘Atlantic fish’, widely distributed in the Atlantic Ocean.

The variety of fish species living together in the North Sea is a clear indication of the pronounced differences in adaptation that have allowed them to coexist [90]. However, communities are dynamic and shifts in dominance are rather a rule than an exception in ecosystems. The initial reported dominance of the Boreal fish fauna during most of the 70 and 80s decades [170, 202], is giving room to an ever more documented shift towards a dominance of Lusitanian fish fauna in the North Sea, especially since the beginning of the new century [25, 113, 118, 315].

The EV2 is likely detecting this shift from Boreal towards Lusitanian fish species in either the northern or the southern areas of the North Sea. Boreal fish such as the starry ray (*Amblyraja radiata*), the four-bearded rockling (*Enchelyopus cimbrius*), the bullrout (*Myoxocephalus scorpius*), the cod (*Gadus morhua*) and the haddock (*Melanogrammus aeglefinus*)



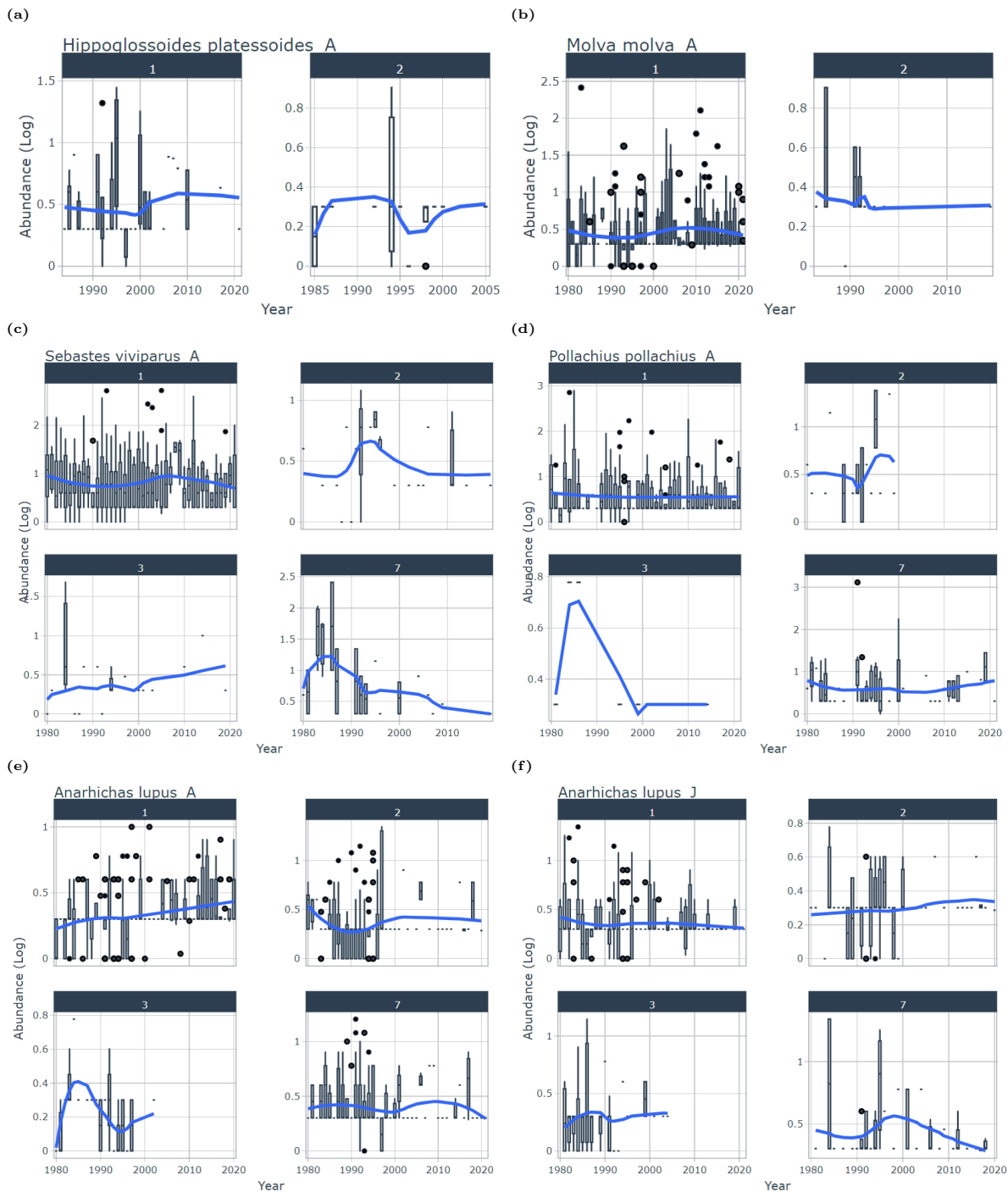


Figure 3.9: Box-plot abundance time series of ‘Boreal fish’ which disappeared from areas of the northern and central North Sea (Areas 1, 2, 3 and 7), including adults of Long rough dab (*Hippoglossoides platessoides*), Ling (*Molva molva*), Norway haddock (*Sebastes viviparus*), Pollack (*Pollachius pollachius*); adults and juveniles of Wolffish (*Anarhichas lupus*). Abundances were yearly Log transformed. The blue line represents the time series trend.

showed a decreasing abundance trend since the beginning of 2000s (Fig. 3.10 and Fig. 3.11), while Lusitanian fish, e.g., the hake (*Merluccius merluccius*), the whiting (*Merlangius merlangus*), the grey gurnard (*Eutrigla gurnardus*), or the dogfish (*Scyliorhinus canicula*), reached abundance peaks during the same period (Fig. 3.12 and Fig. 3.13).

The asymmetry in the EV2 might be related to the lost of some boreal fish species, such as e.g., wolffish (*Anarhichas lupus*), the Norway haddock (*Sebastes viviparus*), the long rough dab (*Hippoglossoides platessoides*) and the ling (*Molva molva*), from northern North Sea areas where they used to be abundant previous to the 2000s (Fig. 3.9). Furthermore, Lusitanian fish,

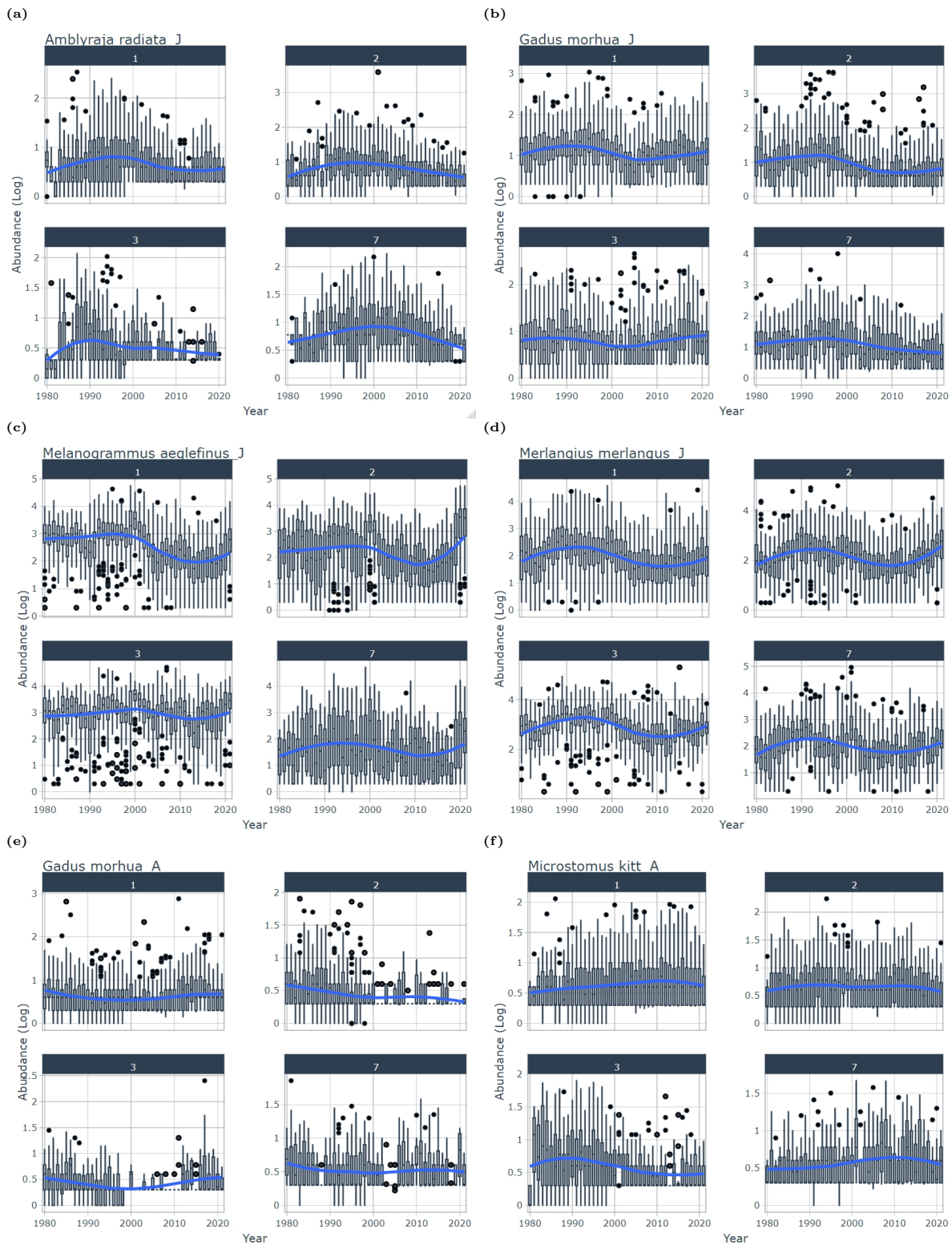


Figure 3.10: Box-plot abundance time series of 'Boreal fish' with decreasing abundance trends in areas of the northern and central North Sea (Areas 1, 2, 3 and 7), including juveniles of Starry ray (*Amblyraja radiata*), Cod (*Gadus morhua*), Haddock (*Melanogrammus aeglefinus*), Whiting (*Merlangius merlangus*); and adults of Cod (*Gadus morhua*), and Lemon sole (*Microstomus kitt*). Abundances were yearly Log transformed. The blue line represents the time series trend.

such as the fork beard (*Phycis blennoides*), the Boar-fish (*Capros aper*), or the Atlantic Blue whiting (*Micromesistius poutassou*), are registered first in the Irish Sea areas and Area 1 and 3 of the northern North Sea, mostly from mid-2000s (Fig. 3.14).

The apparent local extinctions in the northern areas and subsequent fish migrations into the

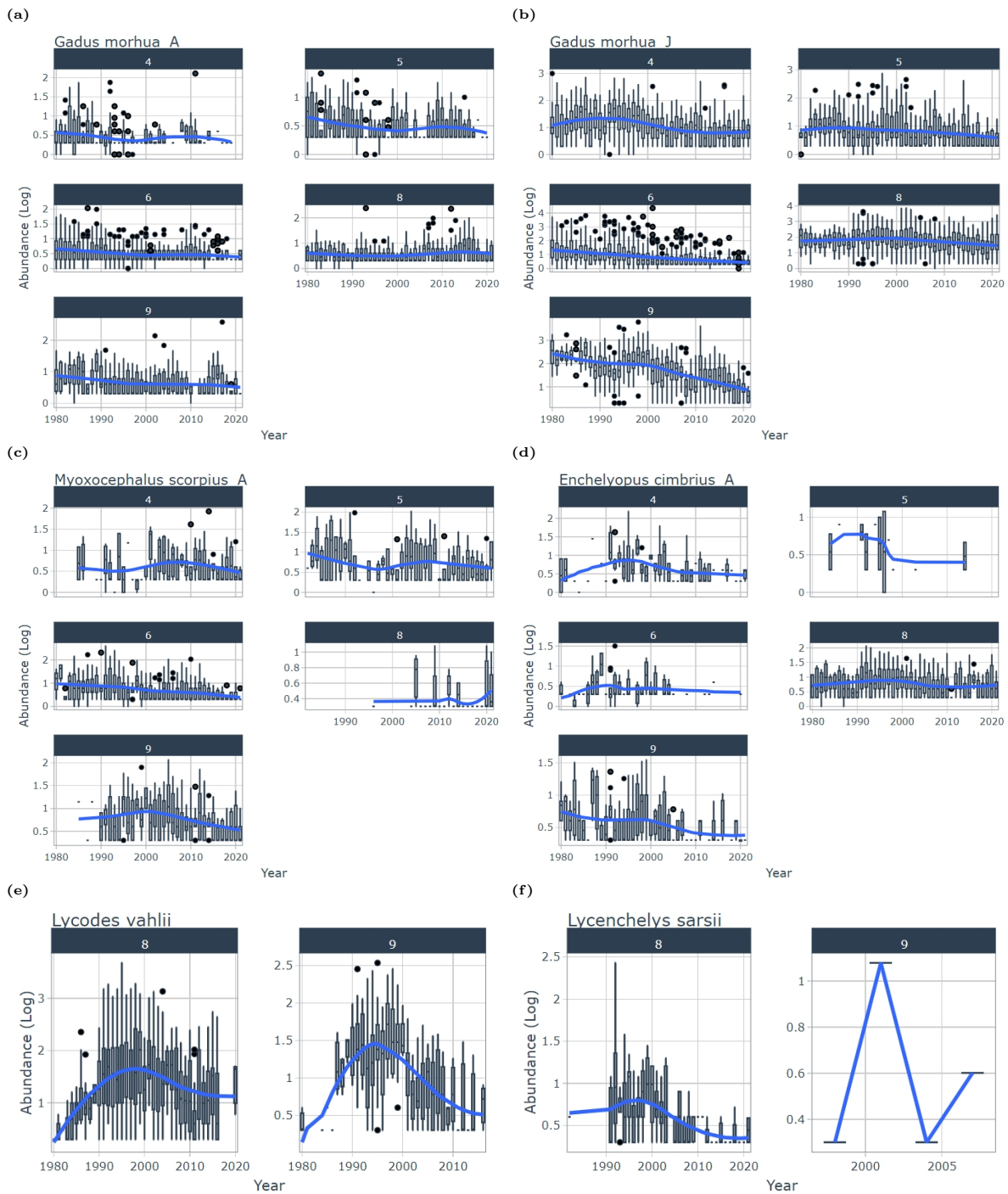


Figure 3.11: Box-plot abundance time series of ‘Boreal fish’ with decreasing abundance trends in areas of the southern North Sea (Areas 4, 5, or 6), the Shagerrak (Area 8), and the Kattegat (Area 9), including juveniles and adults of Cod (*Gadus morhua*); adults of Bullrout (*Myoxocephalus scorpius*) and Four-bearded rockling (*Enchelyopus cimbricus*); Vahl’s eelpout (*Lycodes vahlii*) and Sars’s eelpout (*Lycenchelys sarsii*). Abundances were yearly Log transformed. The blue line represents the time series trend.

North Sea have enhanced the dissimilarities of fish species in the deeper northern rather than in the shallower southern areas. Indeed, relatively warm winter temperatures in the northwestern North Sea might result in stronger inflow of warmer North Atlantic Current waters [116, 324]. Consequently, the main route into the North Sea for southern warm-tolerant species could be stronger via the Shetland–Orkney gap rather than via the English Channel. Besides the species previously mentioned, other warm-water species, such as the john dory (*Zeus faber*) and the snake pipefish (*Entelurus aequoreus*) which also invaded in the beginning of the 2000s, were first reported in the northwestern North Sea before expanding southward [118, 164, 216].

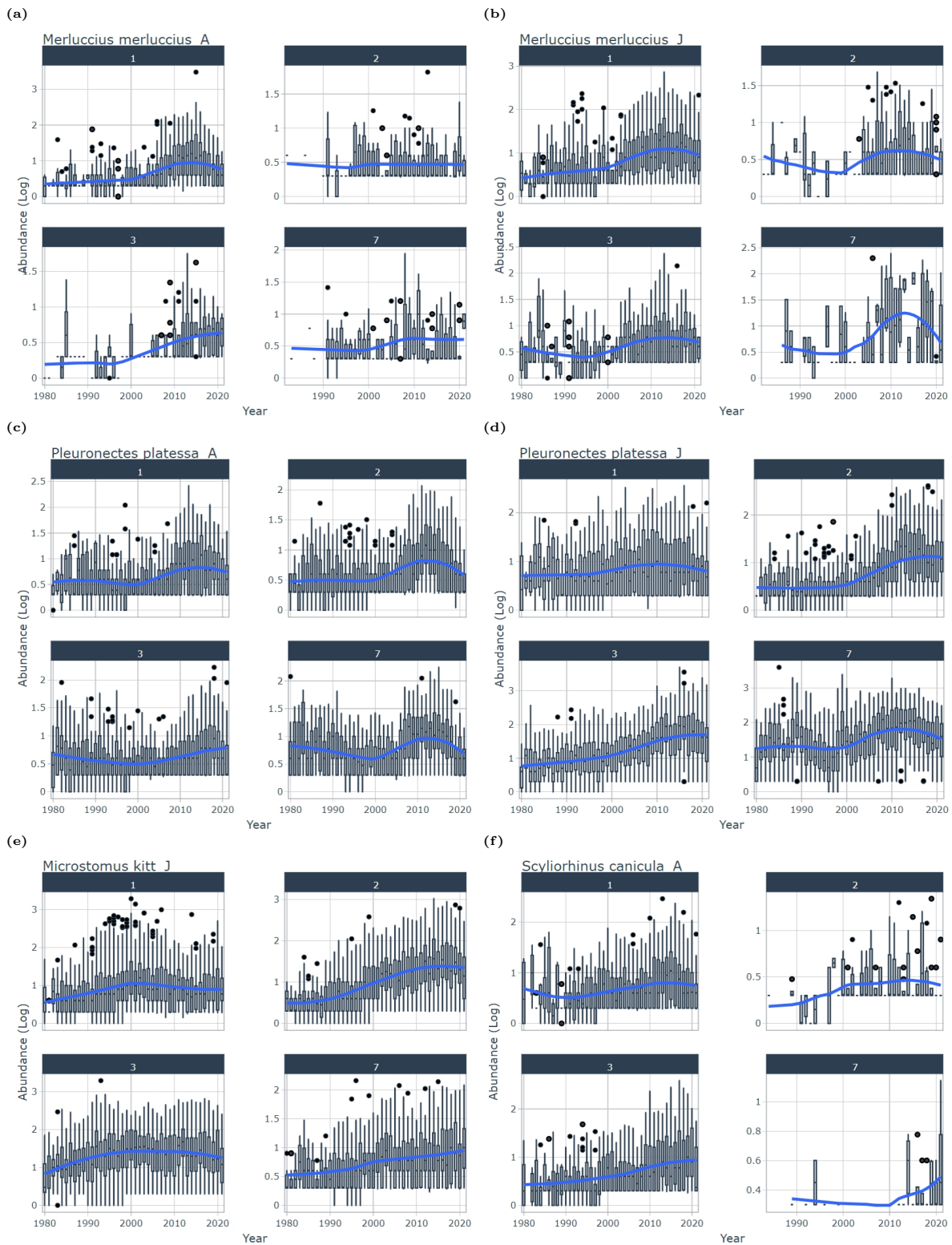


Figure 3.12: Box-plot abundance time series of 'Lusitanian fish' with increasing abundance trends in areas of the northern and central North Sea (Areas 1, 2, 3 and 7), including juveniles and adults of Hake (*Merluccius merluccius*) and Plaice (*Pleuronectes platessa*); juveniles of Lemon sole (*Microstomus kitt*); and adults of Dogfish (*Scyliorhinus canicula*). Abundances were yearly Log transformed. The blue line represents the time series trend.

Nevertheless the examples given so far, some exceptions are present as well. Boreal fish such as the dab (*Limanda limanda*) and the lemon sole (*Microstomus kitt*) had an increasing abundance trend indistinctly of which North Sea area we observe. In other cases, such as the coalfish (*Pollachius virens*) and the greater argentine (*Argentina silus*), a positive abundance trend was maintained only in northern areas, the Shagerrak and the Kattegat, being totally

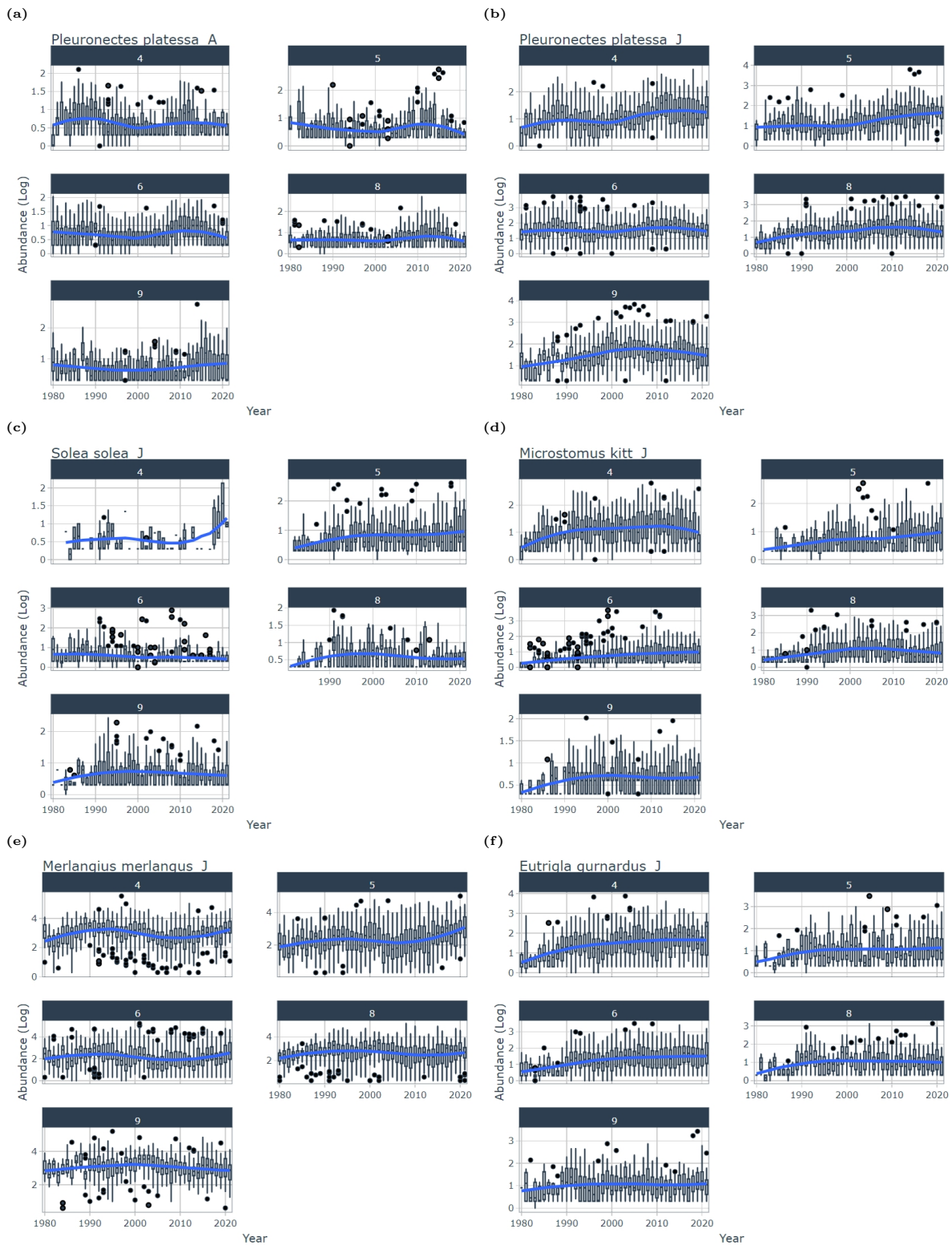


Figure 3.13: Box-plot abundance time series of ‘Lusitanian fish’ with increasing abundance trends in areas of the southern North Sea (Areas 4, 5, or 6), the Shagerrak (Area 8), and the Kattegat (Area 9), including juveniles and adults of Plaice (*Pleuronectes platessa*); juveniles of Sole (*Solea solea*), Lemon sole (*Microstomus kitt*), Whiting (*Merlangius merlangus*) and Grey gurnard (*Eutrigla gurnardus*). Abundances were yearly Log transformed. The blue line represents the time series trend.

absent or barely reported in the shallower southern areas. Few other fish have a relative stationary trend, with oscillatory tendency over periods of 5 to 10 years, e.g., the herring (*Clupea harengus*) and the Norway pout (*Trisopterus esmarkii*). In the case of Lusitanian fish, the few exceptions are related to species which did not have wider expansion, thus, being mostly limited to southern areas, e.g., the anchovy (*Engraulis encrasicolus*), the Pilchard (*Sardina pilchardus*),

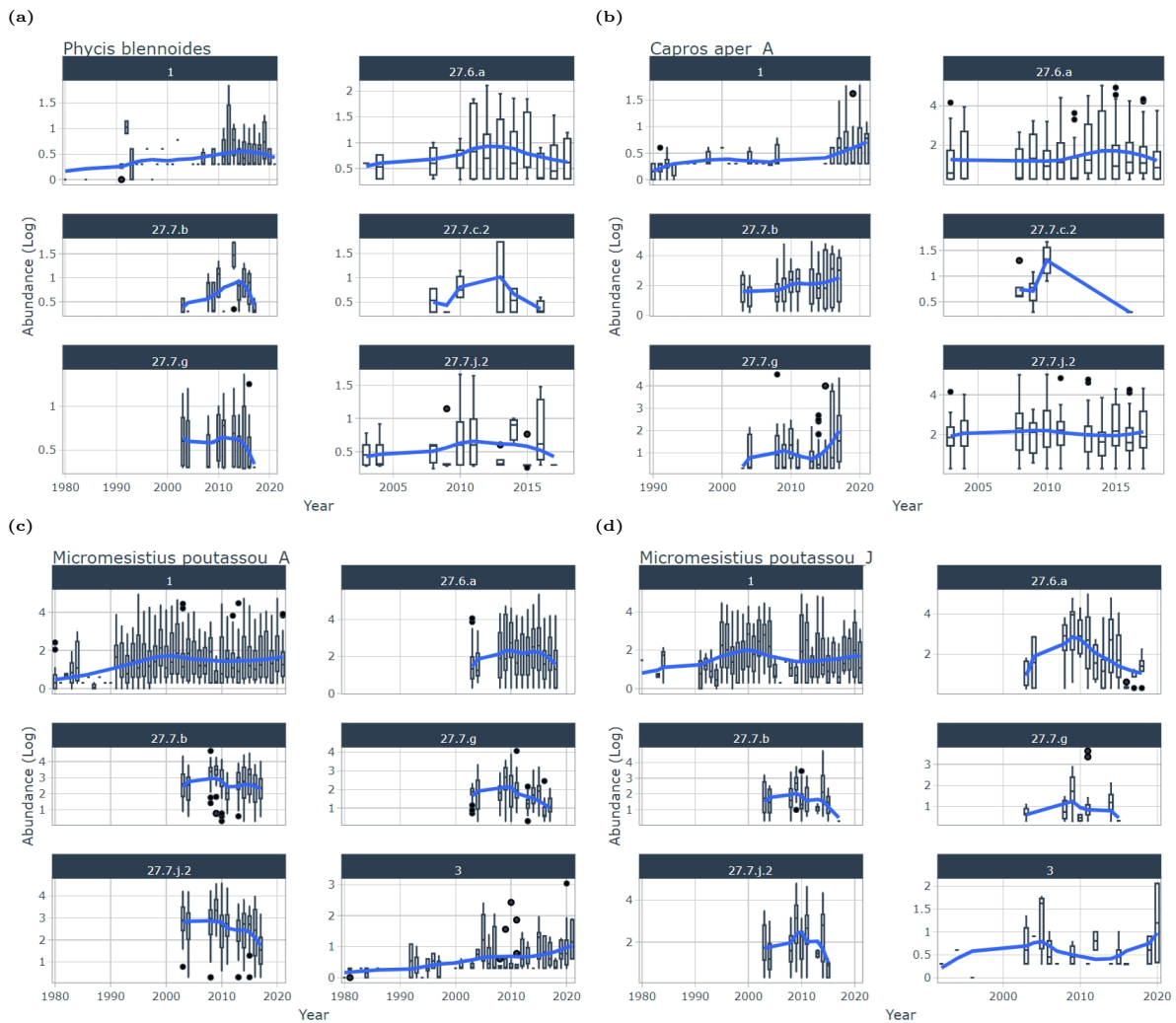


Figure 3.14: Box-plot abundance time series of ‘Lusitanian fish’ which appeared in Area 1 or 3 of the northern North Sea and the Irish Sea (Area 27.6.a, 27.7.b, 27.7.c.2, 27.7.g, 27.7.j.2), including Fork beard (*Phycis blennoides*); adults of Boar-fish (*Capros aper*); adults and juveniles of Blue whiting (*Micromesistius poutassou*). Abundances were yearly Log transformed. The blue line represents the time series trend.

the john dory (*Zeus faber*) and the roker (*Raja clavata*).

Any attempt at grouping tends to be highly artificial, thus, when trying to generalise, exceptions will certainly emerge [90]. The main difficulty in moving from species specific information to generalised insights about the fish community resides not just in the 40 years long-term dataset, but in the vast spatial scale over which such data was gathered [237]. For the North Sea fish community the geographical distinctiveness between north-central and southern North Sea is the major driver of community organization, summarized by variable 1 of diffusion map, i.e., depth of species higher abundance. This influence is as strong as to affect the temporal community changes in different manners according to region under assessment. Precisely, temporal compositional changes, an apparent species turnover or community shift, is detected by variable 2 of the diffusion map, i.e., the year of species higher abundance. Now, a key question is whether the individual responses of species are context-specific phenomena or whether they are symptomatic of a change in the North Sea ecosystem resulting from deterministic or stochastic processes [338]. And if such an ecosystem-scale change has been taking place, how the functional diversity was affected by it?

### 3.3 Functional Diversity

In Sect. 2.5 we have seen that the diffusion distances in the reconstructed i-trait space can be used to robustly estimate functional diversity of ecological communities [267, 359]. Adopting the approach of Ryabov and colleagues [359], functional distinctiveness between all pairs of species can be calculated (Eq. 2.15) from the variables obtained via diffusion maps. These distances are then used to quantify the functional diversity of each sample (Eq. 2.16), by means of the Rao's quadratic entropy [342]. The outcome obtained is a time series with a functional diversity score per sample.

The assessment of functional diversity variation was done in both the temporal and spatial scale. Summarizing all sampling hauls according to area and season, the functional diversity displays different variations over the yearly cycle regarding the area under observation (Fig. 3.15). The northern and central North Sea (Areas 1, 2, 3 and 7) have wider variation in functional diversity, reaching the highest values in winter or spring, and the lowest values mostly in summer. The southern North Sea (Areas 4, 5, or 6), generally displayed narrower variability, except for the winter period in which both highest and lowest values of functional diversity were present. The Shagerrak (Area 8) and the Kattegat (Area 9) also showed wide variability, with winter and summer scoring the highest values of functional diversity, whereas lowest values are usually found in summer or autumn (Fig. 3.15).

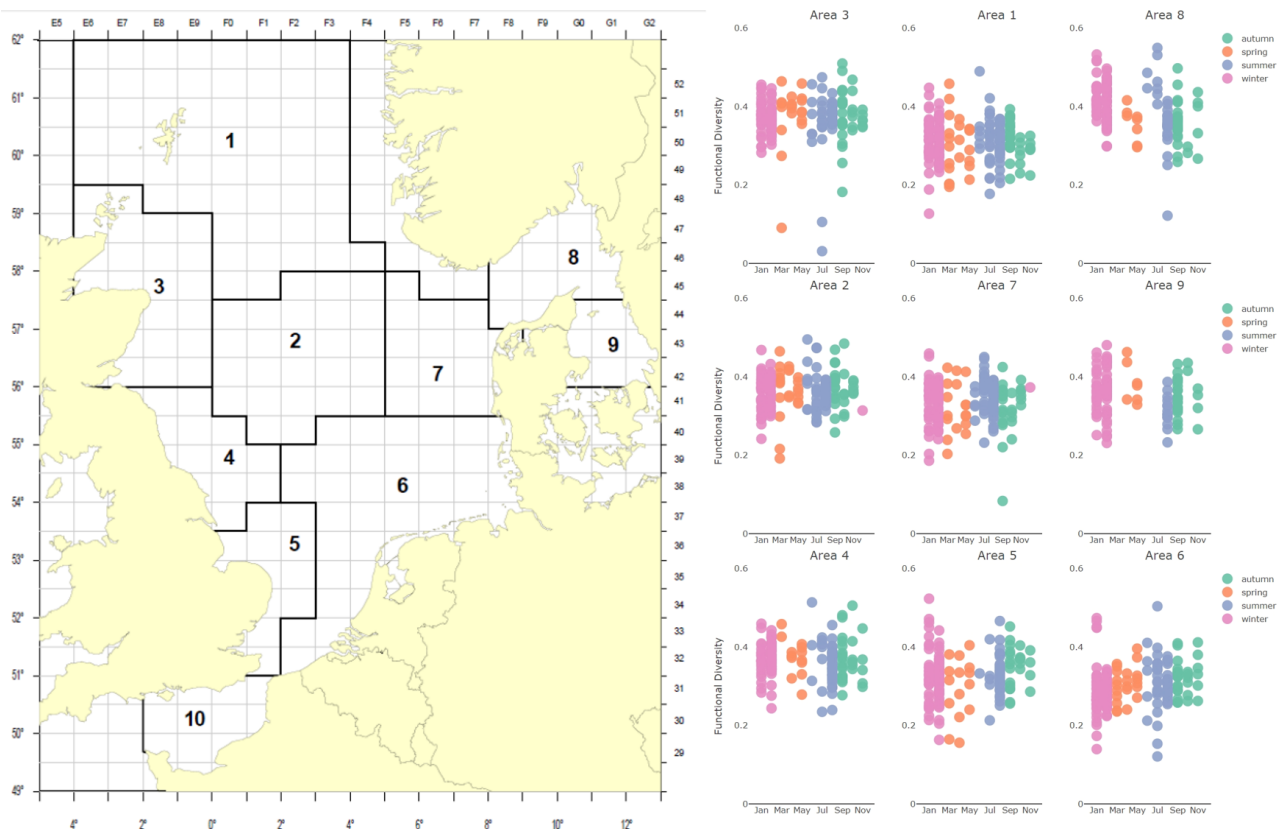


Figure 3.15: North Sea functional diversity variation according to area and season. While the northern and central North Sea (Areas 1, 2, 3 and 7) displays wider seasonal variation, the southern North Sea (Areas 4, 5, or 6) generally displayed narrower variability in functional diversity. Highest functional diversity values are usually found in winter in the north and in summer or autumn in the south. The Shagerrak (Area 8) and the Kattegat (Area 9) also showed wider variability but with highest and lowest values in winter.

Summarizing all sampling hauls according to area and year of sampling, functional diversity displays an oscillatory behaviour from year to year in many areas of the North Sea, however, there are clear trends emerging in some areas. In the northern and central North Sea (Areas

1, 2, 3 and 7) most of the stations show a stationary trend over the entire period, or a mild increasing trend, which nevertheless is not significant. On the contrary, the southern North Sea displayed a significant negative trend for Area 6, and a positive trend for Area 4 (Fig. 3.16). Likewise, the Kattegat (Area 9) shows a significant negative trend, whereas the Shagerrak (Area 8) displayed no significant trend (Fig. 3.16). It is noticeable the major oscillations in Area 1 and Area 5 which take place in periods ranging from 5 to 10 years. In these areas, highest functional diversity values are followed by steep declines in the subsequent years. Less pronounced oscillations are also present in other areas, whether in the northern or southern regions of the North Sea.

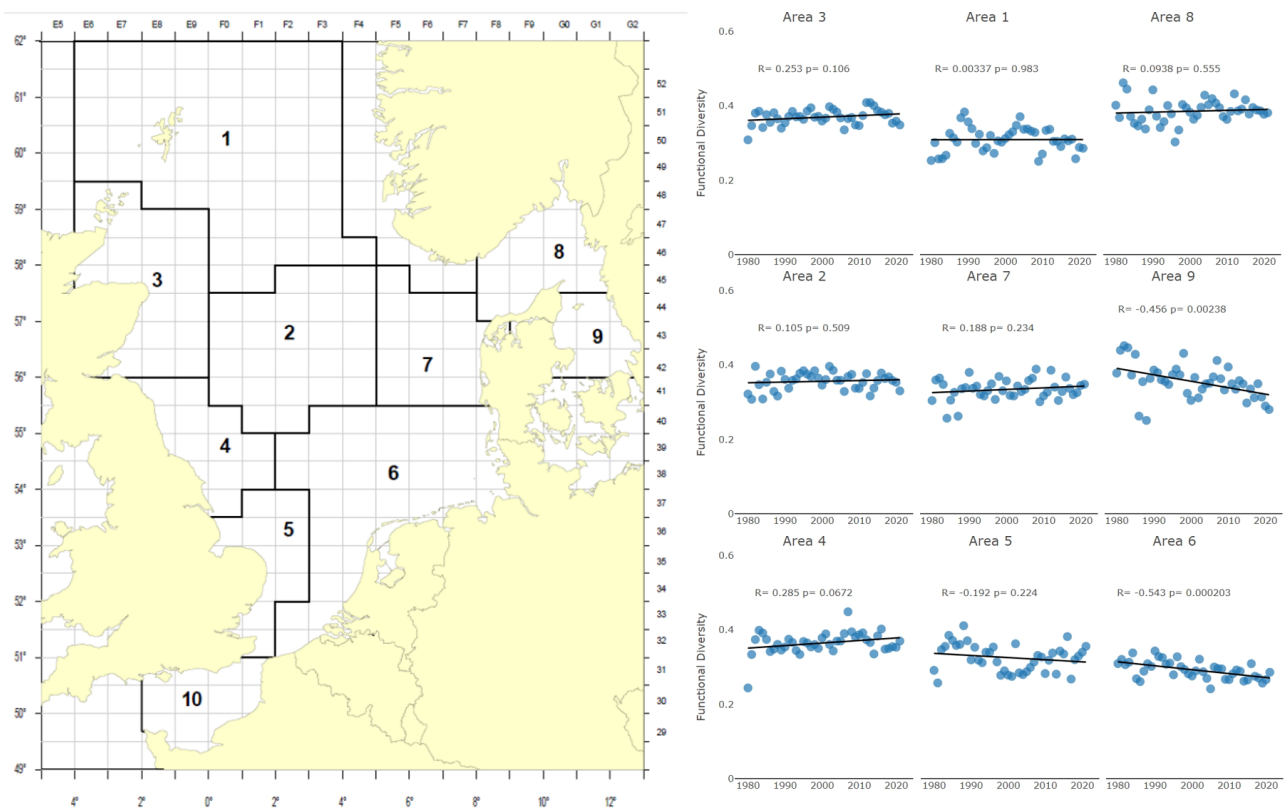


Figure 3.16: North Sea functional diversity variation according to area and year of sampling. While the northern and central North Sea (Areas 1, 2, 3 and 7) displayed either a stationary or a non-significant increasing trend, the southern North Sea did show a significant negative trend for Area 6, and a positive trend for Area 4. In the Kattegat (Area 9) there is a significant negative trend, whereas the Shagerrak (Area 8) showed no significant trend. Major oscillations in Area 1 and Area 5 are happening in time frames of 5 to 10 years.

The oscillations in functional diversity values in Area 1 might be related to the abundance variability of some Gadoid fishes, whose abundance decreased significantly at the end of 1990s, e.g., the haddock (*Melanogrammus aeglefinus*), the Cod (*Gadus morhua*) or the whiting (*Merlangius merlangus*) (Fig. 3.17(a)). The recovery in functional diversity during the decade of 2000s might be related to peaks of abundance of juveniles of various cartilaginous fish, such as the Cuckoo ray (*Leucoraja naevus*), the starry smooth hound (*Mustelus asterias*), the spotted ray (*Raja montagui*) or the dogfish (*Scyliorhinus canicula*) (Fig. 3.17(b)). In the case of Area 4, the positive trend display smaller oscillations. A likely explanation, could be related to a steady positive abundance trend of some Lusitanian species, e.g., the dragonet (*Callionymus lyra*), the grey gurnard (*Eutrigla gurnardus*) or the whiting (*Merlangius merlangus*) (Fig. 3.17(c)); the variability in juvenile abundance of flatfish such as the dab (*Limanda limanda*) or the plaice (*Pleuronectes platessa*); and the invasion of some other juvenile flatfish species, e.g., the scaldfish (*Arnoglossus laterna*) or the solenette (*Buglossidium luteum*) (Fig. 3.17(d)).

The steady decreasing trend in Area 6 might be explained by the abundance reduction of



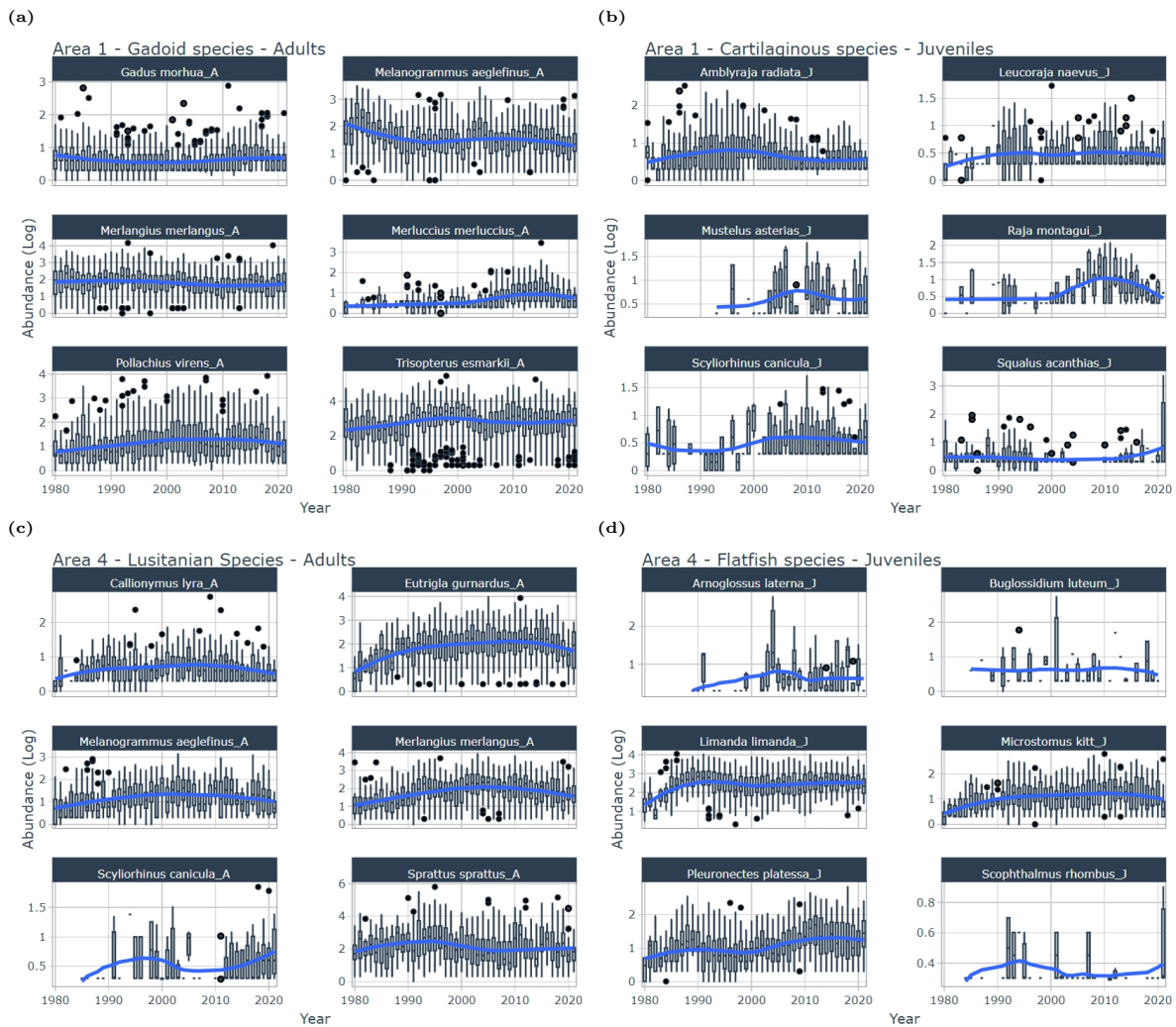


Figure 3.17: Box-plot abundance time series of key species reported in Area 1 and Area 4 of North Sea. Panel (a) shows the 6 most important Gadoid species, the Cod, the Haddock, the Whiting, the Hake, the Coalfish and the Norway pout, whereas panel (b) displays some cartilaginous fish, such as the Cuckoo ray, the Starry smooth hound, the Spotted ray or the dogfish. Panel (c) display Lusitanian fish with a steady positive abundance trend, e.g., the Dragonet, the Grey gurnard or the Whiting, whereas panel (d) shows juvenile abundance of permanent resident flatfish such as the Dab or the Plaice; and invader juvenile of flatfish such as the Scadfish or the Solenette. Abundances were yearly Log transformed. The blue line represents the time series trend.

the Cod (*Gadus morhua*) and the haddock (*Melanogrammus aeglefinus*), or the disappearance of the coalfish (*Pollachius virens*) and the Norway pout (*Trisopterus esmarkii*) (Fig. 3.18(a)). However, multiple juvenile of flatfish species saw a steady increasing trend during the 2000s, e.g., the scadfish (*Arnoglossus laterna*), the solenette (*Buglossidium luteum*), the lemon sole (*Microstomus kitt*) or the dab (*Limanda limanda*) (Fig. 3.18(b)). Conversely, in Area 9 both boreal species and flatfish species presented decreasing trends. For the former the most dramatic cases are the Cod (*Gadus morhua*), the vahl's eelpout (*Lycodes vahlui*), or the herring (*Clupea harengus*) (Fig. 3.18(c)). The latter includes the solenette (*Buglossidium luteum*), the lemon sole (*Microstomus kitt*), and the flounder (*Platichthys flesus*) (Fig. 3.18(d)).

In spite of the adequate summary that the analysis by area provides, it is likely that an assessment over such large spatial scales is masking the local variations within those areas. To accomplish a finer spatial resolution of functional diversity variation, we will summarize all sampling hauls according to subarea and year of sampling. The ICES subareas are standardized boxes ( $0.5^{\circ}\text{Lat} \times 1^{\circ}\text{Lon}$ ), which were sampled homogeneously during the International Bottom Trawl Survey (IBTS). A regression line was run for each subarea, as done for each Area in Fig. 3.16, thus, obtaining a slope that represents the yearly percentage variation of functional



Figure 3.18: Box-plot abundance time series of key species reported in Area 6 and Area 9 of North Sea. Panel (a) shows some Boreal fish species with either a decreasing trend, e.g., the Cod or the Raitt's sandeel; or that totally disappeared by the end of the time series, e.g. the Haddock, the Coalfish and the Norway pout. Panel (b) displays flatfish species with mostly increasing trends, e.g., the Solenette, the Lemon sole or the Dab. Panel (c) displays boreal species with dramatic decreasing trends, such as the Cod, the Vahl's eelpout, or the Herring. Lastly, panel (d) shows a similar scenario for flatfish such as the solenette, the lemon sole, and the flounder. Abundances were yearly Log transformed. The blue line represents the time series trend.

diversity ( $F_{div}/year(\%)$ ). The results are shown in a North Sea map with each sampling haul colored according to its respective subarea. The color-scale assigns warm colors to declining trends and cold colors to increasing trends. Stationary trends are colored in beige (Fig. 3.19).

Observing Fig. 3.19, we notice that, in fact, within areas there are multiple trends happening. Some geographically close subareas present totally different trends, but there are also large regions which share similar positive or negative trends. In the case of Area 1, most of the Fladen Ground displays the greatest losses in functional diversity, having loss around 0.3% yearly over a course of 40 years. However, the area known as Ulsira High and the segment running along the Norwegian Trench display a trend of around 0.4-0.5% yearly increase in functional diversity. For the case of Area 4, most subareas displayed either a positive or stationary trend, particularly in the area west of the Dogger Bank and along the eastern and central English coast. The southern North Sea, especially the Oyster Ground, the German and Southern Bight, had the most dramatic decreasing functional diversity trends. Some subareas reach peaks of 0.4 % yearly loss. Similarly, the Kattegat display mostly losses in functional diversity in its subareas (Fig. 3.19).

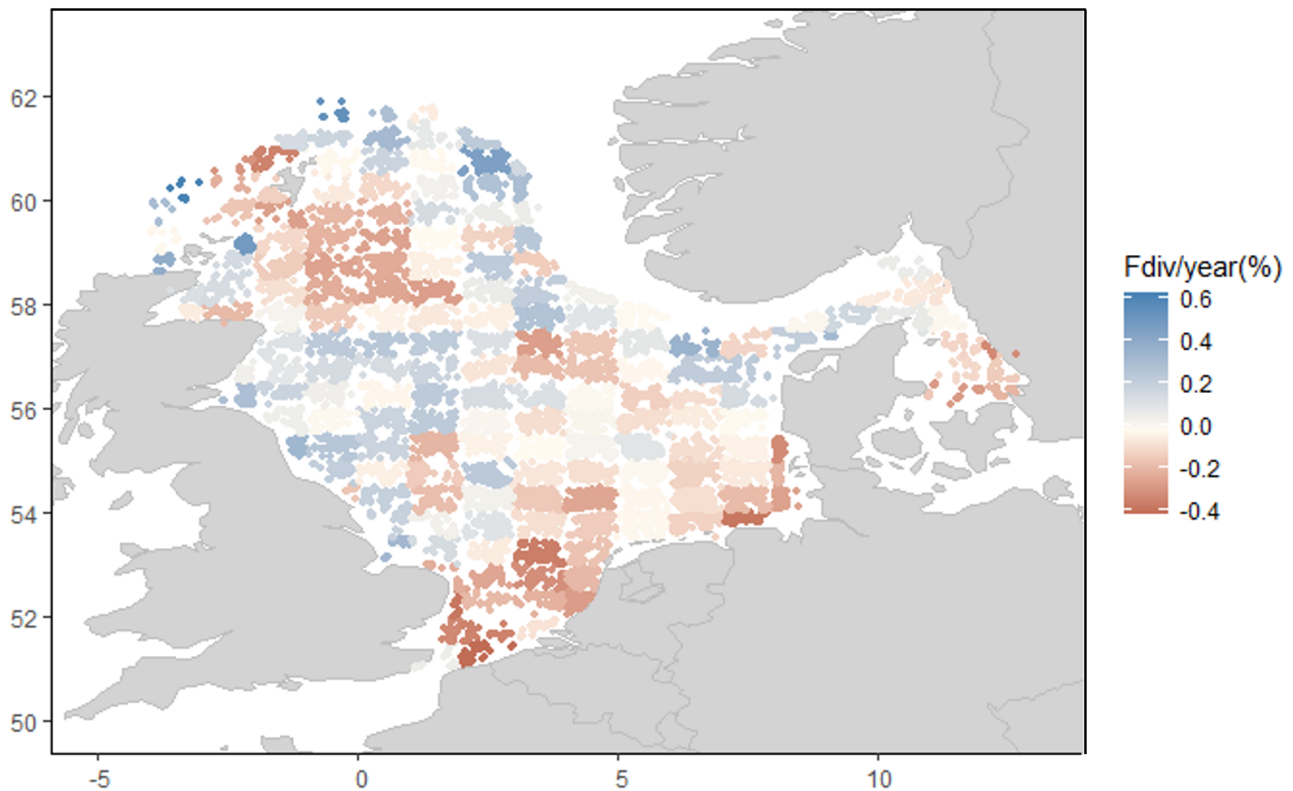


Figure 3.19: North Sea Map displaying the sampling hauls location which were executed as part of the International Bottom Trawl Survey (IBTS) during the period 1980-2021. Each sampling haul is colored by the yearly percentage variation of functional diversity (Fdiv/year(%)) of its respective subarea. The color-scale displays warm colors representing declining trends and cold colors representing increasing trends. Stationary trends are colored in beige

### 3.4 Discussion

In this chapter, I showed that diffusion map is a powerful tool to coarse-grain complex biological communities, such as the North Sea demersal fish community, in terms of certain structuring i-traits. The approach is able to handle multivariate long-term ecological datasets and unfold the lower-dimensional geometrical structure, or manifold, in which the system variation can be explain better. The results show that independent of human interpretation, there are coherent structures in the dataset which act as major explanatory variables. Although, knowledge, and perhaps some level of human intuition, is still necessary to add the labels ‘depth’ and ‘year of higher abundance’ to these structures, the results are not a mere product of interpretation, but rather have a reality which is independent of these labels [21].

The analysis revealed depth as the most important variable structuring the fish community in the North Sea. In fact, depth has been referred in numerous studies as an important determinant of fish distribution and community structure in the North Sea [36, 64, 113, 135, 346]. Perhaps in our diffusion map, depth is working as a summary variable which encapsulates those other variables that co-vary with depth, e.g., temperature, seasonality, salinity and productivity [36]. Moreover, the described two-part system nature of the North Sea basin [124, 228] and the different dynamics taking place in the deeper northern region and the shallower southern one, will either limit or enhance the possibilities of the fish fauna to thrive [116, 193]. Consequently, we see dominance of different groups of fish in either region. While in the northern and central North Sea there are usually uneven communities composed of species with a high mean trophic level and a low proportion of benthivory (e.g. gadoids and generalist/piscivorous species), in the southern North Sea there are uneven communities with lower trophic level species primarily

feeding on benthos (e.g. flatfish) [90, 173, 269].

The second most important variable is the year of highest abundance for the fish. The interpretation of this is less straightforward as with the case of depth. It was detected that species which reached their abundance peak in the period 1980-1999 were mostly associated to the negative side of EV2, whereas those fish reaching abundance peak in the period 2010-2021 did mostly to the positive side of EV2 (Fig. 3.8). Our hypothesis suggests this is correlated to a decreasing abundance trend of boreal fish species, e.g., the starry ray (*Amblyraja radiata*), the four-bearded rockling (*Enchelyopus cimbrius*), the bullrout (*Myoxocephalus scorpius*), the cod (*Gadus morhua*) and the haddock (*Melanogrammus aeglefinus*); while increasing abundance of multiple lusitanian fish species, e.g., the hake (*Merluccius merluccius*), the whiting (*Merlangius merlangus*), the grey gurnard (*Eutrigla gurnardus*), or the dogfish (*Scyliorhinus canicula*). Reports in the literature have extensively suggested the North Sea has been losing temperate-incline species in favor of warm-inclined fish from lower latitudes [8, 118, 315]. Some extreme cases, such as the wolffish and the starry ray, point towards a northward contraction in their distribution [40, 370], or a tendency to expand their range towards deeper areas of the North Sea [113].

However, the ‘lusitanian invasion’ in the North Sea does not seem to be spatially homogeneous. In the deeper northern region, dominant species in either period do not coexist temporally. On the contrary, in the shallower southern region most of dominant species in the 2010s were present as early as the mid-90s. This asymmetry might be due to southern fish establishing first in northern-central areas rather than in southern ones. Such result reinforces the higher relevance of the Shetland–Orkney gap over the English Channel as a main route into the North Sea for southern warm-tolerant species [112]. Moreover, species which invaded from the southern English Channel, such as the anchovy (*Engraulis encrasicolus*) or the Pilchard (*Sardina pilchardus*), did barely expand to central North Sea, but remained absent from northern areas [209, 339].

Whether the lusitanian fish are newcomers or local opportunists, such change implies a shift towards new fish traits [37, 101, 160]. The North Sea fish community is giving room to traits of opportunistic and equilibrium strategists at the expense of traits of periodic strategist [313, 434]. In other words, dominant traits during the first 20 years were a medium to long life span and length, high trophic level, high fecundity but low parental care and offspring size, characteristic of gadoid species or large flatfish [181, 269]. Dominance in the last 10 years has turned to traits such as small size, low trophic level and short life span but with relatively high fecundity and low parental care, e.g., small pelagic fish, gobiidae species and small flatfish; or traits like long length and life span and high trophic level, low fecundity but large offspring size and high parental care, which are characteristic of rays and sharks [36, 125, 393].

Based on the identified i-traits and having reconstructed the trait space in which the species locate, calculation of the functional diversity was done by using the Rao’s quadratic entropy [342]. This method is, in fact, a generalization of the Simpson’s index of diversity [376], and it can be used with various measures of dissimilarity between species. Such index might be viewed as a multivariate measure of functional divergence, i.e., the degree of resource differentiation and thus competition, but also of the predominance of extreme species [265, 364]. High functional divergence indicates a high degree of niche differentiation, and thus low resource competition. Communities with high functional divergence may have increased ecosystem function as a result of more efficient resource use [265].

Our approach addresses common issues when intending to quantify functional diversity. It uses multiple traits to determine species dissimilarity, weights the traits relevance accordingly to their respective eigenvalue, and since the reconstruction of trait space is done from long-term monitoring dataset, it also avoids issues of traits measured at different scales [147, 233, 359]. Although issues such as interspecific overlap in niche space, or within species niche differentiation might be a possibility, we consider the former is an aspect of functional redundancy rather than of functional diversity [265, 356], whereas the later has been addressed by applying the concept of size at maturity [405], considering there might be important differences in the immediate functioning of adult individuals vs. juvenile individuals within the same species [121, 154, 183].

In the North Sea, the last 40 years have seen multiple trends in the variation of functional diversity both in the spatial and temporal scale. Within the year there were no significant differences for functional diversity values. This might be related to the biology of fish itself [90]. Fish have longer timespans, but reproduction tends to vary widely between species, not being restricted to a specific season [9, 158]. Moreover, fish are not immediately recruited into the community, as they spend a significant amount of time in the pelagic environment as larvae, and then in coastal nursery areas as small alevines [23, 43]. Consequently, when compared to highly seasonal organisms, e.g., bacteria and phytoplankton [68, 267, 359], a fish community functional diversity is less subject to seasonal oscillations

Unlike seasonal variations, functional biodiversity values showed notorious fluctuations at inter-annual, quinquennial or even decadal periods. At the same time, such oscillations were not spatially uniform. At the scale of ICES areas, the strongest oscillations were present at Area 1 in the north, Area 5 in the southwest, and the Kattegat (Area 9). With less marked oscillations, but displaying a clear decreasing yearly trend in functional diversity, we find the southern areas 6 and 5. While solely area 4 displayed a significant positive increasing yearly trend of functional diversity. Nevertheless, within those areas such trends were not homogeneous either. Distinct trends are observed at the scale of ICES subareas, with outstanding losses in the areas known as Fladen Ground, the Oyster Ground, the German and Southern Bight; and important gains at the Ulsira High, along the Norwegian Trench and along the eastern and central English coast (Fig. 3.19).

This changes are likely linked to the strong fluctuations in the abundance of juveniles gadoids, the invasion of some cartilaginous fish in north-western areas, as well as the ever more prominent predominance of small flatfish in the southern and central North Sea. Although biological populations will fluctuate in size from time to time, reasons to this are not always the same. An exploration of possible factors influencing such abundance fluctuations, hence, trends in functional diversity, can be divided into two kinds: deterministic and stochastic factors [338]. The former is related to those straightforward cause-and-effect relations that to some extent can be predicted, e.g., usually human activities such as fishing, habitat alteration or destruction, introduction of alien species, blockading migration routes. The later instead is related to those operating in a realm beyond human prediction and control, either because they are truly random or because they are linked to complex geophysical or biological causes [338].

Although this research cannot provide a definitive answer to which factors are driving functional diversity variations, possible explanations can be hypothesized. In our results scale provides an important hint about what factors might be influencing the ecosystem [77]. As the observed changes have impacted the North Sea over large geographical areas, processes inducing them must also operate on a similar large scale, the most evident of which is climatic forcing

[81, 319, 351]. The most important large-scale hydro-meteorological drivers of the North Sea ecosystem are the sea-surface temperatures and the strength of the westerly winds in this region which is linked to the North Atlantic Oscillation [116, 303]. Both have undergone marked changes in the last 50 years, but the most dramatic has been the stepwise increase in the sea surface temperature (SST). The SST has increased about  $1.2^{\circ}\text{C}$  relative to the pre-1980 mean value, with a trend of  $0.4^{\circ}\text{C}/\text{decade}$ , a ten-fold sudden increase [124].

Anthropogenic effects tend to be restricted to a much smaller spatial area, usually around the immediate area of the impact into the marine system (e.g., sewage sludge dumping, mining among others). Although there are exceptions. Fishing exploitation is one of those exceptions. Fishing can influence fish populations because by removing large individuals they alter the age-length population structure, abundance and occupancy of target species. Fish that went through overexploitation, such as the cod, are now rare or are restricted to the northern areas [82]. However, fishing effort is not spatially homogeneous, in fact, fishing effort, particularly by beam trawls, has been greater in the southern North Sea compared to the northern part [201]. Our results have shown southern areas, which have been heavily impacted by bottom trawling fishing, e.g., the Oyster Grounds, The Dogger Bank and the Southern Bight [136, 152], displaying particular losses in functional diversity. Moreover, other activities might have also played a role. It is intriguing the significant decreasing trend in the Fladen Ground, an area of oil production, where some reports of hydrocarbons in sediments have been reported [6, 358].

Ultimately, the long-term changes in the North Sea ecosystem are likely to be a combination of both deterministic and stochastic factors. Nevertheless the reasons, our results suggest a change has been undergoing in the fish community of the North Sea fish community. Such changes in functional diversity might imply either a temporal species turnover [232, 258, 386], or could be considered an ecosystem regime shift [26, 362, 427]. Regime changes in the North Sea phytoplankton community have been detected around 1978, 1989 and 1998. The first two changes were defined as a cold episodic events and the later a regime shift towards a warm dynamic regime [10]. Future research might point towards exploring such scenario and the consequences it might bring [280].

In conclusion, the diffusion maps approach allows us to coarse-grain complex marine fish communities in terms of their functional i-traits. Thus, providing a quantitative framework to reconstruct an i-trait space, a potential niche space over time for fish of the North Sea. This approach enhances our possibilities to assess functional diversity directly from long-term abundance datasets, hence, allowing us to unfold new insights from an ecosystem which has been adequately monitored. The results demonstrate the power of manifold learning approaches to highlight the relationships between community diversity composition and ecosystem functioning, making it feasible to assess variation of functionality in ecosystems over time.

## Chapter 4

# Aggregation of time series: Computing functional diversity of the Wadden Sea and southern North Sea phytoplankton community

Long-term monitoring data is central for the analysis of biodiversity change and its drivers. Known as time series, they have allowed a more accurate evaluation of diversity indices, trait identification and community turnover. This kind of datasets help to resolve both short- and long-term scales of variability, providing context for traditional process-oriented studies [297]. In spite of their relevance, time series of biological communities are not abundant. The few examples expanding over 20 years of sustained monitoring, have faced severe funding difficulties, being forced to either reduce sampling frequency, thus resulting in temporal gaps in the observations, or cease their operations completely [297]. And when new surveys are established, data collecting methods might be modified to better address their primary research questions.

A possible solution to obtain more adequate time series, one that monitors an ecosystem over time spans of a few decades and beyond a local scale, is to aggregate datasets of different origins. This aggregation of datasets is, nonetheless, a non-trivial task. Difficulties in harmonizing the data will arise as variations in the observation of taxa and taxonomic names are the outcome of the discrepancies in the equipment, personnel expertise, and procedures employed by different sampling teams [266].

Here we propose a method for aggregating datasets using diffusion maps. Our approach is illustrated by aggregating long-term phytoplankton abundance datasets from the Wadden Sea and the southern North Sea gathered by two institutions located in Germany and The Netherlands. The aggregated data allowed us to infer species traits (i-traits), to reconstruct the main i-trait axis driving community functionality, to ultimately quantify functional diversity of the individual samples, having used only the co-occurrence of species in samples.

This chapter follows the lines of Carrasco and colleagues [68]. I start with an introduction to the issues of biological time series (Sec. 4.1). In Sec. 4.2, the diffusion maps approach is applied to a new time series generated via a simple aggregation of two phytoplankton abundance datasets. With this approach, diffusion maps fail to identify important i-traits from the newly merged time series. In Sec. 4.3, I apply a different approach to adequately aggregate datasets. In

this case, after diffusion mapping the aggregated time series, a trait space can be reconstructed and relevant i-traits were identified. In addition, I estimate the functional diversity of the phytoplankton community over time (Sec. 4.4). Lastly, a summary of the approach is presented in Sec. 4.5.

## 4.1 Issues of ecological time series

Time series are measurements of a quantity taken over time, with a regular frequency, and covering different spatial scales [140, 297]. Depending on research question, sampling frequency, and temporal and spatial extension, time series can be used for different purposes and are of critical importance to enable or facilitate:

1. the acquisition of ecosystem baselines and rate and scale of environmental change, including climate change and biodiversity loss [165]
2. the understanding of ocean, earth, and climate system processes [375]
3. the monitoring of ecosystem dynamics and its variability [307]
4. the detection of hazards and environmental disturbances and the estimation of recovery times [162]
5. the forecast of ecosystem changes [102]
6. the effective policy-making and sustainable management of the seas and oceans [143]

However, the fact that oceanographic processes occur at multiple scales makes it necessary that observational time series have to be several-fold longer than the time-scale they are trying to resolve [32, 48]. For example, outputs of oceanographic models have suggested that at least three decades would be needed to resolve climate change response in the North Atlantic for certain variables (e.g., primary production) [175], but less for others (e.g., sea surface temperature) [174].

It is paradoxical that while longer time series are necessary to resolve biological processes, they are the most scarce and restricted [240]. In one hand, ecological time series are concentrated in the coastal ocean but almost completely absent in the open ocean. Moreover, within coastal areas there is a clear bias towards coastal zones in North America and Europe, while ecological time series in other coastal regions around the world are less mature [24, 297]. On the other hand, time series are often cross-sectionally wide (e.g., census many interacting species) but short in the time dimension [76, 188].

Ecological time series such as the the ICES coordinated International Bottom Trawl Survey (IBTS) [195], used in Chapt. 3, are extraordinary in terms of both the spatial scale covered and the duration it holds. More often what we find are local or countrywide efforts that sustain observations over periods of a few decades. Such is the case of multiple phytoplankton abundance datasets in the North Sea basin. Aggregation of phytoplankton datasets from multiple origins represent a feasible solution to extend the temporal resolution as well as to improve the spatial coverage that is not reached by individual or isolated time series [197].



Although aggregating datasets seems a trivial and straightforward solution, bringing together datasets from different time series rises questions about the sampling and analytical protocols used at each site. Were the methods used at each site homogeneous, consistent, and intercomparable? What implications does the personnel expertise, or the institutions facilities have on taxa identification? Such questions are relevant considering that, while observations of physical parameters and associated data processing and quality control procedures are well established, biogeochemical parameters, particularly biological and ecological measurements, are less mature [240].

Nevertheless the possible discrepancies between datasets, the development of an adequate and careful aggregation approach is necessary if we want to obtain a time series which can help us to obtain solid insights about ecological communities [266, 359]. In Sec. 2.4, I introduced diffusion maps [79, 80] a manifold learning method which finds new dynamically relevant variables that describe the most important dimensions in a system. In the following section we use this methodology with a simple, almost naive dataset aggregation approach, to then diffusion map the resulting time series.

## 4.2 Failure of simple aggregation

We start by illustrating the diffusion mapping procedure on single datasets. We will follow the algorithm introduced in Sec. 2.4.1 for diffusion mapping the phytoplankton abundance, using the co-occurrence of species as similarity measure, and implementing and executing all diffusion map steps in the Julia programming language [38]. The phytoplankton dataset analyzed next is part of the extensive monitoring program conducted by the Lower Saxony Water Management, Coastal Defence and Nature Conservation Agency (NLWKN), in Germany [291]. We used harmonized data from 04 stations, including 1,664 samples and 249 species. The data harmonization consisted of first removing all species identified as purely heterotrophic [300], and second, homogenizing and updating phytoplankton species nomenclature using the WORMS website taxonomic database [7]. Stations are located on the German area of the Wadden Sea, including the Jade Bay and the Wesser river estuary (Fig. 4.1).

Following we have the phytoplankton dataset which is part of the extensive monitoring program conducted by Rijkswaterstaat, in the Netherlands [20]. Here we used data from 18 stations, including 3,691 samples and 366 species. Data harmonization and taxonomic updating was also performed using the WORMS website taxonomic database [7]. The geographical extension covered by this dataset is significantly larger, as it spreads not just over the Dutch area of the Wadden Sea, but also in the off-shore Terschelling and Rottumerplaat transects in the southern North Sea (Fig. 4.1).

The reconstructed *i*-trait space of the German dataset is slightly skewed towards the negative side of EV1 (Fig. 4.2 (a)). This variable is the most relevant to explain the variation in the dataset. Although when assessing the inverse eigenvalue ranking, the first variable explained poorly such variation (Fig. 4.2 (b)). On the contrary, for the Dutch dataset the first variable gathered most of the variation, as it can be noted from the inverse eigenvalue ranking (Fig. 4.2 (d)). It is also noticeable that the *i*-trait space of the Dutch dataset is slightly skewed towards the positive side of EV1 (Fig. 4.2 (c)).

We observe that the analysis of individual datasets may limit our ability to construct a

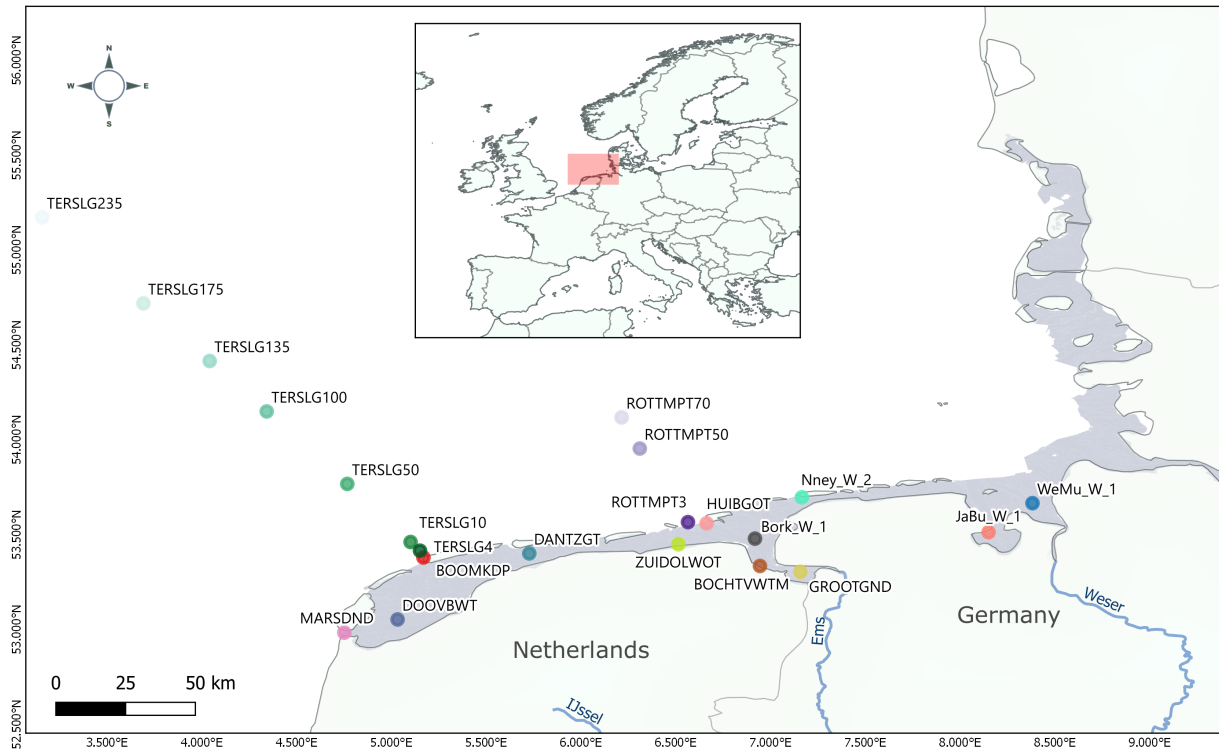


Figure 4.1: Map of the Wadden Sea and Southern North Sea areas including the phytoplankton monitoring stations. Four stations (Bork\_W\_1, Nney\_W\_2, JaBu\_W\_1 and WeMu\_W\_1) were sampled by the Lower Saxony Water Management, Coastal Defence and Nature Conservation Agency (NLWKN). German stations covered only the East Wadden Sea. A total of 18 stations were sampled by Rijkswaterstaat, 10 stations in the off-shore southern North Sea, whereas 8 in the West Wadden Sea.

reliable network, especially for datasets in which the number of samples or the number of species is small. When this happens, we are forcing a comparison between dissimilar species, obtaining a degraded i-trait space quality [21, 127]. Therefore, the aggregation of datasets becomes the best solution, although the procedure to execute it is not yet clear.

Following now, I will demonstrate that datasets cannot be aggregated directly. For this purpose we use the previously introduced dataset by Rijkswaterstaat, in the Netherlands [20] and the dataset collected from the monitoring program of the NLWKN in Germany [291]. The starting point for the analysis is to simply aggregate both datasets into a single table of phytoplankton species (rows) and sampling identifier (columns). When a species is present in a particular sample the entry will be the abundance recorded in the dataset, indistinctly of the host institution. When the species is absent in the sample, the table entry is 0. Next, we will follow the algorithm already introduced in Sec. 2.4.1 to diffusion map the aggregated dataset.

As a result, the EV1, which represents the primary pattern detected by the method in the data, clustered some species into two groups: those phytoplankton species only observed in the Netherlands and those only observed in Germany (Fig. 4.3). This is not a desired result, but rather an artifact from the data gathering. Plankton monitoring is a difficult task. Attribution of different taxonomic identities for similar observations might happen due to the high number of taxa or their sometimes high morphological similarity. Although a certain degree of local endemism is possible [60, 95, 246, 396, 415], the geographical context makes this only a partial explanation. Consequently, what we see here is that diffusion maps picks up on an artefact that

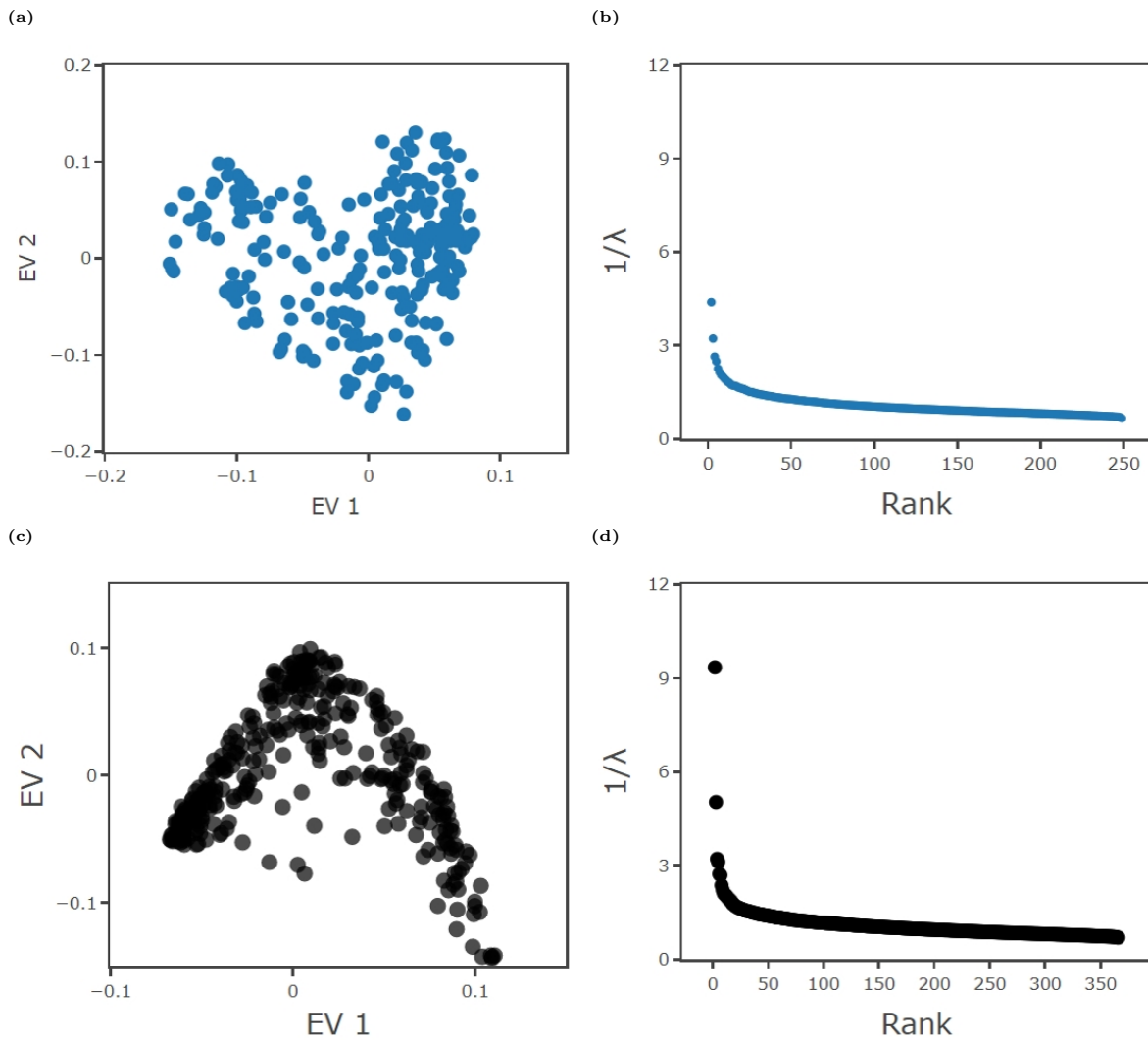


Figure 4.2: Diffusion map displaying the reconstructed  $i$ -traits that explained best the variation of German dataset (a) and Dutch dataset (c). Panels (b) and (d) show the ranked inverse eigenvalue ( $1/\lambda$ ) spectrum across all species datapoints for both datasets. In both cases the first non-zero eigenvalue explains most of the variation, but with lower score for the German data (b) compare to the Dutch data (d).

is rooted in the nature of the data collection and then is exacerbated by the naive aggregation. This defines the need for an aggregation procedure that avoids such artefacts.

### 4.3 Successful aggregation of phytoplankton datasets

To find a better procedure for aggregation, let us analyze why the separation into Dutch and German species occurred in the naive attempt. When considering different monitoring datasets, the list of observed species in the respective areas may be different because some species are absent in one of the areas. Unfortunately, consistency between datasets is compromised by difficulties in identifying many phytoplankton taxa [388]. Severe problems may arise when there is a change in personnel analysing samples during the course of a long-term programme. For instance, major changes in phytoplankton community structure in the Helgoland Roads phytoplankton time series [432], as well as in the Dutch North Sea monitoring programme [314], arose from changes in the staff or laboratory responsible.

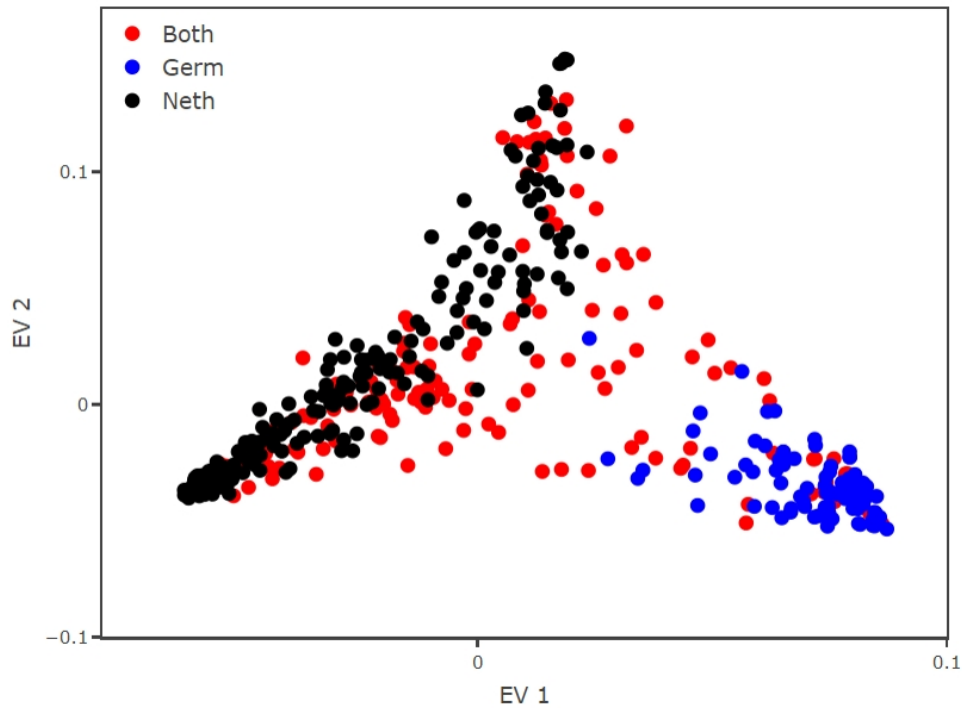


Figure 4.3: Reconstructed i-trait space from the aggregated monitoring dataset using the simple aggregation method. Applying a naive aggregation makes the species (dots) cluster in those observed only in Germany (blue dots) and those observed only in The Netherlands (black dots). The species that are common to both datasets are colored in red. Reproduced from Carrasco and colleagues [68]; open access under the Creative Commons Attribution License - CC BY 4.0.

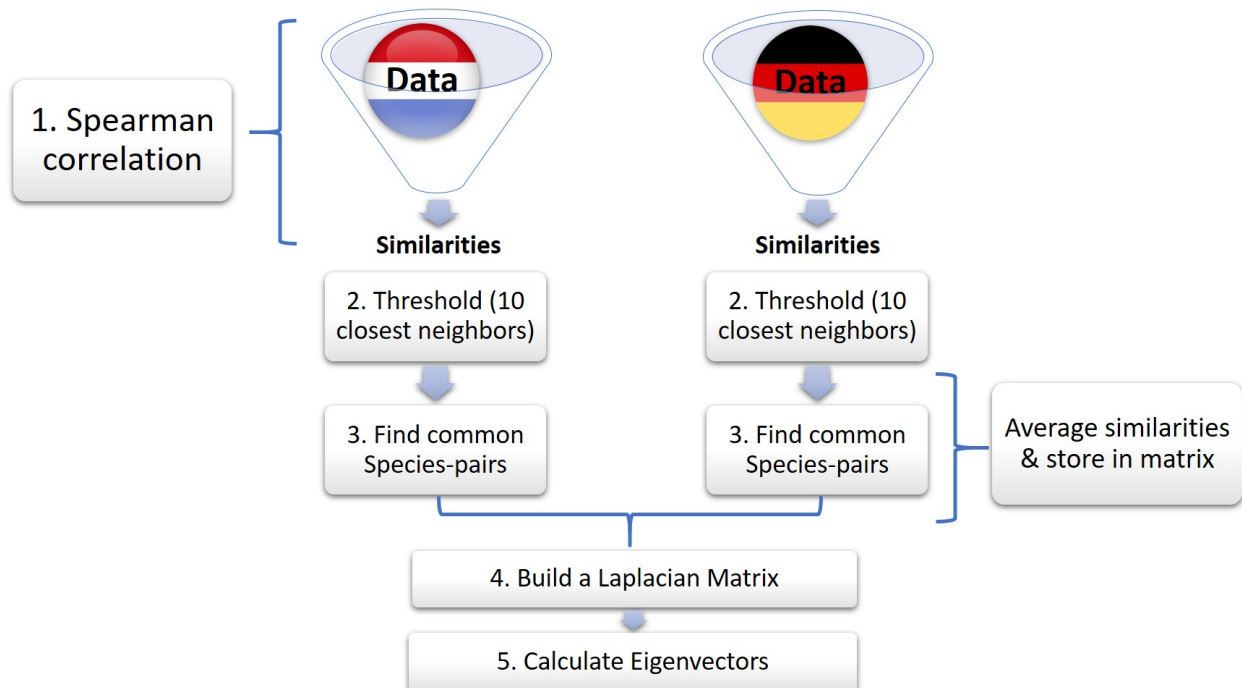


Figure 4.4: Schematic of the proposed method for aggregating monitoring datasets. In step 1, we calculate similarities of German and Dutch phytoplankton abundance data separately. In step 2 we choose the 10 highest similarities (known as threshold). In step 3, after identifying the common species-pairs, we average their similarities and store them in a new matrix. The rest of species-pairs are store with their original similarity values. In step 4 we construct a Laplacian matrix, which is finally used to calculate the eigenvectors in Step 5. Reproduced from Carrasco and colleagues [68]; open access under the Creative Commons Attribution License - CC BY 4.0.

In addition to changes in the scientists responsible, the effects of more subtle modifications in sampling protocol, which are likely to occur during a monitoring programme, are less clear.

First, phytoplankton taxonomy has been under continuous change over the past few decades. New identification keys have appeared since the 1980s; for instance, for diatoms [220–222] and cyanobacteria [217]. Second, technological advances (i.e. better microscopes with higher resolution) might have improved the scrutiny of samples during a programme spanning several decades. Third, the design of the sampling programme (i.e. seasonal or depth resolution of samples) might have changed. Fourth, species identification will have improved with an increase in the experience and knowledge of those concerned [327].

In one hand, while it is clear that these are important problems, which possibly confound interpretation of diversity estimates from long-term data, it is not clear how significant these problems actually are [388]. On the other hand, to just lament such differences between datasets, and call for more standardisation is an easy answer. Different cultures and capabilities may also open up different angles on a complex system that, when properly taken into account, reveal additional information [68].

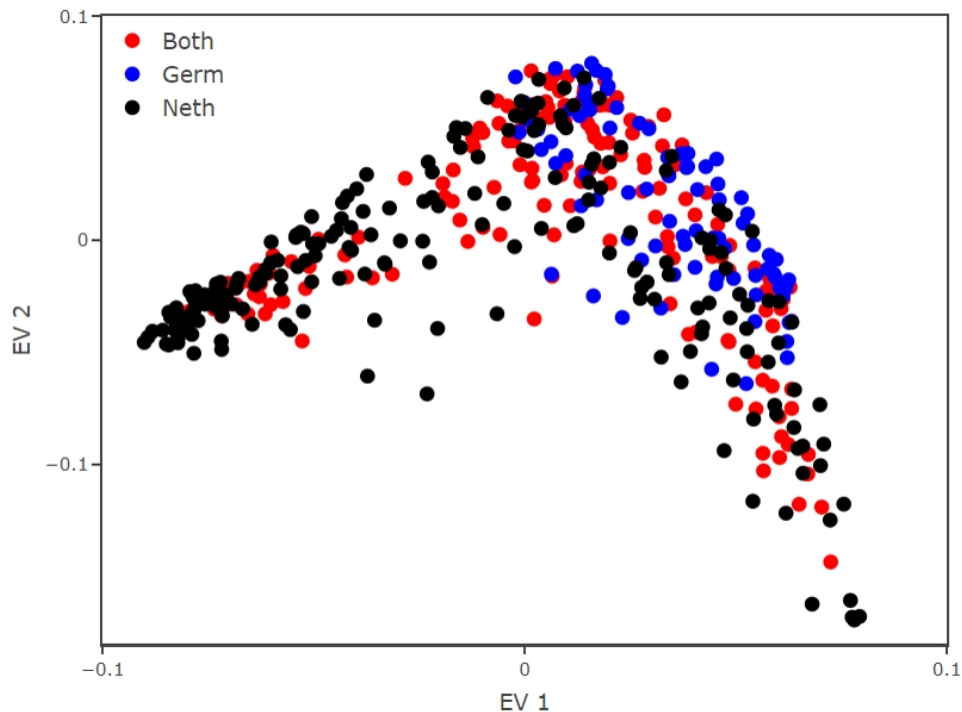


Figure 4.5: Reconstructed i-trait space from the aggregated monitoring dataset using our proposed aggregation method. Applying our method breaks the cluster, having species observed only in Germany (blue dots) and those observed only in The Netherlands (black dots) spreading indistinctly over the manifold. This provides a better reconstructed i-trait space, avoiding data artefacts. The species that are common to both datasets are colored in red. Reproduced from Carrasco and colleagues [68]; open access under the Creative Commons Attribution License - CC BY 4.0.

It must be recognized that if a species is not observed in a given sample, it may indicate the actual absence of the species or it may signal that the species, while objectively present, was not identified or was assigned a different name [231, 316]. In our naive merging procedure we interpreted the absence of an observation as evidence for the absence of the species from the respective sample. This assumption leads to an erroneous matrix of similarities, whose biases make species that occur in only one of the countries appear different from the others.

We propose a more careful approach to dataset aggregation, one that fixes the epistemological shortcomings of the naive procedure. We illustrate this approach using the datasets gathered by Rijkswaterstaat [19] and by NLWKN [291] (Fig. 4.4). After basic data harmonization each of these datasets can be considered as internally consistent regarding the taxa identification.

Thus, we can safely construct and threshold the similarity matrices for the individual datasets as described in Sec. 2.4.1.

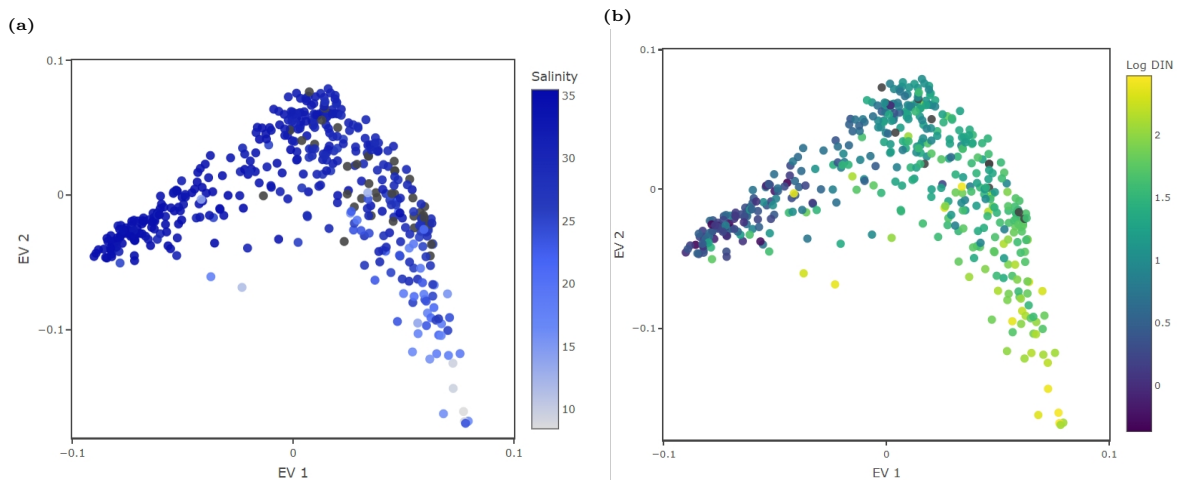


Figure 4.6: Inferred traits from the monitoring dataset. Color coded are environmental conditions under which the species were observed with high relative abundance. The EV1 aligns well with salinity (left) and DIN concentrations, displayed in logarithmic scale (right). This EV1 probably separates species by their adaptation to salinity levels or their nitrogen requirements. Reproduced from Carrasco and colleagues [68]; open access under the Creative Commons Attribution License - CC BY 4.0.

We then merge the processed similarity matrices as follows: We consider all possible pairs of species. For some of these pairs both species exist in both matrices. We interpret that as a sign that the corresponding species are reliably identified by both agencies, and hence average the value of the respective similarities. For some pairs one or both of the species exist only in one of the matrices. We interpret this as an indication that only one of the agencies can make this comparison reliably and hence accept the value from the matrix where the comparison is possible. Finally, some comparisons cannot be made in either of the matrices because one species exists only in one of the matrices while the other one exist only in the other matrix. In this case we set the similarity of the species to zero as no reliable comparison is possible.

The final choice means that we may assign some zeros to comparisons between similar species (or even between the same species which were identified by different taxonomic IDs). However, setting some comparisons wrongly to zero does not degrade the quality of the diffusion map result [359]. The reconstructed trait space shows that the EV1 does no longer clusters the species into country of observation, rather we observe that they spread indistinctly over the manifold (Fig. 4.5). The first i-trait aligns well with Dissolved Inorganic Nitrogen (DIN) as well as with the water salinity (Fig. 4.6). We conclude that this i-trait could represent adaptation to different water basin conditions (nutrient availability and salinity), which are different for the Wadden Sea and the southern part of the North Sea [412, 413]. Such interpretation is likely, as it is being considered in the scientific literature [70, 207].

## 4.4 Functional diversity status of southern North Sea and Wadden Sea

Once the i-trait space has been successfully reconstructed for the aggregated datasets, we can use it to calculate the distance in traits space for each species pair,  $i$  and  $j$ . Such distance, defined as  $d_{ij}$ , is calculated by using the euclidean distance in the reconstructed trait space. The species i-traits are now given by the eigenvector elements corresponding to the species,

re-scaled by the respective eigenvalue, as in Eq. 2.15. These distances are then used to quantify the functional diversity of each sample (Eq. 2.16), by means of the Rao's quadratic entropy [342]. The outcome obtained is a time series with a functional diversity score per sample.

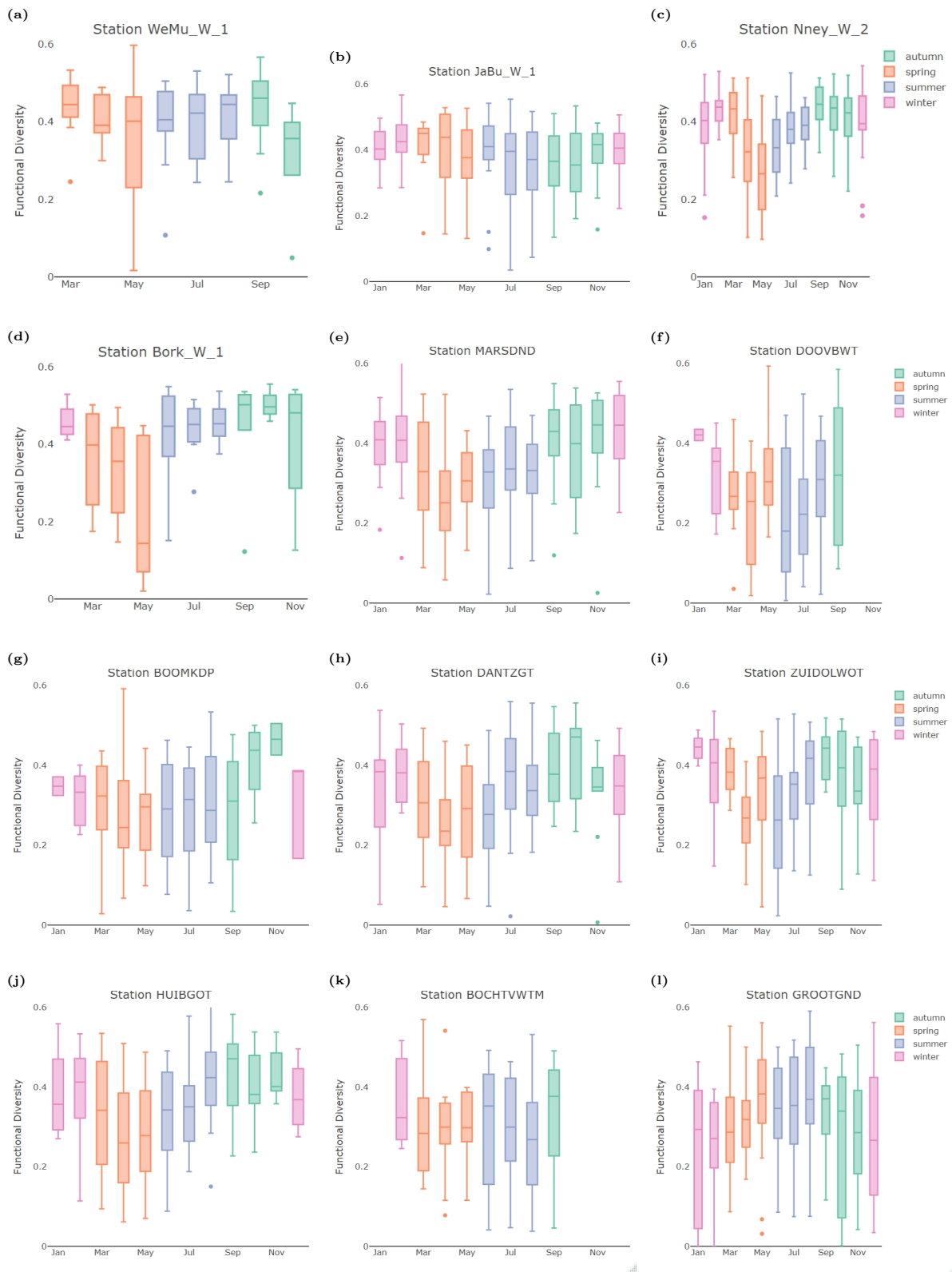


Figure 4.7: Wadden Sea functional diversity variation according to station and season. Although some stations showed higher mean functional diversity values in winter and autumn (MARSNDND, HUIBGOT or DANTZGT), such difference was not extreme. In other stations the differences between seasons were minimum, e.g., WeMu\_W\_1, JaBu\_W\_1.

The assessment of functional diversity variation was evaluated both in the temporal and

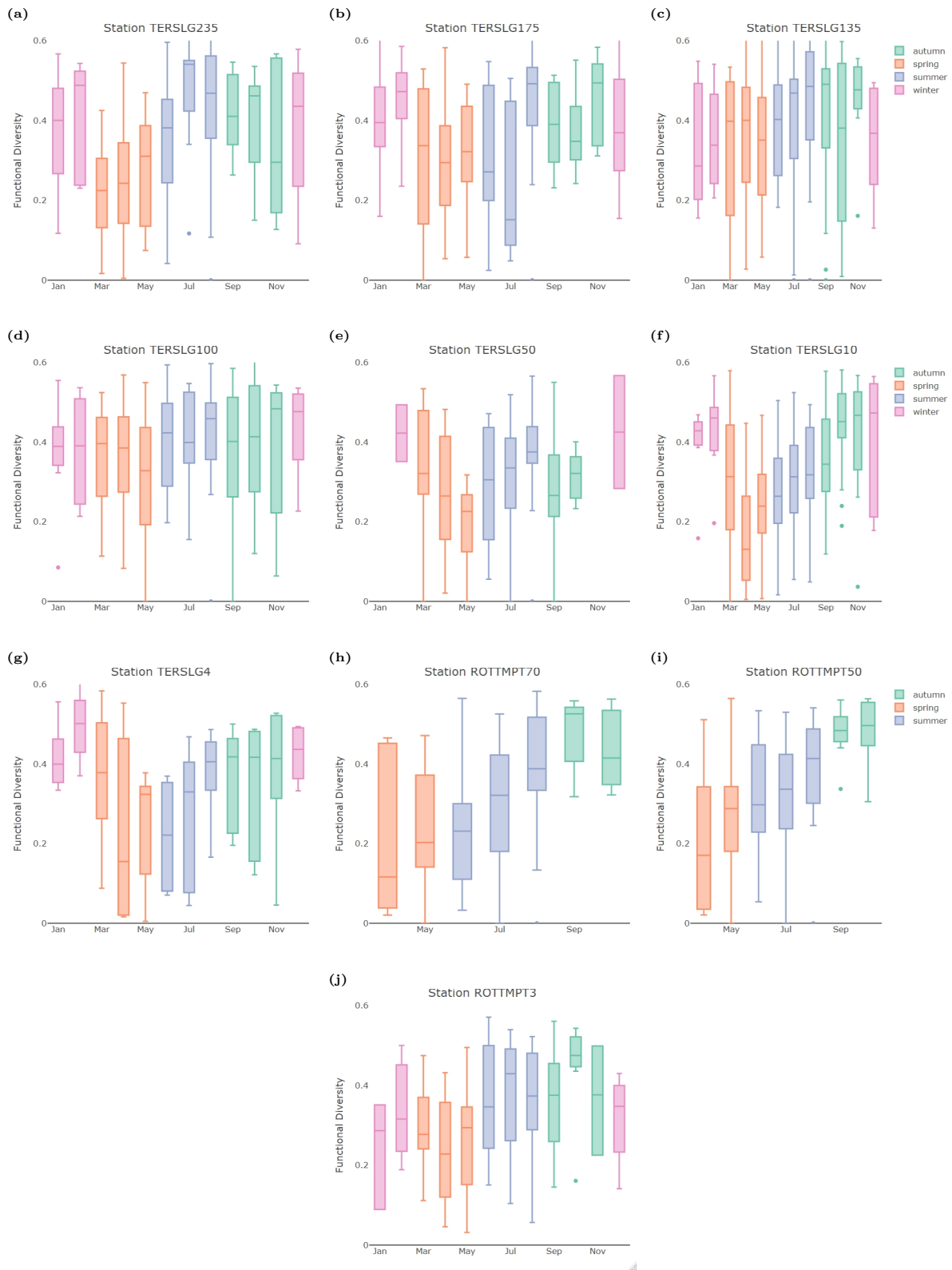


Figure 4.8: Southern North Sea monitoring stations functional diversity variation according to station and season. The stations displayed a wider difference between autumn-winter and the spring season, e.g., TERSLG235, TERSLG10, ROTTMPT3.

spatial scale. For a better visualization the results are displayed in two groups; stations located in the Wadden Sea (Fig. 4.7), and stations located in the southern North Sea (Fig. 4.8). Over the yearly cycle there was a variety of trends among stations in both water basins. In one hand, although some stations at the Wadden Sea showed higher mean functional diversity values in winter and autumn (Station MARSND, HUIBGOT or DANTZGT), such difference was not



extreme. For other stations the differences between seasons were minimum, e.g., WeMu\_W\_1, JaBu\_W\_1 (Fig. 4.7). On the other hand, some stations at the southern North Sea displayed a wider difference between autumn-winter and the spring season, e.g., TERSLG235, TERSLG10, ROTTMPT3 (Fig. 4.8).

There is significant annual, seasonal as well as inter-station variations in functional diversity estimates. However, when considered over the entire period, clearer patterns emerge. On the one hand, significant functional diversity losses occurred at most Dutch Wadden Sea stations, with fastest decrease observed at the Marsdiep basin (MARSDND and DOOVBWT stations) and off the coast of Groningen, Lauwers basin (ZUIDOLWOT station). On the other hand, there is a mild increase of functional diversity in the German Wadden Sea stations, with the fastest increase at the Weser estuary, WeMu\_W\_1 station (Fig. 4.9).

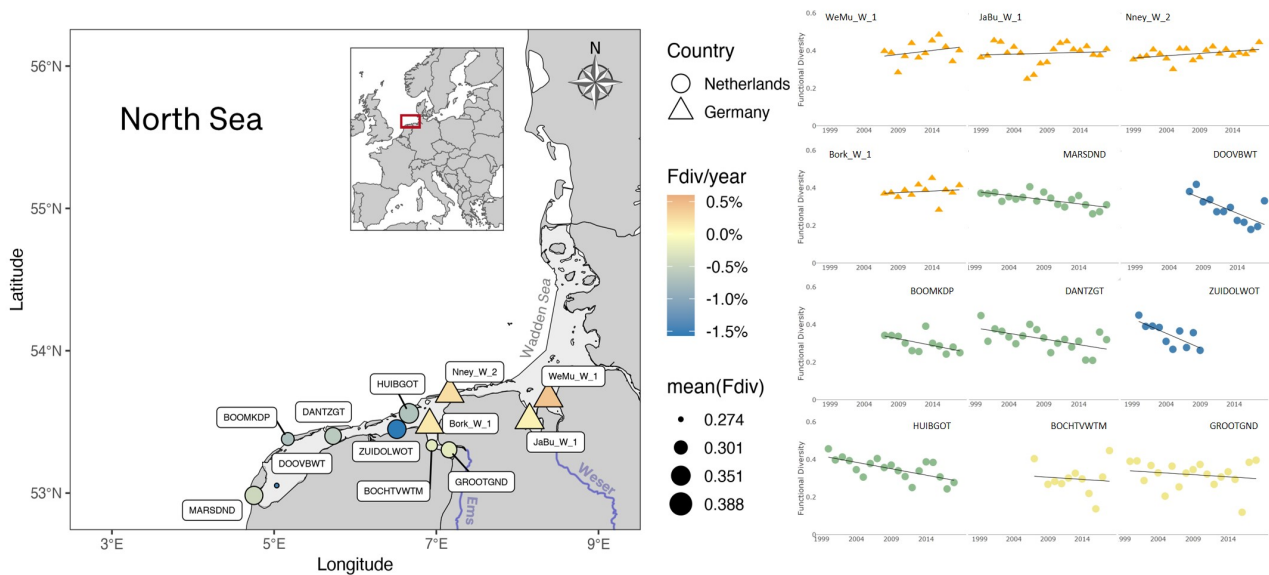


Figure 4.9: Phytoplankton functional diversity on Wadden Sea monitoring stations. A decrease in functional diversity (% Fdiv per year) is observed over the monitoring period at all Dutch stations (circles) whereas a mild increase (warmer colors) can be observed in the German stations (triangles). The fastest decrease rate (colder colors) is found at coastal stations on the Marsdiep and off Groningen. German Wadden Sea stations are in average the most functionally diverse (larger diameter). Reproduced from Carrasco and colleagues [68]; open access under the Creative Commons Attribution License - CC BY 4.0.

Once catalogued as a ‘Changed Ecosystem’ [95], the Wadden Sea experimented a consistent decreasing trend in eutrophication starting in the 1990s [60]. However, contrasting recent reports have found significant signs of increasing eutrophication, persistent algal blooms, and phytoplankton diversity alteration in the the West Wadden Sea [70, 91, 103, 200, 414, 436]. The declining diversity in the Marsdiep basin is likely explained by dominance of the *Phaeocystis globosa* spring and summer blooms [60, 103, 290]. The inter-annual variability among stations also suggest a blooming limitation by nutrients or light, which triggered the prevalence of fast-growing nutrient opportunist, C-strategist or R-strategist phytoplankton species such as *Micromonas pusila*, *Thalassiosira sp.*, *Chaetoceros sp.*, particularly in the second half of the last decade [347, 378, 442].

Stations at the Terschelling transect, in southern North Sea, also showed contrasting estimations of functional diversity between off-shore stations (TERS LG235 to TERS LG100) and in-shore stations (TERS LG50, TERS LG10 and TERS LG4). Whereas off-shore stations had no significant trend variation, the in-shore stations had a clear negative trend (Fig. 4.10). A possible explanation for this is the existence of a ‘line-of-no-return’ off the sand barrier islands of the Wadden Sea [329], which decreases the exchange between water masses and increases the

accumulation of suspended matter in the coastal zone [94]. Jung and colleagues [207] recently estimated this line somewhere between 10 and 100 km off the coastline. At the Terschelling transect, stations inside the ‘line-of-no-return’ will be highly influenced by the Wadden Sea dynamics and its particular environmental conditions. Therefore, the negative trend in functional diversity observed in the in-shore stations, as well as in the Rottumerplaat transect stations (ROTTMP), might be due to seasonal water exchange with the Wadden Sea phytoplanktonic community.

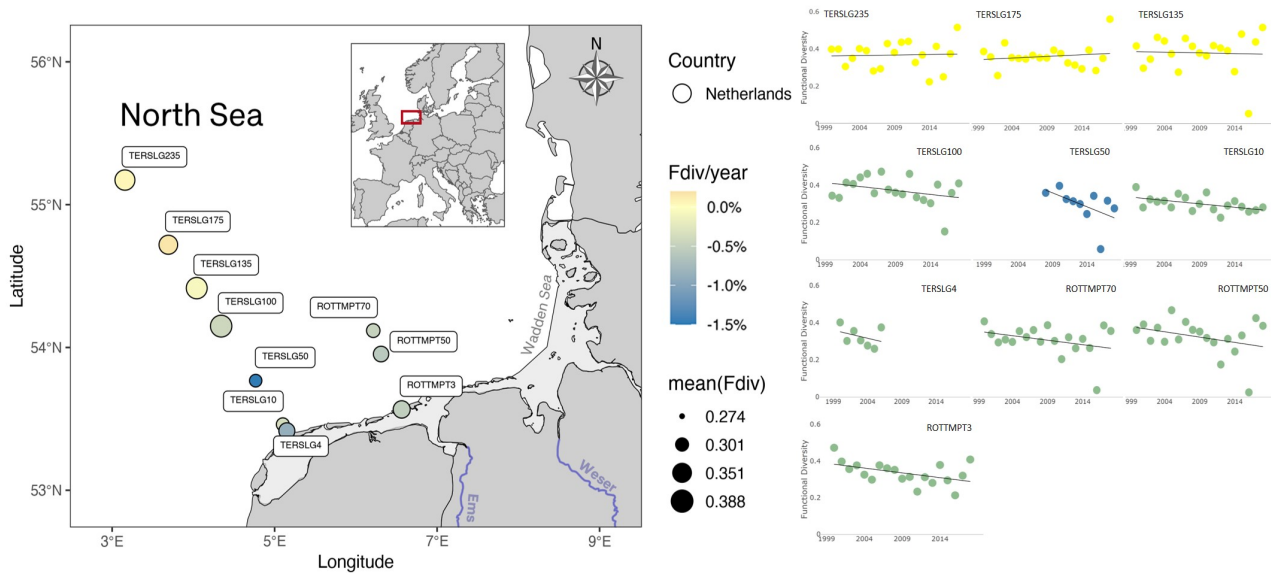


Figure 4.10: Phytoplankton functional diversity in southern North Sea monitoring stations off the Dutch sand barrier islands. Offshore stations (pale-yellow color) show no significant functional diversity trend (% Fdiv per year), contrary to those stations located closer to barrier islands, which show a mild decrease rate (colder color). Offshore stations are in average the most functionally diverse (larger diameter). Reproduced from Carrasco and colleagues [68]; open access under the Creative Commons Attribution License - CC BY 4.0.

Lastly, the estimation of functional diversity were consistent with expectations based on species composition. The low functional diversity in samples of 2006 and 2015 coincides with the dominance of the flagellate *Micromonas pusila*, with numbers over 90% of the total phytoplankton abundance (Fig. 4.9). Similarly, low values of functional diversity in Dutch off-shore waters is due to a major dominance of *Phaeocystis sp.*, whose numbers got to represent up to 99% of the total phytoplankton abundance in 2016 (Fig. 4.10). On the contrary, the period of increased functional diversity in German samples are due to the community being dominated by two to three species constituting together more than 50% of the total abundance. Among this species were *Lithodesmium undulatum*, *Paralia sulcata*, *Leptocylindrus minimus*, *Skeletonema costatum* and other diatoms. The number of non-dominant species with relative abundance less than 10% also increased.

## 4.5 Summary

In this chapter, a method to aggregate phytoplankton abundance datasets from different origins was introduced. By applying diffusion map to the newly aggregated time series, we are able to reconstruct the i-traits which best explain the variation in the community. This aggregated data improved the reconstruction of i-trait axes and the subsequent estimation of functional diversity from monitoring data. Our approach enables a robust estimation of functional diversity within the system based solely on species abundances.

We demonstrated that the failure of naive aggregation is rooted in the nature of the data collection and then exacerbated to the point of clustering those species unique to individual datasets, hence conflicting the trait reconstruction. If some species are not reported in a dataset, it can be assumed that these species were never present there or were not identified, but total certainty for any alternative is unlikely. Our approach to data aggregation avoids assuming a total absence of those no-reported species by averaging similarity values of only those species common to both datasets, obtaining a better reconstructed i-trait space.

The final result is a better estimation of functional diversity for both datasets and for the entire analyzed geographical area. Significant declining estimations of functional diversity in the West Wadden Sea are in line with recent reports [103, 200, 414, 436] and show the prevalence of fast-growing nutrient opportunist phytoplankton species in this ecosystem. Additionally, the difference in the functional diversity trends of the southern North Sea stations might be explained by the existence of a ‘line-of-no-return’ off the sand barrier islands of the Wadden Sea [207, 329] which isolated off-shore stations and their phytoplanktonic community.

## Chapter 5

# Scaling-up the aggregation of time series: Functional diversity of the North Sea phytoplankton community

In the oceans, ubiquitous microscopic phototrophs (phytoplankton) account for approximately half the production of organic matter on Earth [51]. These microcosmic organisms are the base of the marine food web. They are an integral part of the ecosystem, affecting trophic dynamics, nutrient cycling, habitat condition, and fisheries resources [304]. Furthermore, phytoplankton are closely coupled to environmental change [168, 347], making them sensitive indicators of environmental disturbance.

In the North Sea, phytoplankton is still the major contributor to biomass and primary production [66], in spite of the relevance of microphytobenthos production in the shallow tidal flat areas bordering the coasts of England, the Netherlands, Germany and Denmark, or the macroalgae crops along the coast of the U.K. and Norway [345]. However, changes in oceanographic conditions, together with human-induced activities are impacting planktonic species, their productivity, and possibly their capacity to thrive under new conditions.

Changes in phytoplankton of the North Sea can happen on different levels, i.e., new species, changes in dominance, productivity variations; with scientist even suggesting decadal-scale fluctuations linked to climate forcing [397]. Such long-term quantitative changes and trends in the phytoplankton community can only be detected using long-term datasets [93, 433]. But such long-term time series on phytoplankton composition are rare. What we often find are some isolated efforts, which have either focused on productivity estimation, or have had a short-time duration.

This lack of appropriate time series data has made it difficult to identify trends in phytoplankton dynamics and to establish linkages to natural variability or anthropogenic disturbance [115]. However, time series on phytoplankton composition do exist, although mainly restricted to a few stations at the coastline or in the nearshore coastal region [397] (e.g., Helgoland, Marsdiep, Norderney, Rijkswaterstaat monitoring programme). A feasible solution to address this issue is to aggregate those individual efforts into a single time series which overcomes the shortcomings of a low-spatial resolution and a short-time dimension [68, 266].

In Sec. 4.3, we introduced a methodology to successfully aggregate time series from different

origins, overcoming the possible discrepancies between datasets, using diffusion maps [68]. In this Chapter I apply this approach to aggregate multiple phytoplankton abundance datasets into a single time series. We start with an introduction to the phytoplankton datasets used for the analysis, and the particularities of their monitoring programmes (Sec. 5.1). In Sec. 5.2, we build the aggregation of datasets one by one, to evaluate the reconstruction of the i-trait space. In Sec. 5.3, using the best aggregated time series, I estimate the functional diversity variation of the North Sea phytoplankton community over time. Finally, I discuss the results in Sec. 5.4.

## 5.1 Phytoplankton monitoring programs in the North Sea

All individual datasets which have been used in this work were processed via the *Phytomerge* Julia programme [38]. *Phytomerge* was developed in this research to ensure the suitability of the aggregation process and the standardization of the datasets. The programme used the database *WORMS* [7] to extract updated phytoplankton taxonomic information. When a new dataset is submitted to *Phytomerge*, the programme uses *AphiaIDs* to perform a full processing of all of phytoplankton species within the dataset. Processing of the data included (i) the exclusion of instances of “double counting” where, for example, a taxon is included in both higher and lower taxonomic groupings within the same dataset and (ii) the correction or removal of taxa that have not been correctly recorded over the entire time period, for example, due to spelling mistakes or non-updated phytoplankton taxonomic names. Following the processing of the data, a master list is generated which can be used to check the submitted phytoplankton species, thus highlighting the final list of species to be included. Any phytoplankton species that does not count with a proper *AphiaIDs* could be included or not in the final dataset, e.g., for cases in which identification to the species level was not possible.

A total of seven long-term datasets were obtained and included into the diffusion map analysis. Each of them is sampled, maintained and curated by different research institutions. The focus of the assessment will be on the period 1999 to 2020, since most of the datasets covered such time period, in spite of minor gaps. Each monitoring programme analysed their samples for a full community of phytoplankton species using the Utermohl method on high-power inverted microscopes, unless indicated otherwise. Phytoplankton cell counts are abundance values reported in Cells per Litre.

First dataset included was collected in UK waters by the Centre for Environment Fisheries and Aquaculture Science (CEFAS) for the years 2001-2017 [278]. Phytoplankton samples have been collected using *SmartBuoys* at Dowsing, West Gabbard, Warp and Liverpool Bay stations. After processing with *Phytomerge* we obtained a dataset including 824 samples and 311 species. Second dataset included was collected by the Marine Scotland Science (MSS) in Scottish Waters for the period 2000-2020. Phytoplankton samples have been collected at stations Stonehaven, Scalloway, Scapa, Loch Ewe, and Millport (Clyde Sea), and reported as *Fields of View*, i.e., phytoplankton species abundance [53]. The processed dataset includes 3,932 samples and 179 species.

The third dataset was collected at the Helgoland Roads station by the Alfred Wegener Institute (AWI). This dataset has an extraordinary time resolution, as samples are collected 5 days a week since 1962 [432]. Our dataset covers the period 1998-2018, including phytoplankton

species which are confidently identified and certified by AWI scientists, i.e., 87 phytoplankton species and a total of 8,401 samples. The fourth dataset was collected by the Flanders Marine Institute (VLIZ) in the Belgian Exclusive Economic Zone using a continuous recording, or ship of opportunity, covering the period 2010-2016 [1]. Processed dataset includes 75 phytoplankton species and 513 samples. The fifth dataset was collected by the Havforskninginstituttet, covering the time period 2017-2022. The samples are collected from a CTD and no pre-filtration was performed [134]. The processed dataset includes 141 phytoplankton species and 114 samples.

The two last datasets were already introduced in Chapt. 4. One dataset was gathered at 04 stations in the German Wadden Sea, and was conducted by the Lower Saxony Water Management, Coastal Defence and Nature Conservation Agency (NLWKN), in Germany for the period 1999 to 2018 [291]. We used a processed dataset including 1,664 samples and 249 species. The last dataset corresponds to the extensive monitoring program conducted by Rijkswaterstaat, in the Netherlands [20]. The programme collected data in 18 stations spread over the Dutch Wadden Sea and the Dutch area of North Sea. The processed dataset includes 3,691 samples and 366 species.

## 5.2 Reconstructing the functional i-trait space

We use the Phytomerge programme to perform the aggregation of datasets following the algorithm introduced in Sec. 4.3 for a successful aggregation [68]. Then, we diffusion map the aggregated dataset. To illustrate the effect of datasets topology on the diffusion maps results, we start by aggregating datasets with distinct characteristics, i.e., lower number of species and high amount of samples, or the opposite. Subsequently we will aggregate one by one the datasets to assess the variation in the reconstructed i-trait space as we increase the amount of data available for the diffusion map analysis.

First diffusion map analysis was performed on the aggregated datasets of the Centre for Environment Fisheries and Aquaculture Science (CEFAS) and the Marine Scotland Science (MSS) (Fig. 5.1(a)). These datasets have opposite topologies. While CEFAS data has high number of phytoplankton species and low amount of samples, the MSS dataset includes almost half the amount of species than the CEFAS dataset, gathered in almost 4,000 samples. Second diffusion map was done with the aggregated datasets of NLWKN and the Helgoland Roads (Fig. 5.1(b)). These datasets also have opposite topologies, with NLWKN having three times the amount of phytoplankton species than Helgoland Roads data. However, Helgoland Roads has double the amount of samples compared to the NLWKN dataset.

The reconstructed i-trait space for both diffusion maps look quiet different, despite the aggregated datasets for both cases display similar differences. Noticeable is the difference in symmetry for the second diffusion map (Fig. 5.1(b)), which is highly skewed to the negative side of EV1. While the first map is more symmetric, it spreads over a smaller area (Fig. 5.1(a)). Furthermore, the first map counts with a higher number of phytoplankton species, with a total of 320 phytoplankton species, whereas the second map has a total of 259 species. Nevertheless the higher amount of species, the second map explains better the variation in the dataset, as can be deduced from the ranked inverse eigenvalue ( $1/\lambda$ ) (Table. 5.1).

Diffusion map is a method which delivers better results the more data is fed on [21, 359]. But datasets come in different topologies, something intrinsic to the geographical area in which the

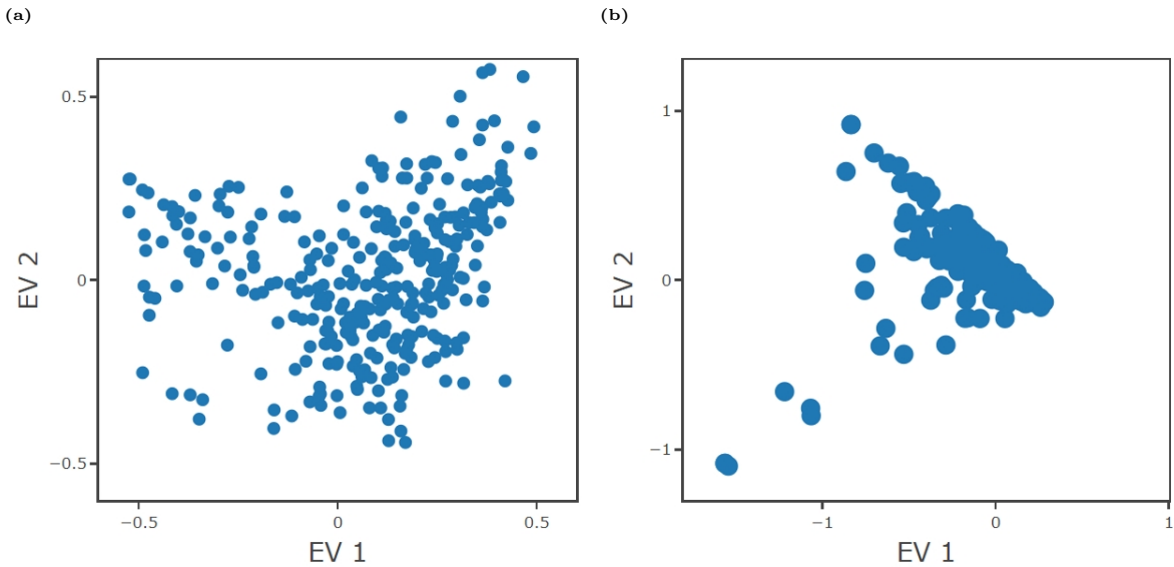


Figure 5.1: Diffusion map displaying the reconstructed i-traits that explained best the variation of the aggregated CEFAS and MSS dataset (a) and the aggregated NLWKN and Helgoland Roads dataset (b). Blue dots represent the phytoplankton species located in the reconstructed i-trait space.

monitoring takes place, i.e., number of species, and the capacity of the monitoring program in terms of the frequency and duration of sampling. The nature of this datasets will influence the quality of the reconstructed i-trait space, therefore, the best solution is to aggregate datasets in order to balance the shortcomings of individual datasets. The question that arises then is, how much data is enough? To approximate an answer, we will evaluate the reconstructed i-trait space performing multiple diffusion map analysis over the aggregated datasets, adding one dataset into the analysis at a time.

Table 5.1: Table with the ranking of diffusion maps first variable. Each ranking value ( $1/\lambda$ ) is shown according to the aggregated datasets used for the analysis.

Aggregated datasets	Rank ( $1/\lambda$ )
CEFAS, MSS	4.4583
NLWKN, HRoads	5.1546
NLWKN, HRoads, FLM	5.1759
NLWKN, HRoads, FLM, CEFAS	6.0096
NLWKN, HRoads, FLM, CEFAS, NOR	5.5066
NLWKN, HRoads, FLM, CEFAS, NOR, MSS	5.3850
NLWKN, HRoads, FLM, CEFAS, NOR, MSS, Rijk	6.2853

Since the aggregation of the NLWKN and Helgoland Roads datasets delivers better results (Table. 5.1) we will continue the aggregation of datasets from here. The third diffusion map (Fig. 5.2(a)) includes the two previous datasets and the Flanders dataset (FLM). The shape of the reconstructed i-trait space changes significantly, having a reduced skewness in the first variable. However, the ranked inverse first eigenvalue ( $1/\lambda$ ) indicates a similar variation was gathered by this variable when compared with the previous diffusion map (Table. 5.1).

For the subsequent diffusion map analysis we use the previously aggregated dataset and progressively add one dataset at the time. Thus, we will first add the CEFAS dataset (Fig. 5.2(b)), then the CEFAS and Norway dataset (Fig. 5.2(c)), and lastly the CEFAS, Norway and MSS dataset (Fig. 5.2(d)). For this 3 diffusion maps, the i-trait space preserves a similar triangular

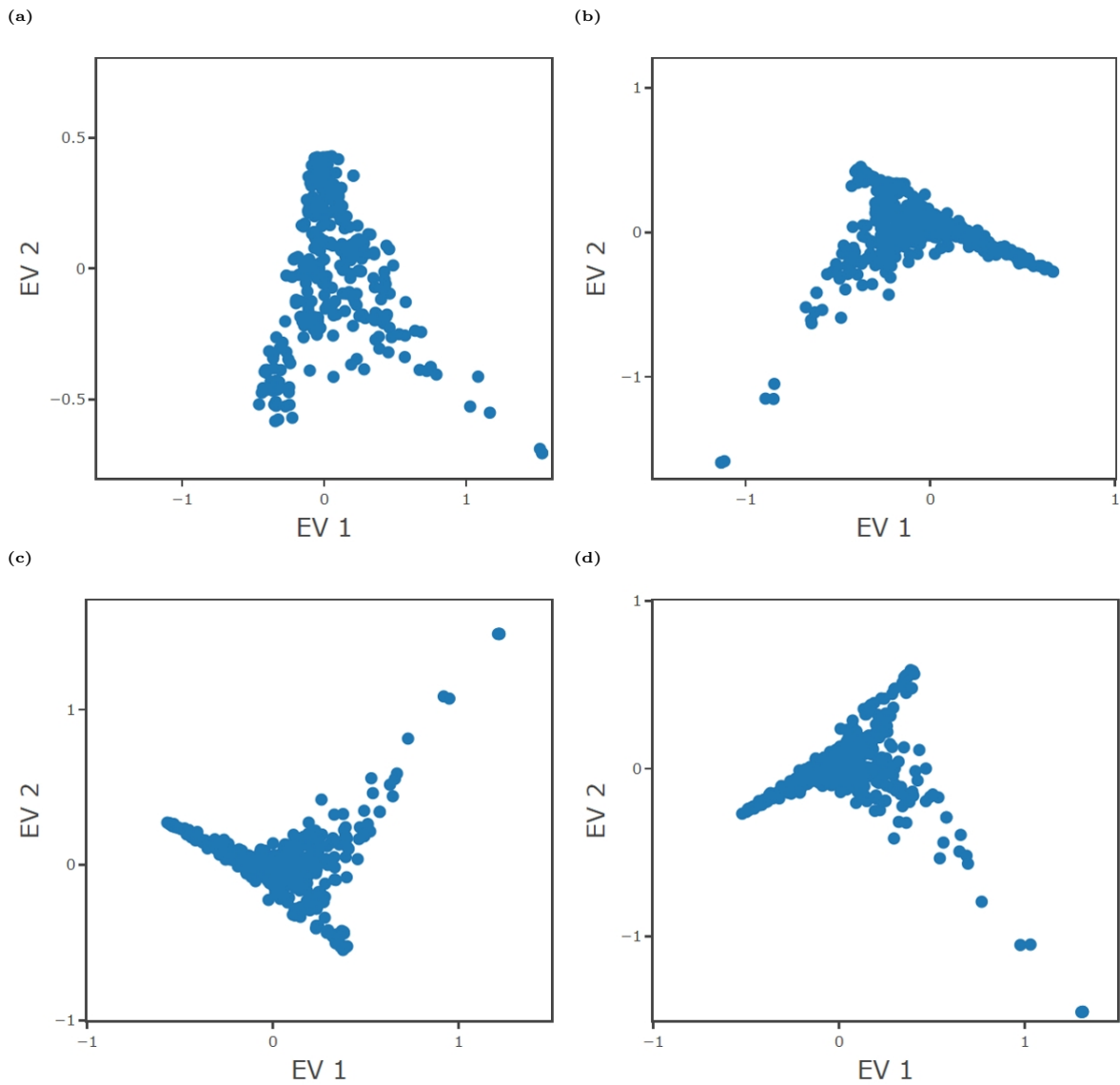


Figure 5.2: Diffusion map displaying the reconstructed i-traits that explained best the variation of the aggregated NLWKN, Helgoland Roads plus the Flanders dataset (a). Subsequent maps are based on the previous aggregated dataset plus the CEFAS dataset (b), plus CEFAS and Norway dataset (c), and finally plus CEFAS, Norway and MSS dataset (d).

shape with an elongated branch, producing skewness for the first variable (EV1). According to the ranked inverse first eigenvalue ( $1/\lambda$ ) however, the best results are given in the diffusion map four, with aggregated datasets of NLWKN, Helgoland Roads, Flanders and CEFAS (Table. 5.1). Subsequent dataset aggregation generated a slight reduction in the EV1 capacity to explain the dataset variation (Table. 5.1).

Finally, the Rijkswaterstaat dataset (Rijk), is aggregated into a single long-term dataset of phytoplankton abundances, formed by seven datasets which were gathered by seven distinct monitoring programmes (Fig. 5.3). The shape of the reconstructed trait space remains in its triangular shape, having the elongated branch reduced, but keeping a skewness for the first variable. However, according to the ranked inverse first eigenvalue ( $1/\lambda$ ) we obtain the best results so far [147], with the EV1 gathering most of the variation in the reconstructed i-trait space (Table. 5.1).

Having no environmental data available for all the samples within the datasets, it is not possible to compute a weighted average of each environmental parameter (Eq. 3.1), to assess the nature of the EV1. As seen in Sec. 3.2.3, another approach is to verify the extreme values and



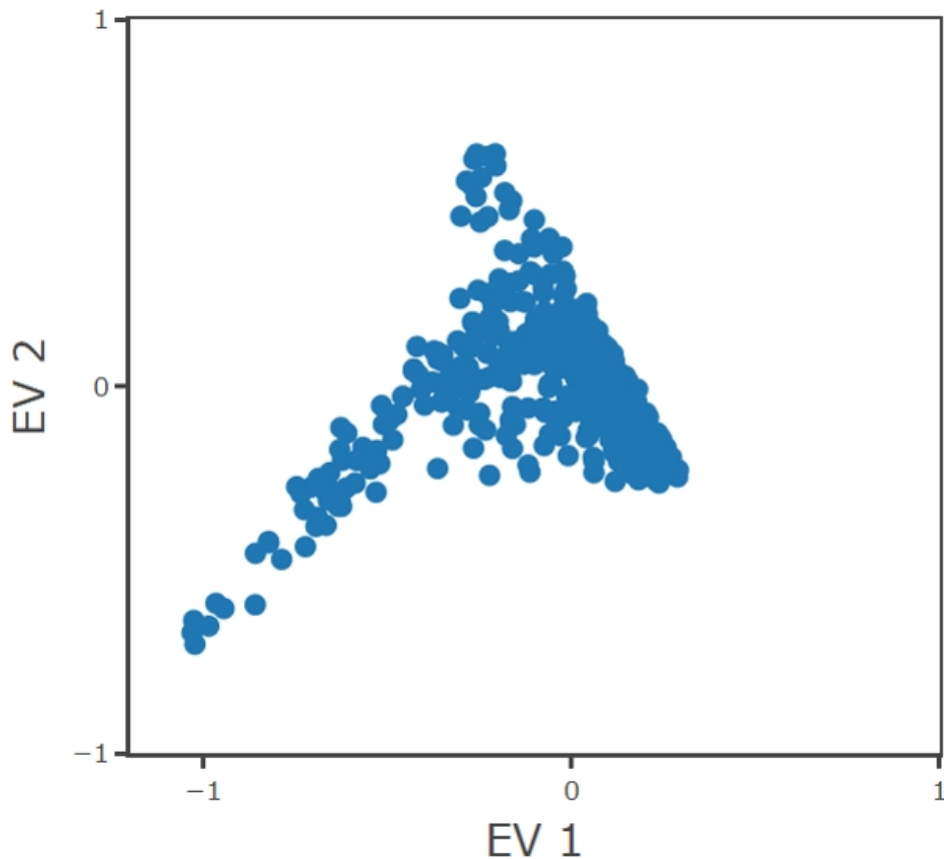


Figure 5.3: Diffusion map displaying the reconstructed i-traits that explained best the variation of the aggregated NLWKN, Helgoland Roads, Flanders, CEFAS, Norway, MSS, and Rijkswaterstaat datasets.

check for biological or ecological characteristics of the associated phytoplankton species. When doing this, the species associated to the negative extreme of EV1 are some cyanobacteria (e.g., *Aphanizomenon sp.*, *Pseudanabaena sp.*, and *Limnothrix sp.*); chlorophytes (e.g., *Closterium sp.*, *Koliella sp.*, and *Gloeotila sp.*); and small pennate diatoms (e.g., *Synedra acus f. acus*, *Melosira varians*, and *Aulacoseira sp.*). Three different types of phytoplankton, which are usually reported in freshwater or brackish environments, and that in our dataset are present only in the Wadden Sea area. The positive side of EV1 is associated to dinoflagelates (e.g., *Amphidinium longum*, *Dinophysis hastata*, *Scrippsiella trochoidea*, and *Protoperidinium breve*); and marine chain-forming diatoms (e.g., *Chaetoceros throssenii*, *Thalassiosira conferta*, and *Guinardia blavyana var. blavyana*). Both of these groups are strict marine species that were reported only in the CEFAS or the MSS datasets.

### 5.3 Functional diversity of North Sea and Wadden Sea phytoplankton community

Using the reconstructed i-trait space obtained after diffusion mapping the aggregated seven long-term datasets, we will now assess the functional diversity in each sample, by means of the Rao's quadratic entropy [342], and following the approach introduced in Sec. 4.4. The outcome obtained is a time series with a functional diversity score per sample. Short datasets, with less than 7 years duration, such as the Flanders dataset (FML) or the Norwegian dataset (NOR),

as well as sampling stations within datasets with less than 7 years monitoring time, won't be considered for this assessment.



Figure 5.4: Phytoplankton functional diversity variation on the CEFAS U.K. monitoring stations of Warp, West Gabbard, Liverpool Bay and Dowsing (stars); as well as in the Scottish MSS monitoring station of Stonehaven, Scalloway, Scapa, Millport and Loch Ewe (dots). The black line represent a linear regression trend which runs over the yearly functional diversity values at each station. A decrease in functional diversity is observed over the measurement period at Liverpool Bay, whereas a mild increase can be observed in Scottish stations of Scapa and Loch Ewe.

At the CEFAS monitoring stations, functional diversity in the Warp and West Gabbard stations had a increasing trend, whereas in the Liverpool Bay station, a negative trend is evident during the time period 2001-2017 (Fig. 5.4). Inter-annual oscillations are evident in this stations, in contrast to what can be seen at the Scottish MSS monitoring stations. These stations displayed functional diversity values with low oscillations, and an almost stationary trend for the period 2000-2020. Only the stations Scapa and Loch Ewe presented a slight increasing trend in annual functional diversity values (Fig. 5.4). Additionally, it is quite evident the higher values in functional diversity found at CEFAS stations compared to the lower values at the Scottish stations.

When we assess the functional diversity in the Wadden Sea basin, major trends have remained similar to those obtained in Chapt. 4 [68], despite some minor changes in functional

diversity inter-annual values at German stations (Fig. 5.5). Two stations, JaBu\_W\_1 and Nney\_W\_2, show a mild, non-significant negative trend, whereas station WeMu\_W\_1 displays now a significant gain in functional diversity. Functional diversity at Dutch stations have experienced decreasing trends (Fig. 5.5). Among the stations with the highest losses we have ZUIDOLWOT, DOOVBWT and DANTZGT. Oscillations in functional diversity values are present in all stations, being more evident in quinquennial periods.

A similar scenario with functional diversity is displayed in the southern North Sea basin, with major trends similar to those obtained in Chapt. 4 [68] in spite of minor changes in inter-annual values for some stations (Fig. 5.6). One station, TERSLG175, shows a significant negative trend, whereas the Helgoland Roads station displays a significant gain in functional diversity. For the rest of off-shore stations, functional diversity has experienced either mild decreasing or mild increasing trends (Fig. 5.6). In general, it can be considered that the southern North Sea area has had a stationary trend during the period 2000-2018. Less pronounced oscillations in functional diversity values were present at these stations, especially when compared to those located in the Wadden Sea.

Although the spatial resolution covered by monitoring stations is not ideal to draw a strong conclusion about patterns of functional diversity variation on the entire North Sea, we can draw tentative conclusions for some of the areas covered by the monitoring datasets. In one hand, it is clear that the West Wadden Sea has the higher losses in functional diversity, together with the Liverpool Bay area and few off-shore stations at the southern North Sea (Fig. 5.7). On the other hand, clear increases in functional diversity can be seen at the German Bay (Helgoland Roads) and in the East coast of the UK (CEFAS stations). Finally, most of the Scottish monitoring stations, as well as the southern North Sea tend to have an stationary trend (Fig. 5.7).

## 5.4 Discussion

In this chapter, I aggregated seven phytoplankton abundance datasets of the North Sea basin gathered by different monitoring programmes. Applying the dataset aggregation approach, which was introduced in Chapt. 4 [68], I was able to explore the effects of dataset aggregation on the quality of the reconstructed i-trait space. It is evident that for diffusion maps the more data the better, but the topology of datasets is also relevant. Whether our dataset is crosssectionally larger (i.e., number of phytoplankton species larger, or almost similar, than the number of samples), or crosssectionally wider (i.e., number of phytoplankton species smaller than the number of samples) the diffusion map delivers results successfully but the shape of the reconstructed trait space varies.

Diffusion mapping two crosssectionally different datasets such as CEFAS and MSS reconstructed a well covered i-trait space, in spite of the short area it covered. Such result might be intrinsic to the community under study, since the total number of species (320 phytoplankton species) was high enough to provide trustworthy results. However, the number of samples seems to be more relevant for an adequate i-trait reconstruction. Diffusion maps obtained from datasets with high number of samples, such as the Helgoland Roads dataset, covered wider i-trait space, gathering more of the variation in the community in the first variable, thus providing slightly better results.

Ecological datasets have always struggle with such issues. The number of species that can

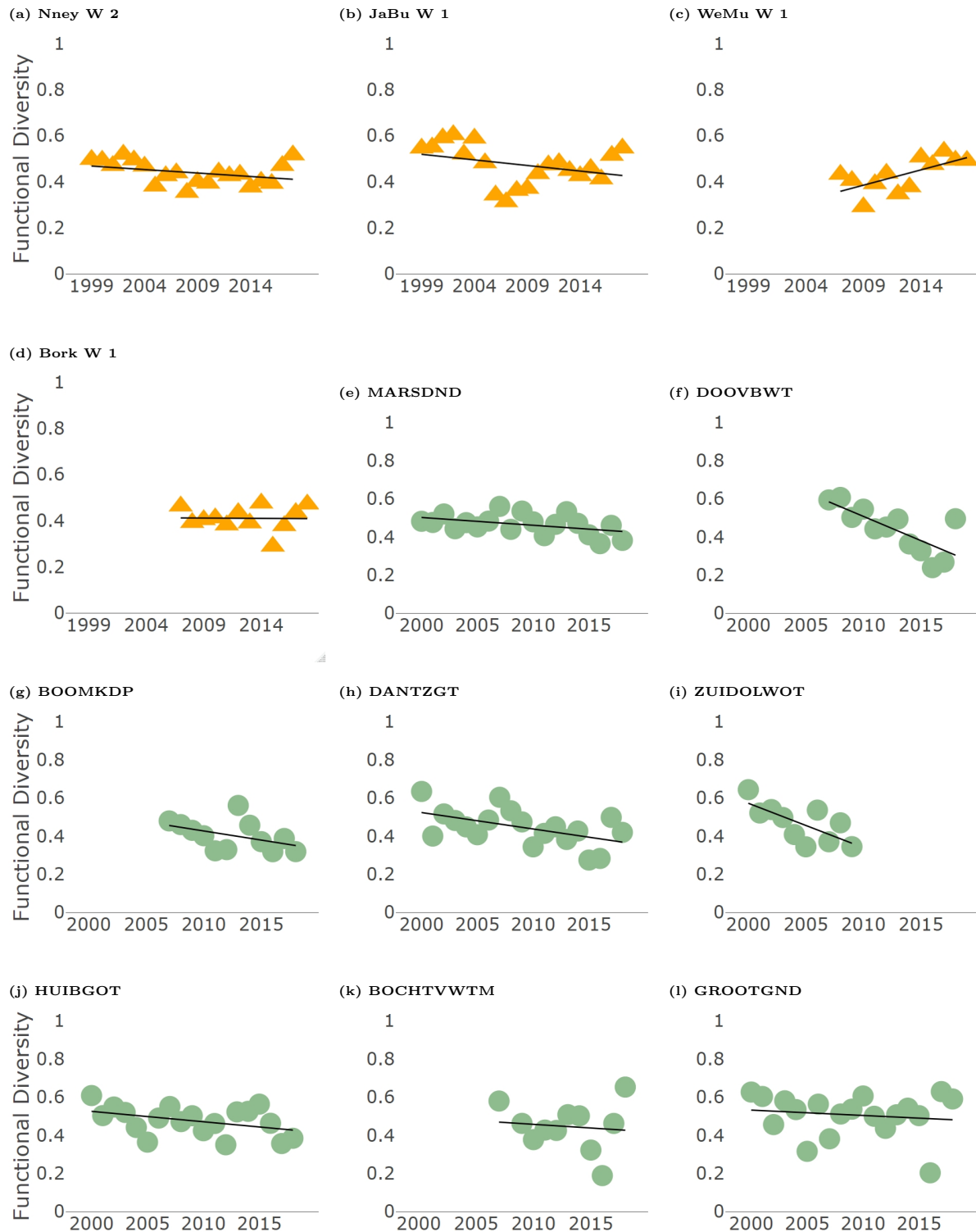


Figure 5.5: Phytoplankton functional diversity variation on the Wadden Sea basin, including stations of the NLWKN (triangles) and Rijkswaterstaat (dots) extensive coastal monitoring programmes. The black line represent a linear regression trend which runs over the yearly functional diversity values at each station. A mild decrease in functional diversity is observed at JaBu\_W\_1 and Nney\_W\_2, whereas there is a mild increase in WeMu\_W\_1 in the East German Wadden Sea. Most of the stations at the West Dutch Wadden Sea displayed a decreasing trend.

be gathered in a sample is highly dependent in sampling effort, spatial resolution, local diversity and the targeted group of organisms [279, 297]. Sampling fish in two sampling stations will surely provide a reduced number of species, no matter the time extension of the monitoring programme. However, the same sampling design, targeting a phytoplankton or a bacterial com-



Figure 5.6: Phytoplankton functional diversity variation on the Southern North Sea basin, including stations of the Rijkswaterstaat (dots) and of the Helgoland Roads (triangles) extensive monitoring programmes. The black line represent a linear regression trend which runs over the yearly functional diversity values at each station. A significant decrease in functional diversity is observed at TERSLG175, whereas a significant increasing trend is observed at Helgoland Roads. Most of the off-shore stations displayed either a mild decreasing or increasing trend. Oscillations at this stations are less pronounced than those in the Wadden Sea basin.

munity could provide a representative amount of species, if monitoring is adequately extended and frequency is consistent. In this sense, we must consider carefully those conclusions obtained from a diffusion map of aggregated short-term datasets, or datasets with a low sampling frequency. As results have showed, it might be better to aggregate at least one dataset with large number of samples.

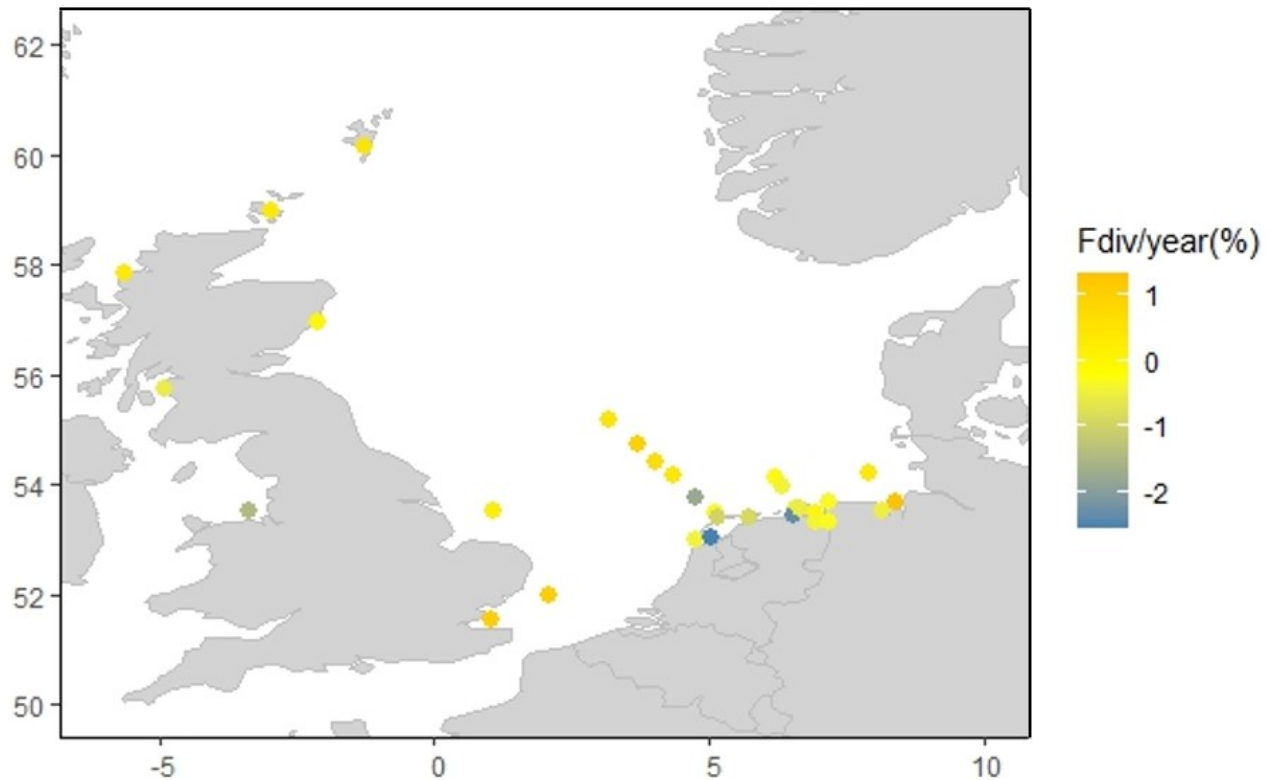


Figure 5.7: Phytoplankton functional diversity on the North Sea and Wadden Sea. A decrease in functional diversity ( $\Delta F_{div}$  per year(%)) is observed over the measurement period at most Dutch stations at west Wadden Sea, whereas a mild increase (warmer colors) can be observed in some CEFAS stations, the Helgoland Roads and few off-shore stations in southern North Sea. The fastest decrease rate (colder colors) is found at coastal stations on the Marsdiep and off Groningen, and at the Liverpool Bay.

Another relevant question addressed here was the issue of the amount of data needed. This can be understood from the perspective already touched, i.e., the topology of datasets, or from the number of datasets that can be aggregated. In terms of the area covered by the reconstructed trait space, it did not grow larger after the aggregation of the fourth dataset (CEFAS dataset). This means that the added species did not fill a different space of the reconstructed i-trait, but rather the species were functionally similar to the ones already present in the aggregated dataset. Evidence tend to suggest the former. When aggregating the fifth (Norwegian dataset) and later the sixth (MSS dataset) into the diffusion map, the i-trait space not only did not grow larger, but the first variable decrease in quality, thus gathering less of the variation in the community.

The latter could be related to a certain degree of endemism in the phytoplankton species within those datasets [115, 133, 345]. Up to that point the aggregated datasets only monitored the Wadden Sea basin and some parts of the southern North Sea. The marked differences in the North Sea oceanographic conditions [124], might influence the phytoplanktonic community composition and the dominance of certain species. In one hand, the shallow, well-mixed, and high nutrient concentrations of the southern region, becomes ideal for fast-growing planktonic species, including some microphytobenthos diatoms which become suspended in the water column [96, 97]. In the other hand, the seasonally stratified and nutrient-limited in the surface mixed layer of the central and northern region, favors strict marine phytoplankton, larger cells and chain-former diatoms [244].

Another possible explanation is the presence of more off-shore sampling stations in the

Norway and MSS datasets. Most of the phytoplankton biomass develops in coastal areas [245], where nutrients are abundant and water column is well mixed, thus allowing faster development of phytoplankton blooms [200]. Therefore, the species gathered in off-shore stations tend to be different to those in coastal areas, due to faster changing oceanographic conditions and limiting nutrient availability [207, 369]. This explanation is reinforced by the increase in performance of the first variable when the Rijkswaterstaat dataset is aggregated in the analysis. The Rijkswaterstaat monitored the Terschelling transect, with seven stations extending up to 235 Km off-shore, and the Rottumerplaat transect, with three stations located at 3, 50 and 70 km from the West Frisian barrier islands.

In spite of the lack of environmental factors to evaluate the nature of the diffusion map variables, the species associated to extreme values in the i-trait space suggested a marked difference between coastal-brackish and marine phytoplankton species. In the negative extreme of EV1 we have three cyanobacteria, three chlorophyte and three small pennate diatom. All of this phytoplankton are associated to high concentrated nutrients in water, and water-mass with low salinity, characteristics of the Wadden Sea and coastal areas [60, 379, 420]. On the contrary, in the positive extreme we find four dinoflagellates and four marine chain-forming diatoms, species that are characteristic of marine waters [180]. This conclusion is supported by the results of Chapt. 4, in which the aggregation of NLWKN and Rijkswaterstaat datasets showed a phytoplanktonic community driven mainly by the different oceanographic conditions of the Wadden Sea and the southern North Sea [68].

Nevertheless the tentative conclusion, the recovery of environmental data associated to each sample in the dataset will be a major issue to solve for the future of this research. For the second most relevant variable, it was not possible to identify a key ecological characteristic in the species associated to extreme values, hence, no hypothesis can be drawn at this stage. Another limitation is the low-spatial coverage reached up to this point. Although this might be an issue intrinsically linked to the phytoplankton monitoring design [297, 302], innovative monitoring programmes, such as the Continuous Plankton Recorder (CPR) [348], will offer a feasible solution to increase the spatial resolution in the North Sea off-shore area, thus improving the results of the diffusion map.

In terms of the functional diversity assessment, no major conclusion can be drawn at the spatial scale, although similar trends were found for the Wadden Sea as those found in Chapt. 4 [68]. In particular the west Wadden Sea had showed major losses in functional diversity. Such results are supported by Di Carvahlo and colleagues [103] who working with Wadden Sea data from NLWKN and Rijkswaterstaat found that the Dutch stations have loss around 20 % of phytoplankton species, or 17.1 species per decade. In the German stations, however, there is an increasing trend, corresponding to a gain of 13.5 species in a decade [103]. For the German stations, our results showed that after a progressive decline in the first half of 2000s decade (evident in stations JaBu\_W\_1 and Nney\_W\_2), there was a recovery of functional diversity up to the end of the monitoring period.

Other stations with clear gains were the CEFAS stations, in particular those located in the southeast British Coast (Warp, West Gabbard, and Dowsing ), as well as the Helgoland Roads. It is quite interesting to point out that CEFAS stations had in average the highest values of functional diversity, whereas Helgoland Roads had the lowest. The geographical location of CEFAS stations, at the end of the Dover Channel could favor the mixing of their phytoplankton community with the Atlantic water masses entering into the North Sea [55, 56], thus, enhancing the phytoplankton diversity around this region. For the case of Helgoland

Roads no clear answer is suggested. However, it might be necessary to add a raw dataset, which includes all phytoplankton species which were identified per sample, with the purpose of discarding that the functional diversity results for Helgoland Roads are an artefact of the dataset used. Even though taxonomic accuracy is quiet relevant for any diversity analysis [388], exclusion of species might be detrimental for the quality of results [327].

In summary, the aggregation of long-term monitoring datasets has improved the results obtained from diffusion maps, allowing a better reconstruction of the i-trait space, and a trustworthy assessment of functional diversity in individual samples. This approach allows us to translate complex ecological information store in time series into a straightforward value which can be interpreted and compared within the region of study. In the future, to obtain a higher spatial resolution coverage, new datasets might need to be aggregated into the analysis. Given that individual time series are distributed across different areas and managed by different countries, open collaboration with national institutions managing the time series is essential [210, 297].



## Chapter 6

# Conclusions

Endings are always enlightening. An end is the point at which something no longer continues to happen. An end is also a goal or purpose. Both meanings imply dimensions; a physical, a temporal, a moral, an ethical. The end of this research was to understand biodiversity from a functional perspective, in an ecological, spatial and temporal dimension. I focused on ecological non-individual entities, like species, ecosystems, and on the processes which promote and maintain those entities [299]. The concern are "the wholes". But the path to know, is via its constituents [409]. The means to this end implied to use novel approaches which consider the complexity of ecological communities. Diffusion maps provide such possibility. On the one hand, this approach have enabled us to make sense of large datasets of ecological communities, while on the other hand, we have not dismissed the interactions of the species forming them. Nevertheless, in science there is always room for improvement.

In this work, I focused on assessing the feasibility of this approach, and filling in the knowledge gaps which remain, specially in terms of its application to ecology. Consequently, I have applied diffusion maps to ecologically and biologically more complex organisms, as the fish demersal community from the North Sea. Moreover, the limitations imposed by the scarcity and adequacy of ecological long-term datasets, led to propose datasets aggregation as an adequate solution. Having an appropriate methodology to aggregate datasets, to scale-up the analysis was the right step to take. Therefore, I last focused on trying the limits of diffusion maps when aggregating datasets from different North Sea research institutions. In this section, I outline the main conclusions of this work, at the time I give a brief outlook for future research directions.

### **Diffusion map making sense of demersal fish community**

Diffusion map enabled us to make sense of high-dimensional datasets of demersal fish communities by coarse-graining the 229 fish taxa in terms of their functional i-traits, providing a quantitative framework to reconstruct an i-trait space, a potential niche space over time for fish of the North Sea. With this approach, we achieved a dimensionality reduction in terms of the most important variables, which for this community were 'depth' and 'year of higher abundance'. These results are not a mere product of interpretation, but rather have a reality which is independent of these labels. Depth has been referred in multiple research as an important determinant of fish species distribution and community structure in the North Sea [36, 64, 113, 135, 346]. The year of abundance is correlated to a decreasing abundance trend of some boreal

fish species, while increasing abundance of multiple lusitanian fish species. The North Sea has been losing temperate-incline species in favor of warm-inclined fish from lower latitudes [8, 118, 315].

This is one of the earliest application of diffusion maps to fish communities, a biologically and ecologically more complex group of species. Remarkable was how the addition of the concept of size at maturity, a proxy to differentiate juveniles from adults in fishes [405], provided better results when reconstructing the functional i-trait space. However, a similar effect was obtained when adding the datasets from the Scottish West Coast Bottom Trawl Survey (SWC-IBTS) and The Irish Groundfish Survey (IGFS). This suggests that, equally important as to have more data it is to incorporate relevant ecological characteristics which capture changes in spatial distribution of the taxa under analysis. Lacking this information would not totally degenerate the quality of results but might cause inadequate interpretations. Application of diffusion maps in long-timespan species, such as some invertebrates, birds or mammals, would benefit from the incorporation in the analysis of migration patterns, habitat segregation, or reproductive preferences.

### Aggregation of datasets as solution to data limitations

An important limitation, which diffusion map has for its application in ecology, is the availability of data on the ecosystem we want to investigate. Previous diffusion map applications have used bacteria genomic time series [127, 267], phytoplankton biomass time series [359], and in this work, demersal fish abundance time series. However, we need a certain amount of data to ensure the applicability of the diffusion map. Massing suggested that this method should not be applied to datasets with less than ca. 30 samples for the case of marine bacteria communities [266]. In the case of phytoplankton communities, it is more relevant to see the amount of species, which should be ca. 100 species. However, for the case of long timespan organisms, i.e., fish, birds or mammals, the focus should also be in the temporal resolution of the dataset. Having an adequate number of species will matter little if we have a time series with a temporal resolution lower than the average timespan of the targeted species group.

We have proposed an approach to address this issue, one that will help us to obtain adequate time series, the aggregation of datasets from different origins. Although it might seem a trivial task, a direct aggregation is not possible due to likely inconsistencies between datasets. These inconsistencies are the outcome of differences in the equipment, personnel expertise, and procedures employed by different sampling teams [266], which results in a successful, or not, taxonomic identification of species. Our approach does not neglect that whether a species is observed or not in a given sample, is an indicative of either the actual absence of the species, or a signal that the species, while objectively present, was not identified or was assigned a different name [231, 316].

Applying our approach to aggregate two datasets, one gathered by Rijkswaterstaat [19] and another by NLWKN [291], provided a better reconstruction of the i-trait space than when datasets were analyzed independently. Diffusion maps allowed the dimensionality reduction in terms of the most important variable, which for this community was ‘salinity’ and ‘dissolved inorganic nitrogen’. A clear signal obtained from the diffusion maps analysis was the marked influence that water masses have on the structuring of phytoplankton communities. Moreover, assessment of functional diversity was also consistent with reports of significant declining es-

timations of functional diversity in the west Wadden Sea [103, 200, 414, 436] and showed the ever prevalence of fast-growing nutrient opportunist phytoplankton species in this ecosystem.

### **From local to regional: Scaling-up the aggregation of time series**

Assessment of functional diversity at regional scales requires a wider spatial coverage, an effort that few monitoring programmes can afford. Ecological time series such as the the ICES coordinated International Bottom Trawl Survey (IBTS) are extraordinary [195], both in the spatial scale covered and the duration it holds. More often what we find are local or countrywide efforts that sustain observations over periods of few decades. The last part of this work, gathered seven long-term phytoplankton abundance time series, aggregated them into a single long-term dataset, thus providing a starting point to move from local to regional assessment of functional diversity by means of datasets aggregation.

It is important to remark that datasets will have different topologies, with either a high number of species and low amount of samples (i.e., crosssectionally wider), or a low number of species gathered in a large number of samples (i.e., crosssectionally larger). In order to test the implications of such differences, it is recommended to perform a gradual aggregation of datasets. This will allow the researcher to assess the variation in the reconstructed i-trait space as well as its capacity to explain community variation. Based on this experience, it is recommended to aggregate at least one dataset with large number of samples.

Another important insight is that aggregation of geographically close datasets is preferable, as aggregation of datasets from very distant geographical areas might decreased the quality of the i-trait space reconstruction, or the quality of the first variable. Having aggregated mostly southern North Sea datasets, the aggregation of the Norwegian dataset reduced the capacity of the first variable to explain the variation in the community. Moreover, if we seek for a regional assessment, the best is to avoid large gaps in the data spatial resolution.

At this point the spatial coverage of the southern North Sea looks quiet optimal, but there remain large areas of the central and northern North Sea in which no datasets were obtained. The natural solution is to find and aggregate other monitoring programmes from those areas. A feasible solution under the current positive trends for international collaboration and initiatives to open data to research. Long-term programmes, such as the Continuous Plankton Recorder (CPR) [302], are a great option to increased the spatial coverage. Moreover, it is necessary the incorporation of more off-shore monitoring points, a regular weak point for many marine phytoplankton monitoring programmes, that have focused mainly on coastal areas [397].

At the end of this work, the usage of diffusion maps in ecology and functional diversity assessment is just beginning. There remain still to explore the applicability of this method in other complex organisms, not just biologically but also in terms of their ecological behavior and their population dynamics. Moreover, the application of diffusion maps to mixed group of organisms is a tentative option. Important collaborations were developed in the past years to gather and organize ecological time series of a variety of biota in online servers, such as BioTIME [109] or CRITTERBASE [394], thus providing a unique opportunity. Expanding the boundaries of this novel approach is also important, either by making use of yet unexplored sources of data (e.g. eDNA), which could expand the scale for datasets aggregation, or by incorporating other tools, such as machine learning, that could lead to universally accepted

trait spaces for important groups of organisms. A final goal is to develop an online engine which, powered by diffusion maps, is able to measure functional diversity in new samples of a well-assessed region.

# Bibliography

- [1] (VLIZ), F. M. I. “LifeWatch observatory data: phytoplankton observations by imaging flow cytometry (FlowCam) in the Belgian Part of the North Sea.” In: *VLIZ, Ostende. Belgium* (2022).
- [2] Aas, Ø. et al. *Atlantic salmon ecology*. John Wiley & Sons, 2010.
- [3] Aber, J. D. and Melillo, J. M. *Terrestrial ecosystems*. Vol. 429. Saunders College Publishing Philadelphia, 1991.
- [4] Abou-Seedo, F. and Potter, I. “The estuarine phase in the spawning run of the river lamprey *Lampetra fluviatilis*”. In: *Journal of Zoology* 188.1 (1979), pp. 5–25.
- [5] Ahern, M. et al. “State of the Ocean Report 2024.” In: (2024).
- [6] Ahmed, A. S. et al. “The distribution and composition of hydrocarbons in sediments from the Fladen Ground, North Sea, an area of oil production”. In: *Journal of Environmental Monitoring* 8.2 (2006), pp. 307–316.
- [7] Ahyong, S. et al. *World Register of Marine Species*. Accessed 2023-02-06. 2023. URL: <https://www.marinespecies.org>.
- [8] Alheit, J. et al. “Climate variability drives anchovies and sardines into the North and Baltic Seas”. In: *Progress in Oceanography* 96.1 (2012), pp. 128–139.
- [9] Alvarez-Fernandez, S. et al. “Effect of zooplankton on fish larval abundance and distribution: a long-term study on North Sea herring (*Clupea harengus*)”. In: *ICES Journal of Marine Science* 72.9 (2015), pp. 2569–2577.
- [10] Alvarez-Fernandez, S., Lindeboom, H., and Meesters, E. “Temporal changes in plankton of the North Sea: community shifts and environmental drivers”. In: *Marine Ecology Progress Series* 462 (2012), pp. 21–38.
- [11] Anderson, E. and Abbe, E. C. “A quantitative comparison of specific and generic differences in the Betulaceae”. In: *Journal of the Arnold Arboretum* 15.1 (1934), pp. 43–49.
- [12] Anderson, R. M. and May, R. M. *Population biology of infectious diseases*. Springer, 1982.
- [13] Andrew, A. L., Chu, K.-W. E., and Lancaster, P. “Derivatives of eigenvalues and eigenvectors of matrix functions”. In: *SIAM journal on matrix analysis and applications* 14.4 (1993), pp. 903–926.
- [14] Arthur, W., Durlauf, S., and Lane, D. “Introduction. In The Economy as an Evolving Complex System”. In: *Santa Fe Institute Studies in the Science of Complexity, Vol. 27*. Addison-Wesley: Reading, MA. 1997, 10 pp.
- [15] Aslan, C. E. et al. “The role of honey bees as pollinators in natural areas”. In: *Natural Areas Journal* 36.4 (2016), pp. 478–488.

- [16] Attrill, M. J. and Power, M. “Climatic influence on a marine fish assemblage”. In: *Nature* 417.6886 (2002), pp. 275–278.
- [17] Auber, A. et al. “Decline of cold-water fish species in the Bay of Somme (English Channel, France) in response to ocean warming”. In: *Estuarine, Coastal and Shelf Science* 189 (2017), pp. 189–202.
- [18] Auber, A. et al. “International Bottom Trawl Survey Working Group (IBTSWG)”. In: (2023).
- [19] Barabási, A.-L. and Albert, R. “Emergence of scaling in random networks”. In: *science* 286.5439 (1999), pp. 509–512.
- [20] Baretta-Bekker, J. et al. “Description of the long-term (1991–2005) temporal and spatial distribution of phytoplankton carbon biomass in the Dutch North Sea”. In: *Journal of Sea Research* 61.1-2 (2009), pp. 50–59.
- [21] Barter, E. and Gross, T. “Manifold cities: social variables of urban areas in the UK”. In: *Proceedings of the Royal Society A* 475.2221 (2019), p. 20180615.
- [22] Bascompte, J. “Networks in ecology”. In: *Basic and applied ecology* 8.6 (2007), pp. 485–490.
- [23] Bastrokin, D. K. et al. “Settlement length and temporal settlement patterns of juvenile cod (*Gadus morhua*), haddock (*Melanogrammus aeglefinus*), and whiting (*Merlangius merlangus*) in a northern North Sea coastal nursery area”. In: *ICES Journal of Marine Science* 71.8 (2014), pp. 2101–2113.
- [24] Batten, S. D. et al. “A global plankton diversity monitoring program”. In: *Frontiers in Marine Science* 6 (2019), p. 321.
- [25] Beare, D. et al. “Long-term increases in prevalence of North Sea fishes having southern biogeographic affinities”. In: *Marine Ecology Progress Series* 284 (2004), pp. 269–278.
- [26] Beaugrand, G. “The North Sea regime shift: evidence, causes, mechanisms and consequences”. In: *Progress in Oceanography* 60.2-4 (2004), pp. 245–262.
- [27] Beezer, R. A. *A first course in linear algebra*. Independent, 2015.
- [28] Belkin, M. and Niyogi, P. “Towards a theoretical foundation for Laplacian-based manifold methods”. In: *International conference on computational learning theory*. Springer. 2005, pp. 486–500.
- [29] Bellmann, R. *Adaptive Control Processes: A Guided Tour*. Princeton University Press, 1961.
- [30] Bellwood, D. et al. “Functional versatility supports coral reef biodiversity”. In: *Proceedings of the Royal Society B: Biological Sciences* 273.1582 (2006), pp. 101–107.
- [31] Ben-Naim, E., Frauenfelder, H., and Toroczkai, Z. *Complex networks*. Vol. 650. Springer Science & Business Media, 2004.
- [32] Benway, H. M. et al. “Ocean time series observations of changing marine ecosystems: an era of integration, synthesis, and societal applications”. In: *Frontiers in Marine Science* 6 (2019), p. 393.
- [33] Berg, F. et al. “Comparative biology and population mixing among local, coastal and offshore Atlantic herring (*Clupea harengus*) in the North Sea, Skagerrak, Kattegat and western Baltic”. In: *PLoS One* 12.10 (2017), e0187374.
- [34] Berke, S. K. “Functional groups of ecosystem engineers: a proposed classification with comments on current issues”. In: *Integrative and comparative biology* 50.2 (2010), pp. 147–157.

- [35] Berkeley, S. A. et al. “Fisheries sustainability via protection of age structure and spatial distribution of fish populations”. In: *Fisheries* 29.8 (2004), pp. 23–32.
- [36] Beukhof, E. et al. *A trait collection of marine fish species from North Atlantic and North-east Pacific continental shelf seas [dataset]*. PANGAEA, 2019. DOI: [10.1594/PANGAEA.900866](https://doi.org/10.1594/PANGAEA.900866).
- [37] Beukhof, E. et al. “Marine fish traits follow fast-slow continuum across oceans”. In: *Scientific Reports* 9.1 (2019), p. 17878.
- [38] Bezanson, J. et al. “Julia: A fresh approach to numerical computing”. In: *SIAM review* 59.1 (2017), pp. 65–98.
- [39] Blonder, B. “Hypervolume concepts in niche-and trait-based ecology”. In: *Ecography* 41.9 (2018), pp. 1441–1455.
- [40] Bluemel, J. K. et al. “Decline in Atlantic wolffish *Anarhichas lupus* in the North Sea: Impacts of fishing pressure and climate change”. In: *Journal of Fish Biology* 100.1 (2022), pp. 253–267.
- [41] Blum, M. J. “Environmental Pressures on Top-Down and Bottom-Up Forces in Coastal Ecosystems”. In: *Diversity* 13.9 (2021), p. 444.
- [42] Boccaletti, S. et al. “Complex networks: Structure and dynamics”. In: *Physics reports* 424.4-5 (2006), pp. 175–308.
- [43] Bolle, L. J. et al. “Nursery grounds of dab (*Limanda limanda* L.) in the southern North Sea”. In: *Netherlands Journal of Sea Research* 32.3-4 (1994), pp. 299–307.
- [44] Bollobás, B. “Random graphs academic press”. In: *New York* (1985).
- [45] Bollobás, B. and Bollobás, B. *Random graphs*. Springer, 1998.
- [46] Bond, E. M. and Chase, J. M. “Biodiversity and ecosystem functioning at local and regional spatial scales”. In: *Ecology letters* 5.4 (2002), pp. 467–470.
- [47] Bongers, F. J. et al. “Functional diversity effects on productivity increase with age in a forest biodiversity experiment”. In: *Nature Ecology & Evolution* 5.12 (2021), pp. 1594–1603.
- [48] Borja, A. *Grand challenges in marine ecosystems ecology*. 2014.
- [49] Bornholdt, S. and Schuster, H. G. ( *Handbook of Graphs and Networks: From the Genome to the Internet*. Wiley-VCH, 2002.
- [50] Botta-Dukát, Z. “Rao’s quadratic entropy as a measure of functional diversity based on multiple traits”. In: *Journal of vegetation science* 16.5 (2005), pp. 533–540.
- [51] Boyce, D. G., Lewis, M. R., and Worm, B. “Global phytoplankton decline over the past century”. In: *Nature* 466.7306 (2010), pp. 591–596.
- [52] Bremner, J., Rogers, S., and Frid, C. “Assessing functional diversity in marine benthic ecosystems: a comparison of approaches”. In: *Marine Ecology Progress Series* 254 (2003), pp. 11–25.
- [53] Bresnan, E. et al. “The Scottish coastal observatory 1997–2013. Part 2-description of Scotland’s coastal waters”. In: *Scottish Marine and Freshwater Science* 7.26 (2016).
- [54] Brooks, D. R. et al. “Entropy and information in evolving biological systems”. In: *Biology and Philosophy* 4 (1989), pp. 407–432.
- [55] Brunet, C. et al. “Phytoplankton dynamics during the spring bloom in the south-eastern English Channel”. In: *Estuarine, Coastal and Shelf Science* 43.4 (1996), pp. 469–483.

- [56] Brylinski, J. and Aelbrecht, D. “Plankton transfers and coastal front in the Dover Strait”. In: *Oceanologica acta* 16.5-6 (1993), pp. 671–676.
- [57] Bunse, C. and Pinhassi, J. “Marine bacterioplankton seasonal succession dynamics”. In: *Trends in microbiology* 25.6 (2017), pp. 494–505.
- [58] Buyse, J. et al. “European plaice movements show evidence of high residency, site fidelity, and feeding around hard substrates within an offshore wind farm”. In: *ICES Journal of Marine Science* (2023), fsad179.
- [59] Byers, J. E. “Using ecosystem engineers to enhance multiple ecosystem processes”. In: *Functional Ecology* 38.1 (2024), pp. 22–36.
- [60] Cadée, G. C. and Hegeman, J. “Phytoplankton in the Marsdiep at the end of the 20th century; 30 years monitoring biomass, primary production, and Phaeocystis blooms”. In: *Journal of Sea Research* 48.2 (2002), pp. 97–110.
- [61] Cadotte, M. W., Carscadden, K., and Mirotchnick, N. “Beyond species: functional diversity and the maintenance of ecological processes and services”. In: *Journal of applied ecology* 48.5 (2011), pp. 1079–1087.
- [62] Cain, A. J. and Harrison, G. “An analysis of the taxonomist’s judgment of affinity”. In: *Proceedings of the Zoological Society of London*. Vol. 131. 1. Wiley Online Library. 1958, pp. 85–98.
- [63] Caley, M. J. “Community dynamics of tropical reef fishes: local patterns between latitudes”. In: *Marine Ecology Progress Series* 129 (1995), pp. 7–18.
- [64] Callaway, R. et al. “Diversity and community structure of epibenthic invertebrates and fish in the North Sea”. In: *ICES Journal of Marine Science* 59.6 (2002), pp. 1199–1214.
- [65] Campos, W. L., Kloppmann, M., and Westernhagen, H. von. “Inferences from the horizontal distribution of dab *Limanda limanda* (L.) and flounder *Platichthys flesus* (L.) larvae in the southeastern North Sea”. In: *Netherlands Journal of Sea Research* 32.3-4 (1994), pp. 277–286.
- [66] Capuzzo, E. et al. “A decline in primary production in the North Sea over 25 years, associated with reductions in zooplankton abundance and fish stock recruitment”. In: *Global change biology* 24.1 (2018), e352–e364.
- [67] Cardinale, B. J. et al. “Biodiversity loss and its impact on humanity”. In: *Nature* 486.7401 (2012), pp. 59–67.
- [68] Carrasco De La Cruz, P. M. et al. “Aggregation of monitoring datasets for functional diversity estimation”. In: *Frontiers in Ecology and Evolution* 11 (2023), p. 1285115.
- [69] Carreira-Perpinán, M. A. “A review of dimension reduction techniques”. In: *Department of Computer Science. University of Sheffield. Tech. Rep. CS-96-09* 9 (1997), pp. 1–69.
- [70] Carstensen, J., Klais, R., and Cloern, J. E. “Phytoplankton blooms in estuarine and coastal waters: Seasonal patterns and key species”. In: *Estuarine, Coastal and Shelf Science* 162 (2015), pp. 98–109.
- [71] Castro, C. de et al. “Evidence for substock dynamics within whiting (*Merlangius merlangus*) management regions”. In: *ICES Journal of Marine Science* 70.6 (2013), pp. 1118–1127.
- [72] Chase, J. M. et al. “Species richness change across spatial scales”. In: *Oikos* 128.8 (2019), pp. 1079–1091.
- [73] Chiu, C.-H. and Chao, A. “Distance-based functional diversity measures and their decomposition: a framework based on Hill numbers”. In: *PloS one* 9.7 (2014), e100014.



- [74] Chung, F. R. *Spectral graph theory*. Vol. 92. American Mathematical Soc., 1997.
- [75] Ciannelli, L. et al. “Theory, consequences and evidence of eroding population spatial structure in harvested marine fishes: a review”. In: *Marine Ecology Progress Series* 480 (2013), pp. 227–243.
- [76] Clark, A. T. et al. “Spatial convergent cross mapping to detect causal relationships from short time series”. In: *Ecology* 96.5 (2015), pp. 1174–1181.
- [77] Clark, R. A. and Frid, C. L. “Long-term changes in the North Sea ecosystem”. In: *Environmental reviews* 9.3 (2001), pp. 131–187.
- [78] Cody, M. L. and Diamond, J. M. *Ecology and evolution of communities*. Harvard University Press, 1975.
- [79] Coifman, R. R. and Lafon, S. “Diffusion maps”. In: *Applied and computational harmonic analysis* 21.1 (2006), pp. 5–30.
- [80] Coifman, R. R. et al. “Geometric diffusions as a tool for harmonic analysis and structure definition of data: Diffusion maps”. In: *Proceedings of the national academy of sciences* 102.21 (2005), pp. 7426–7431.
- [81] Colebrook, J. “Environmental influences on long-term variability in marine plankton”. In: *Long-Term Changes in Coastal Benthic Communities: Proceedings of a Symposium, held in Brussels, Belgium, December 9–12, 1985*. Springer, 1987, pp. 309–325.
- [82] Cook, R., Sinclair, A., and Stefansson, G. “Potential collapse of North Sea cod stocks”. In: *Nature* 385.6616 (1997), pp. 521–522.
- [83] Córdova-Tapia, F. and Zambrano, L. “La diversidad funcional en la ecología de comunidades”. In: *Ecosistemas* 24.3 (2015), pp. 78–87.
- [84] Corten, A. “On the causes of the recruitment failure of herring in the central and northern North Sea in the years 1972–1978”. In: *ICES Journal of Marine Science* 42.3 (1986), pp. 281–294.
- [85] Cortés, J. “Marine biodiversity baseline for Área de Conservación Guanacaste, Costa Rica: published records”. In: *ZooKeys* 652 (2017), p. 129.
- [86] D’Alelio, D. et al. “The time for sex: A biennial life cycle in a marine planktonic diatom”. In: *Limnology and Oceanography* 55.1 (2010), pp. 106–114.
- [87] Da Silva, V. E. et al. “Functional traits of fish species: Adjusting resolution to accurately express resource partitioning”. In: *Frontiers in Marine Science* 6 (2019), p. 303.
- [88] Da Silva Vieira, R. “Complex Adaptive Systems”. In: *Industry, Innovation and Infrastructure. Encyclopedia of the UN Sustainable Development Goals*. Springer, 2021.
- [89] Daan, N. “A quantitative analysis of the food intake of North Sea cod, *Gadus morhua*”. In: *Netherlands Journal of Sea Research* 6.4 (1973), pp. 479–517.
- [90] Daan, N. et al. “Ecology of North sea fish”. In: *Netherlands Journal of Sea Research* 26.2-4 (1990), pp. 343–386.
- [91] Dajka, J.-C. et al. “Modeling drivers of biodiversity change emphasizes the need for multivariate assessments and rescaled targeting for management”. In: *Conservation Science and Practice* 4.10 (2022), e12794.
- [92] Darwin, C. *Journal of researches into the geology and natural history of the countries visited during the voyage of HMS Beagle round the world, under the command of Captain Fitzroy, RN from 1832 to 1836*. London: John Murray, 1839.

- [93] Davies, C. H. et al. “A database of marine phytoplankton abundance, biomass and species composition in Australian waters”. In: *Scientific data* 3.1 (2016), pp. 1–12.
- [94] De Jong, V. and De Jong, D. “Global change’impact of inter-annual variation in water discharge as a driving factor to dredging and spoil disposal in the river Rhine system and of turbidity in the Wadden Sea”. In: *Estuarine, Coastal and Shelf Science* 55.6 (2002), pp. 969–991.
- [95] De Jonge, V., Essink, K., and Boddeke, R. “The Dutch Wadden Sea: a changed ecosystem”. In: *Hydrobiologia* 265 (1993), pp. 45–71.
- [96] De Jonge, V. and Van Beuselom, J. “Contribution of resuspended microphytobenthos to total phytoplankton in the Ems estuary and its possible role for grazers”. In: *Netherlands Journal of Sea Research* 30 (1992), pp. 91–105.
- [97] De Jorge, V. and Van Beusekom, J. “Wind-and tide-induced resuspension of sediment and microphytobenthos from tidal flats in the Ems estuary”. In: *Limnology and oceanography* 40.4 (1995), pp. 776–778.
- [98] De la Porte, J. et al. “An introduction to diffusion maps”. In: *Proceedings of the 19th symposium of the pattern recognition association of South Africa (PRASA 2008), Cape Town, South Africa*. 2008, pp. 15–25.
- [99] DeAngelis, D. L. and Yurek, S. “Equation-free modeling unravels the behavior of complex ecological systems”. In: *Proceedings of the National Academy of Sciences* 112.13 (2015), pp. 3856–3857.
- [100] DeLaplante, K. and Picasso, V. “The biodiversity-ecosystem function debate in ecology”. In: *Philosophy of ecology*. Elsevier, 2011, pp. 169–200.
- [101] Dencker, T. S. et al. “Temporal and spatial differences between taxonomic and trait biodiversity in a large marine ecosystem: Causes and consequences”. In: *PLoS One* 12.12 (2017), e0189731.
- [102] Deyle, E. R. et al. “Tracking and forecasting ecosystem interactions in real time”. In: *Proceedings of the Royal Society B: Biological Sciences* 283.1822 (2016), p. 20152258.
- [103] Di Cavalho, J. A. et al. “Temporal change in phytoplankton diversity and functional group composition”. In: *Marine Biodiversity* 53.6 (2023), p. 72.
- [104] Dickie, L. “Management of fisheries; ecological subsystems”. In: *Transactions of the American Fisheries Society* 102.2 (1973), pp. 470–480.
- [105] Diaz, S. and Cabido, M. “Vive la différence: plant functional diversity matters to ecosystem processes”. In: *Trends in ecology & evolution* 16.11 (2001), pp. 646–655.
- [106] Docker, M. F. and Potter, I. C. “Chapter 4. Life history evolution in lampreys: alternative migratory and feeding types”. In: *Lampreys: Biology, Conservation and Control*. Springer, 2019.
- [107] Dominguez-Petit, R. et al. “Changes in size at maturity of European hake Atlantic populations in relation with stock structure and environmental regimes”. In: *Journal of Marine Systems* 71.3-4 (2008), pp. 260–278.
- [108] Donoho, D. L. et al. “High-dimensional data analysis: The curses and blessings of dimensionality”. In: *AMS math challenges lecture* 1.2000 (2000), p. 32.
- [109] Dornelas, M. et al. “BioTIME: A database of biodiversity time series for the Anthropocene”. In: *Global Ecology and Biogeography* 27.7 (2018), pp. 760–786.
- [110] Dorogovtsev, S. N. and Mendes, J. F. *Evolution of networks: From biological nets to the Internet and WWW*. Oxford university press, 2003.

- [111] Drake, J. A. “The mechanics of community assembly and succession”. In: *Journal of Theoretical Biology* 147.2 (1990), pp. 213–233.
- [112] Dulvy, N. K., Sadovy, Y., and Reynolds, J. D. “Extinction vulnerability in marine populations”. In: *Fish and fisheries* 4.1 (2003), pp. 25–64.
- [113] Dulvy, N. K. et al. “Climate change and deepening of the North Sea fish assemblage: a biotic indicator of warming seas”. In: *Journal of Applied Ecology* 45.4 (2008), pp. 1029–1039.
- [114] Edwards, K. F. “Community trait structure in phytoplankton: seasonal dynamics from a method for sparse trait data”. In: *Ecology* 97.12 (2016), pp. 3441–3451.
- [115] Edwards, M., Reid, P., and Planque, B. “Long-term and regional variability of phytoplankton biomass in the Northeast Atlantic (1960–1995)”. In: *ICES Journal of Marine Science* 58.1 (2001), pp. 39–49.
- [116] Edwards, M. et al. “Ocean climate anomalies and the ecology of the North Sea”. In: *Marine Ecology Progress Series* 239 (2002), pp. 1–10.
- [117] Ehrich, S., Stelzenmueller, V., and Adlerstein, S. “Linking spatial pattern of bottom fish assemblages with water masses in the North Sea”. In: *Fisheries Oceanography* 18.1 (2009), pp. 36–50.
- [118] Ehrich, S. and Stransky, C. “Spatial and temporal changes in the southern species component of North Sea bottom fish assemblages”. In: *Senckenbergiana maritima* 31 (2001), pp. 143–150.
- [119] Ehrlich, P. and Ehrlich, A. *Extinction: the causes and consequences of the disappearance of species*. New York: Random House, 1981.
- [120] Eklöf, A. et al. “The dimensionality of ecological networks”. In: *Ecology letters* 16.5 (2013), pp. 577–583.
- [121] Elliott, S. A. et al. “Shedding light on the river and sea lamprey in western European marine waters”. In: *Endangered Species Research* 44 (2021), pp. 409–419.
- [122] Ellis, J., Rogers, S., and Freeman, S. “Demersal assemblages in the Irish Sea, St George’s channel and Bristol channel”. In: *Estuarine, Coastal and Shelf Science* 51.3 (2000), pp. 299–315.
- [123] Elton, C. S. *Animal Ecology*. Sigdwick and Jackson, 1927.
- [124] Emeis, K.-C. et al. “The North Sea—A shelf sea in the Anthropocene”. In: *Journal of Marine Systems* 141 (2015), pp. 18–33.
- [125] Engelhard, G. H. et al. “Effort reduction and the large fish indicator: spatial trends reveal positive impacts of recent European fleet reduction schemes”. In: *Environmental Conservation* 42.3 (2015), pp. 227–236.
- [126] Estes, J. A. and Palmisano, J. F. “Sea otters: their role in structuring nearshore communities”. In: *Science* 185.4156 (1974), pp. 1058–1060.
- [127] Fahimipour, A. K. and Gross, T. “Mapping the bacterial ways of life”. In: *arXiv preprint arXiv:1908.07631* (2019).
- [128] Faria, D. et al. “The breakdown of ecosystem functionality driven by deforestation in a global biodiversity hotspot”. In: *Biological Conservation* 283 (2023), p. 110126.
- [129] Felden, J. et al. *PANGAEA – Data Publisher for Earth Environmental Science. Scientific Data*. PANGAEA, 2023. DOI: [10.1038/s41597-023-02269-x](https://doi.org/10.1038/s41597-023-02269-x).

- [130] Ferrier, S. et al. “Spatial turnover in species composition of ground-dwelling arthropods, vertebrates and vascular plants in north-east New South Wales: implications for selection of forest reserves”. In: *The other* 99 (1999), pp. 68–76.
- [131] Fiedler, M. “Algebraic connectivity of graphs”. In: *Czechoslovak mathematical journal* 23.2 (1973), pp. 298–305.
- [132] Fonseca, C. R. and Ganade, G. “Species functional redundancy, random extinctions and the stability of ecosystems”. In: *Journal of Ecology* (2001), pp. 118–125.
- [133] Ford, D. A. et al. “Observing and modelling phytoplankton community structure in the North Sea”. In: *Biogeosciences* 14.6 (2017), pp. 1419–1444.
- [134] Franzè, G. et al. “North Sea Ecosystem-Report\_2023002006”. In: *Toktrapport* (2024).
- [135] Fraser, H. M. et al. “Mapping spatial variation in demersal fish species diversity and composition in the North Sea: accounting for species- and size-related catchability in survey trawls”. In: *ICES Journal of Marine Science* 65.4 (2008), pp. 531–538.
- [136] Frid, C. J. et al. “Long-term changes in the benthic communities on North Sea fishing grounds”. In: *ICES Journal of Marine Science* 57.5 (2000), pp. 1303–1309.
- [137] Froese, R. P. *Fish Base. World Wide Web electronic publication*. Accessed 2024-02-06. 2019. URL: <https://www.fishbase.org>.
- [138] Fuhrman, J. A., Cram, J. A., and Needham, D. M. “Marine microbial community dynamics and their ecological interpretation”. In: *Nature Reviews Microbiology* 13.3 (2015), pp. 133–146.
- [139] Fujii, T. “Potential influence of offshore oil and gas platforms on the feeding ecology of fish assemblages in the North Sea”. In: *Marine Ecology Progress Series* 542 (2016), pp. 167–186.
- [140] Fulcher, B. D., Little, M. A., and Jones, N. S. “Highly comparative time-series analysis: the empirical structure of time series and their methods”. In: *Journal of the Royal Society Interface* 10.83 (2013), p. 20130048.
- [141] Fung, T. et al. “Why the size structure of marine communities can require decades to recover from fishing”. In: *Marine Ecology Progress Series* 484 (2013), pp. 155–171.
- [142] Funk, V. A. “Collections-based science in the 21st century”. In: *Journal of Systematics and Evolution* 56.3 (2018), pp. 175–193.
- [143] Furlan, E. et al. “Tools and methods to support adaptive policy making in marine areas: Review and implementation of the Adaptive Marine Policy Toolbox”. In: *Ocean & Coastal Management* 151 (2018), pp. 25–35.
- [144] Gamarra, J. G. and Solé, R. V. “Bifurcations and chaos in ecology: lynx returns revisited”. In: *Ecology Letters* 3.2 (2000), pp. 114–121.
- [145] Garnier, E. et al. “Plant functional markers capture ecosystem properties during secondary succession”. In: *Ecology* 85.9 (2004), pp. 2630–2637.
- [146] Gaston, K. J. and Fuller, R. A. “Commonness, population depletion and conservation biology”. In: *Trends in ecology & evolution* 23.1 (2008), pp. 14–19.
- [147] Gault, J. A. et al. “Dissimilarity analysis based on diffusion maps”. In: *Oikos* (2023), e10249.
- [148] Gause, G. F. *The Struggle for Existence*. Willians and Wilkins, 1934.
- [149] Ghilarov, A. “The paradox of the plankton reconsidered; or, why do species coexist?” In: *Oikos* (1984), pp. 46–52.

- [150] Ghojogh, B. et al. “Laplacian-based dimensionality reduction including spectral clustering, Laplacian eigenmap, locality preserving projection, graph embedding, and diffusion map: Tutorial and survey”. In: *arXiv preprint arXiv:2106.02154* (2021).
- [151] Gibson, R. “The Atlantic salmon in fresh water: spawning, rearing and production”. In: *Reviews in fish biology and fisheries* 3 (1993), pp. 39–73.
- [152] Gislason, H. “Ecosystem effects of fishing activities in the North Sea”. In: *Marine Pollution Bulletin* 29.6-12 (1994), pp. 520–527.
- [153] Golley, F. B. *A history of the ecosystem concept in ecology: more than the sum of the parts*. Yale University Press, 1993.
- [154] González-Irusta, J. M. and Wright, P. “Spawning grounds of whiting (*Merlangius merlangus*)”. In: *Fisheries Research* 195 (2017), pp. 141–151.
- [155] Gray, J. et al. “Coastal and deep-sea benthic diversities compared”. In: *Marine ecology progress series* 159 (1997), pp. 97–103.
- [156] Grime, J. “Benefits of plant diversity to ecosystems: immediate, filter and founder effects”. In: *Journal of Ecology* 86.6 (1998), pp. 902–910.
- [157] Grinnell, J. *An account of the mammals and birds of the Lower Colorado Valley*. University California Publishing, 1914.
- [158] Gröger, J. P., Kruse, G. H., and Rohlf, N. “Slave to the rhythm: how large-scale climate cycles trigger herring (*Clupea harengus*) regeneration in the North Sea”. In: *ICES Journal of Marine Science* 67.3 (2010), pp. 454–465.
- [159] Gross, N. et al. “Functional trait diversity maximizes ecosystem multifunctionality”. In: *Nature ecology & evolution* 1.5 (2017), p. 0132.
- [160] Hadj-Hammou, J., Mouillot, D., and Graham, N. A. “Response and effect traits of coral reef fish”. In: *Frontiers in Marine Science* 8 (2021), p. 640619.
- [161] Haila, Y. and Kouki, J. “The phenomenon of biodiversity in conservation biology”. In: *Annales Zoologici Fennici*. JSTOR. 1994, pp. 5–18.
- [162] Hanewinkel, M. et al. “Seventy-seven years of natural disturbances in a mountain forest area—the influence of storm, snow, and insect damage analysed with a long-term time series”. In: *Canadian Journal of Forest Research* 38.8 (2008), pp. 2249–2261.
- [163] Hanski, I. “Metapopulation dynamics”. In: *Nature* 396.6706 (1998), pp. 41–49.
- [164] Harris, M. P. et al. “A major increase in snake pipefish (*Entelurus aequoreus*) in northern European seas since 2003: potential implications for seabird breeding success”. In: *Marine Biology* 151 (2007), pp. 973–983.
- [165] Harvey, C. J. et al. “The importance of long-term ecological time series for integrated ecosystem assessment and ecosystem-based management”. In: *Progress in Oceanography* 188 (2020), p. 102418.
- [166] Hatcher, M. J. and Dunn, A. M. *Parasites in ecological communities: from interactions to ecosystems*. Cambridge University Press, 2011.
- [167] Hawksworth, D. L. *Biodiversity: measurement and estimation*. Vol. 345. Springer Science & Business Media, 1995.
- [168] Hays, G. C., Richardson, A. J., and Robinson, C. “Climate change and marine plankton”. In: *Trends in ecology & evolution* 20.6 (2005), pp. 337–344.
- [169] Heessen, H. J. “Time-series data for a selection of forty fish species caught during the International Bottom Trawl Survey”. In: *ICES Journal of Marine Science* 53.6 (1996), pp. 1079–1084.

- [170] Heessen, H. J. and Daan, N. “Long-term trends in ten non-target North Sea fish species”. In: *ICES Journal of Marine Science* 53.6 (1996), pp. 1063–1078.
- [171] Heessen, H. J., Daan, N., and Ellis, J. R. *Fish atlas of the Celtic Sea, North Sea, and Baltic Sea: Based on international research-vessel surveys*. BRILL, 2023.
- [172] Heincke, F. “Naturgeschichte des Herings. Abhandl. d. deut”. In: *Seefish. Vereins. Bd. II* (1898).
- [173] Heip, C. et al. “Trends in biomass, density and diversity of North Sea macrofauna”. In: *ICES Journal of Marine Science* 49.1 (1992), pp. 13–22.
- [174] Henson, S. A., Beaulieu, C., and Lampitt, R. “Observing climate change trends in ocean biogeochemistry: when and where”. In: *Global change biology* 22.4 (2016), pp. 1561–1571.
- [175] Henson, S. A. et al. “Detection of anthropogenic climate change in satellite records of ocean chlorophyll and productivity”. In: *Biogeosciences* 7.2 (2010), pp. 621–640.
- [176] Heydari, B. and Wade, J. *Complexity: Definition and Reduction Techniques, Some Simple Thoughts on Complex Systems*. 2014.
- [177] Hill, M. O. “Diversity and evenness: a unifying notation and its consequences”. In: *Ecology* 54.2 (1973), pp. 427–432.
- [178] Hillebrand, H. and Matthiessen, B. “Biodiversity in a complex world: consolidation and progress in functional biodiversity research”. In: *Ecology letters* 12.12 (2009), pp. 1405–1419.
- [179] Hillebrand, H. et al. “Biodiversity change is uncoupled from species richness trends: Consequences for conservation and monitoring”. In: *Journal of applied ecology* 55.1 (2018), pp. 169–184.
- [180] Hinder, S. L. et al. “Changes in marine dinoflagellate and diatom abundance under climate change”. In: *Nature Climate Change* 2.4 (2012), pp. 271–275.
- [181] Hislop, J. R. “Changes in North Sea gadoid stocks”. In: *ICES Journal of Marine Science* 53.6 (1996), pp. 1146–1156.
- [182] Hislop, J. and MacKenzie, K. “Population studies of the whiting *Merlangius merlangus* (L.) of the northern North Sea”. In: *ICES Journal of Marine Science* 37.1 (1976), pp. 98–110.
- [183] Hislop, J. et al. “The diet and food consumption of whiting (*Merlangius merlangus*) in the North Sea”. In: *ICES Journal of Marine Science* 48.2 (1991), pp. 139–156.
- [184] Hoffmann, M. et al. “The impact of conservation on the status of the world’s vertebrates”. In: *science* 330.6010 (2010), pp. 1503–1509.
- [185] Holland, J. H. *Hidden order: How adaptation builds complexity*. Reading (MA): Addison-Wesley, 1995.
- [186] Hoogenboom, M. O. and Connolly, S. R. “Defining fundamental niche dimensions of corals: synergistic effects of colony size, light, and flow”. In: *Ecology* 90.3 (2009), pp. 767–780.
- [187] Hopfield, J. “Physics, computation, and why biology looks so different”. In: *Journal of Theoretical Biology* 171.1 (1994), pp. 53–60.
- [188] Hsieh, C.-h., Anderson, C., and Sugihara, G. “Extending nonlinear analysis to short ecological time series”. In: *The American Naturalist* 171.1 (2008), pp. 71–80.
- [189] Hsieh, C.-H. and Chiu, T.-S. “Summer spatial distribution of copepods and fish larvae in relation to hydrography in the northern Taiwan Strait”. In: *Zoological Studies* 41.1 (2004), pp. 85–98.

- [190] Hudson, L. N. et al. “The PREDICTS database: a global database of how local terrestrial biodiversity responds to human impacts”. In: *Ecology and evolution* 4.24 (2014), pp. 4701–4735.
- [191] Hutchinson, G. E. “Concluding remarks”. In: *Cold Spring Harbor symposia on quantitative biology*. Vol. 22. Cold Spring Harbor Laboratory Press. 1957, pp. 415–427.
- [192] Hutchinson, G. E. *Introduction to population ecology*. Yale University Press, 1978.
- [193] Huthnance, J. “Physical oceanography of the North Sea”. In: *Ocean and shoreline management* 16.3-4 (1991), pp. 199–231.
- [194] Hutton, J., Adams, W. M., and Murombedzi, J. C. “Back to the barriers? Changing narratives in biodiversity conservation”. In: *Forum for development studies*. Vol. 32. 2. Taylor & Francis. 2005, pp. 341–370.
- [195] ICES. *Database on Trawl Surveys (DATRAS)*. Data Accessed online on November 2023. ICES, Copenhagen, Denmark. 2024. URL: <https://datras.ices.dk>.
- [196] Iles, T. and Sinclair, M. “Atlantic herring: stock discreteness and abundance”. In: *Science* 215.4533 (1982), pp. 627–633.
- [197] IOCCG. “Ocean-Colour Data Merging.” In: *Report of the International Ocean-Color Coordinating Group, No. 6*. IOCCG, Dartmouth, Canada., 2007.
- [198] Isbell, F. et al. “High plant diversity is needed to maintain ecosystem services”. In: *Nature* 477.7363 (2011), pp. 199–202.
- [199] Jaccard, P. “The distribution of the flora in the alpine zone. 1”. In: *New phytologist* 11.2 (1912), pp. 37–50.
- [200] Jacobs, P. et al. “Planktonic primary production in the western Dutch Wadden Sea”. In: *Marine Ecology Progress Series* 639 (2020), pp. 53–71.
- [201] Jennings, S. et al. “Spatial and temporal patterns in North Sea fishing effort”. In: *Effects of Fishing on Non-Target Species and Habitats: Biological Conservation and Socio-Economic Issues*. Blackwell Science, 2000, pp. 3–14.
- [202] Jiming, Y. “The dominant fish fauna in the North Sea and its determination”. In: *Journal of Fish Biology* 20.6 (1982), pp. 635–643.
- [203] Johnson, R. H. *Determine Evolution in the Color-Pattern of the Lady-Beetles*. Carnegie Institution of Washington, 1910.
- [204] Jolliffe, I. *Principal component analysis*. Springer-Verlag New York, 1986.
- [205] Jones, G. P. “Competitive interactions among adults and juveniles in a coral reef fish”. In: *Ecology* 68.5 (1987), pp. 1534–1547.
- [206] Jost, L. “Entropy and diversity”. In: *Oikos* 113.2 (2006), pp. 363–375.
- [207] Jung, A. et al. “Spatial and temporal trends in order richness of marine phytoplankton as a tracer for the exchange zone between coastal and open waters”. In: *Journal of the Marine Biological Association of the United Kingdom* 97.3 (2017), pp. 477–489.
- [208] Jürgens, N. et al. “The BIOTA Biodiversity Observatories in Africa—a standardized framework for large-scale environmental monitoring”. In: *Environmental monitoring and assessment* 184 (2012), pp. 655–678.
- [209] Kanstinger, P. and Peck, M. A. “Co-occurrence of European sardine (*Sardina pilchardus*), anchovy (*Engraulis encrasicolus*) and sprat (*Sprattus sprattus*) larvae in southern North Sea habitats: abundance, distribution and biochemical-based condition”. In: *Scientia Marina* 73.S1 (2009), pp. 141–152.

- [210] Karl, D. M. “Oceanic ecosystem time-series programs: ten lessons learned”. In: *Oceanography* 23.3 (2010), pp. 104–125.
- [211] Karlsson-Vinkhuyzen, S. et al. “Mainstreaming biodiversity in economic sectors: An analytical framework”. In: *Biological Conservation* 210 (2017), pp. 145–156.
- [212] Kauffman, K. P. “Emotional sentience and the nature of phenomenal experience”. In: *Progress in biophysics and molecular biology* 119.3 (2015), pp. 545–562.
- [213] Keddy, P. A. “Assembly and response rules: two goals for predictive community ecology”. In: *Journal of vegetation science* 3.2 (1992), pp. 157–164.
- [214] Kikkawa, J. and Anderson, D. J. *Community ecology: Pattern and process*. Blackwell Scientific, Melbourne, Australia, 1986.
- [215] Kiørboe, T., Visser, A., and Andersen, K. H. “A trait-based approach to ocean ecology”. In: *ICES Journal of Marine Science* 75.6 (2018), pp. 1849–1863.
- [216] Kirby, R. R., Johns, D. G., and Lindley, J. A. “Fathers in hot water: rising sea temperatures and a Northeastern Atlantic pipefish baby boom”. In: *Biology letters* 2.4 (2006), pp. 597–600.
- [217] Komárek, J. and Anagnostidis, K. *Süßwasserflora von Mitteleuropa. Bd. 19 (1). Cyanoprokaryota 1. Teil: Chroococcales*. 1998.
- [218] Körner, C. “Scalling from Species to Vegetation: The Usefulness of Functional Groups”. In: *Biodiversity and Ecosystem Function*. Springer, 19993.
- [219] Kossinna, E. “Die tiefen des weltmeeres”. In: (*No Title*) (1921).
- [220] Krammer, K. and Lange-Bertalot, H. “Die Süßwasserflora von Mitteleuropa 2: Bacillariophyceae. 1 Teil: Naviculaceae”. In: *Gustav-Fischer Verlag, Stuttgart* (1991), pp. 1–876.
- [221] Krammer, K. and Lange-Bertalot, H. “Die Süßwasserflora von Mitteleuropa 2: Bacillariophyceae. 2 Teil: Bacillariaceae, Epithemiaceae, Surriellaceae”. In: *Gustav-Fischer Verlag, Stuttgart* (1988), pp. 1–569.
- [222] Krammer, K. and Lange-Bertalot, H. “Die Süßwasserflora von Mitteleuropa 2: Bacillariophyceae. 3 Teil: Centrales, Fragilariaceae, Eunotiaceae”. In: *Gustav Fischer-Verlag, Stuttgart* (1991).
- [223] Krebs, C. J. “Population cycles revisited”. In: *Journal of mammalogy* 77.1 (1996), pp. 8–24.
- [224] Laliberté, E. and Legendre, P. “A distance-based framework for measuring functional diversity from multiple traits”. In: *Ecology* 91.1 (2010), pp. 299–305.
- [225] Larsen, B. B. et al. “Inordinate fondness multiplied and redistributed: the number of species on earth and the new pie of life”. In: *The Quarterly Review of Biology* 92.3 (2017), pp. 229–265.
- [226] Laughlin, D. C. “Nitrification is linked to dominant leaf traits rather than functional diversity”. In: *Journal of Ecology* 99.5 (2011), pp. 1091–1099.
- [227] Lawton, J. H. “Are there general laws in ecology?” In: *Oikos* (1999), pp. 177–192.
- [228] Lee, A. “North Sea: physical oceanography”. In: *Elsevier Oceanography Series*. Vol. 24. Elsevier, 1980, pp. 467–493.
- [229] Legendre, P., Galzin, R., and Harmelin-Vivien, M. L. “Relating behavior to habitat: solutions to thefourth-corner problem”. In: *Ecology* 78.2 (1997), pp. 547–562.



- [230] Legendre, P. and Legendre, L. “Complex ecological data sets”. In: *Developments in environmental modelling*. Vol. 24. Elsevier, 2012, pp. 1–57.
- [231] Legras, G. et al. “Assessing functional diversity: the influence of the number of the functional traits”. In: *Theoretical Ecology* 13.1 (2020), pp. 117–126.
- [232] Leibold, M. A. et al. “Species turnover and the regulation of trophic structure”. In: *Annual review of ecology and systematics* 28.1 (1997), pp. 467–494.
- [233] Leps, J. et al. “Quantifying and interpreting functional diversity of natural communities: practical considerations matter”. In: *Preslia* 78.4 (2006), pp. 481–501.
- [234] Letten, A. D., Ke, P.-J., and Fukami, T. “Linking modern coexistence theory and contemporary niche theory”. In: *Ecological Monographs* 87.2 (2017), pp. 161–177.
- [235] Levin, S. A. “Ecosystems and the biosphere as complex adaptive systems”. In: *Ecosystems* (1998).
- [236] Levin, S. A. *Fragile dominion: complexity and the commons*. Helix Books, 1999.
- [237] Levin, S. A. “The problem of pattern and scale in ecology: the Robert H. MacArthur award lecture”. In: *Ecology* 73.6 (1992), pp. 1943–1967.
- [238] Lieberman, E., Hauert, C., and Nowak, M. A. “Evolutionary dynamics on graphs”. In: *Nature* 433.7023 (2005), pp. 312–316.
- [239] Lindeman, R. L. “The trophic-dynamic aspect of ecology”. In: *Ecology* 23.4 (1942), pp. 399–417.
- [240] Lindstrom, E. et al. *A Framework for Ocean Observing*. IOC Information Document. UNESCO, 2012.
- [241] Linquist, S. et al. “Yes! There are resilient generalizations (or “laws”) in ecology”. In: *The Quarterly Review of Biology* 91.2 (2016), pp. 119–131.
- [242] Litchman, E. et al. “The role of functional traits and trade-offs in structuring phytoplankton communities: scaling from cellular to ecosystem level”. In: *Ecology letters* 10.12 (2007), pp. 1170–1181.
- [243] Liu, H.-L. et al. “Effect of climate change on the vulnerability of a socio-ecological system in an arid area”. In: *Global and Planetary Change* 137 (2016), pp. 1–9.
- [244] Llope, M. et al. “Effects of environmental conditions on the seasonal distribution of phytoplankton biomass in the North Sea”. In: *Limnology and Oceanography* 54.2 (2009), pp. 512–524.
- [245] Loebel, M., Dolch, T., and Beusekom, J. E. van. “Annual dynamics of pelagic primary production and respiration in a shallow coastal basin”. In: *Journal of Sea Research* 58.4 (2007), pp. 269–282.
- [246] Loebel, M. and Van Beusekom, J. E. “Seasonality of microzooplankton grazing in the northern Wadden Sea”. In: *Journal of Sea Research* 59.4 (2008), pp. 203–216.
- [247] Loh, J. et al. “The Living Planet Index: using species population time series to track trends in biodiversity”. In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 360.1454 (2005), pp. 289–295.
- [248] Loots, C. et al. “Understanding what controls the spawning distribution of North Sea whiting (*Merlangius merlangus*) using a multi-model approach”. In: *Fisheries Oceanography* 20.1 (2011), pp. 18–31.
- [249] Loreau, M. “Biodiversity and ecosystem functioning: a mechanistic model”. In: *Proceedings of the National Academy of Sciences* 95.10 (1998), pp. 5632–5636.

- [250] Loreau, M. et al. “Biodiversity and ecosystem functioning: current knowledge and future challenges”. In: *science* 294.5543 (2001), pp. 804–808.
- [251] Lorenzano, P. “On biological laws and the laws of biological sciences”. In: *Revista Patagónica de Filosofía* 2 (2001), pp. 27–42.
- [252] Lovejoy, T. E. “The Forestry Projection and The Environment”. In: *The global 2000 report to the President: A report*. US Government Printing Office, 1980.
- [253] Ma, Y. and Fu, Y. *Manifold learning theory and applications*. Vol. 434. CRC press Boca Raton, 2012.
- [254] MacArthur, R. and Levins, R. “The limiting similarity, convergence, and divergence of coexisting species”. In: *The american naturalist* 101.921 (1967), pp. 377–385.
- [255] Maclaurin, J. and Sterelny, K. *What is biodiversity?* University of Chicago Press, 2008.
- [256] Magnussen, E. “Demersal fish assemblages of Faroe Bank: species composition, distribution, biomass spectrum and diversity”. In: *Marine Ecology Progress Series* 238 (2002), pp. 211–225.
- [257] Magurran, A. E. *Measuring biological diversity*. Blackwell Science, USA, 2005.
- [258] Magurran, A. E. and Henderson, P. A. “Temporal turnover and the maintenance of diversity in ecological assemblages”. In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 365.1558 (2010), pp. 3611–3620.
- [259] Mamie, J. et al. “Aspects of the distribution and growth of bluemouth (*Helicolenus dactylopterus*, Delaroche 1809) since its invasion of the northern North Sea in 1991”. In: *Fisheries Oceanography* 16.1 (2007), pp. 85–94.
- [260] Margalef, R. “Diversity, stability and maturity in natural ecosystems”. In: *Unifying Concepts in Ecology: Report of the plenary sessions of the First international congress of ecology, The Hague, the Netherlands, September 8–14, 1974*. Springer. 1975, pp. 151–160.
- [261] Mariani, P. et al. “Control of plankton seasonal succession by adaptive grazing”. In: *Limnology and Oceanography* 58.1 (2013), pp. 173–184.
- [262] Martin, L. J., Blossey, B., and Ellis, E. “Mapping where ecologists work: biases in the global distribution of terrestrial ecological observations”. In: *Frontiers in Ecology and the Environment* 10.4 (2012), pp. 195–201.
- [263] Martinez-Gordón, R. et al. “A review of the role of spatial resolution in energy systems modelling: Lessons learned and applicability to the North Sea region”. In: *Renewable and Sustainable Energy Reviews* 141 (2021), p. 110857.
- [264] Mason, N. W. et al. “An index of functional diversity”. In: *Journal of Vegetation Science* 14.4 (2003), pp. 571–578.
- [265] Mason, N. W. et al. “Functional richness, functional evenness and functional divergence: the primary components of functional diversity”. In: *Oikos* 111.1 (2005), pp. 112–118.
- [266] Maßing, J. C. “Dynamics of Complex ecological communities”. PhD thesis. Carl von Ossietzky Universität Oldenburg, 2024.
- [267] Massing, J. C. et al. “Quantification of metabolic niche occupancy dynamics in a Baltic Sea bacterial community”. In: *Msystems* 8.3 (2023), e00028–23.
- [268] Mateus, C. S. et al. “European lampreys: new insights on postglacial colonization, gene flow and speciation”. In: *PLoS One* 11.2 (2016), e0148107.

- [269] Maureaud, A. et al. “Biodiversity–ecosystem functioning relationships in fish communities: biomass is related to evenness and the environment, not to species richness”. In: *Proceedings of the Royal Society B* 286.1906 (2019), p. 20191189.
- [270] May, R. “Larval mortality in marine fishes and the critical period concept”. In: *The Early Life History of Fish: The Proceedings of an International Symposium Held at the Dunstaffnage Marine Research Laboratory of the Scottish Marine Biological Association at Oban, Scotland, from May 17–23, 1973*. Springer, 1974, pp. 3–19.
- [271] May, R. M. “Network structure and the biology of populations”. In: *Trends in Ecology & Evolution* 21.7 (2006), pp. 394–399.
- [272] McCann, K. “Protecting biostructure”. In: *Nature* 446.7131 (2007), pp. 29–29.
- [273] McCann, K. S., Rasmussen, J., and Umbanhowar, J. “The dynamics of spatially coupled food webs”. In: *Ecology letters* 8.5 (2005), pp. 513–523.
- [274] McGeoch, M. A. et al. “Monitoring biological invasion across the broader Antarctic: a baseline and indicator framework”. In: *Global Environmental Change* 32 (2015), pp. 108–125.
- [275] McGill, B. J. et al. “Rebuilding community ecology from functional traits”. In: *Trends in ecology & evolution* 21.4 (2006), pp. 178–185.
- [276] McNaughton, S. J. “Diversity and stability of ecological communities: a comment on the role of empiricism in ecology”. In: *The American Naturalist* 111.979 (1977), pp. 515–525.
- [277] Melián, C. J. et al. “Deciphering the interdependence between ecological and evolutionary networks”. In: *Trends in ecology & evolution* 33.7 (2018), pp. 504–512.
- [278] Milligan, A. and Bear, B. “SmartBuoy Marine Observational Network - Phytoplankton Data 2001-2017.” In: *Cefas, UK. V2* (2019).
- [279] Miloslavich, P. et al. “Essential ocean variables for global sustained observations of biodiversity and ecosystem changes”. In: *Global change biology* 24.6 (2018), pp. 2416–2433.
- [280] Möllmann, C. and Diekmann, R. “Marine ecosystem regime shifts induced by climate and overfishing: a review for the Northern Hemisphere”. In: *Advances in ecological research* 47 (2012), pp. 303–347.
- [281] Mouillot, D., Dumay, O., and Tomasini, J. A. “Limiting similarity, niche filtering and functional diversity in coastal lagoon fish communities”. In: *Estuarine, Coastal and Shelf Science* 71.3-4 (2007), pp. 443–456.
- [282] Mouillot, D. et al. “Functional regularity: a neglected aspect of functional diversity”. In: *Oecologia* 142 (2005), pp. 353–359.
- [283] Murphy, D. W. “Theorizing climate change,(im) mobility and socio-ecological systems resilience in low-elevation coastal zones”. In: *Climate and Development* 7.4 (2015), pp. 380–397.
- [284] Murray, J. et al. *Report on the scientific results of the voyage of HMS Challenger during the years 1873-76 under the command of Captain George S. Nares and the late Captain Frank Tourle Thomson*. Vol. 2. HM Stationery Office, 1885.
- [285] Mutshinda, C. M. et al. “Phytoplankton traits from long-term oceanographic time-series”. In: *Marine Ecology Progress Series* 576 (2017), pp. 11–25.
- [286] Naeem, S. “Ecosystem consequences of biodiversity loss: the evolution of a paradigm”. In: *Ecology* 83.6 (2002), pp. 1537–1552.

- [287] Naeem, S. et al. “Declining biodiversity can alter the performance of ecosystems”. In: *Nature* 368.6473 (1994), pp. 734–737.
- [288] Newman, M. *Networks*. Oxford university press, 2010.
- [289] Newman, M. E. “The structure and function of complex networks”. In: *SIAM review* 45.2 (2003), pp. 167–256.
- [290] Niu, L. et al. “Statistical analysis of phytoplankton biomass in coastal waters: Case study of the Wadden Sea near Lauwersoog (The Netherlands) from 2000 to 2009”. In: *Ecological Informatics* 30 (2015), pp. 12–19.
- [291] NLWKN. “Gewässerüberwachungssystem Niedersachsen, Gütemessnetz Übergangs- und Küstengewässer”. In: *Küstengewässer und Ästuare* 6 (2013).
- [292] Norse, E. A. and McManus, R. E. “Ecology and Living Resources - Biological Resources”. In: *Environmental Quality, 1980: The Eleventh Annual Report of the Council on Environmental Quality*. Executive Office of the President, Council on Environmental Quality, 1980.
- [293] Noy-Meir, I., Walker, D., and Williams, W. “Data transformations in ecological ordination: II. On the meaning of data standardization”. In: *The Journal of Ecology* (1975), pp. 779–800.
- [294] Ntiba, M. and Harding, D. “The food and the feeding habits of the long rough dab, *Hippoglossoides platessoides* (Fabricius 1780) in the North Sea”. In: *Netherlands journal of sea research* 31.2 (1993), pp. 189–199.
- [295] Nunnally, C. C., Benfield, M. C., and McClain, C. R. “Trait-based diversity of deep-sea benthic megafauna communities near the Deepwater Horizon oil spill site”. In: *Marine Ecology* 41.5 (2020), e12611.
- [296] Nye, J. A. et al. “Changing spatial distribution of fish stocks in relation to climate and population size on the Northeast United States continental shelf”. In: *Marine Ecology Progress Series* 393 (2009), pp. 111–129.
- [297] O’Brien, T. D. *What are Marine Ecological Time Series telling us about the ocean? A status report*. UNESCO-IOC, 2017.
- [298] Odum, H. T. “Primary production in flowing waters 1”. In: *Limnology and oceanography* 1.2 (1956), pp. 102–117.
- [299] Oksanen, M. “The moral value of biodiversity”. In: *Ambio* (1997), pp. 541–545.
- [300] Olenina, I. “Biovolumes and size-classes of phytoplankton in the Baltic Sea”. In: *Balt. Sea Environ. Proc.* 106 (2006), p. 144.
- [301] OSPAR. *Quality Status Report 2010*. OSPAR Commission, London, 2010.
- [302] Ostle, C. et al. “The Plankton Lifeform Extraction Tool: a digital tool to increase the discoverability and usability of plankton time-series data”. In: *Earth System Science Data Discussions* 2021 (2021), pp. 1–73.
- [303] Ottersen, G. et al. “Ecological effects of the North Atlantic oscillation”. In: *Oecologia* 128 (2001), pp. 1–14.
- [304] Paerl, H. W. et al. “Phytoplankton photopigments as indicators of estuarine and coastal eutrophication”. In: *BioScience* 53.10 (2003), pp. 953–964.
- [305] Paine, R. T. “Food web complexity and species diversity”. In: *The American Naturalist* 100.910 (1966), pp. 65–75.

- [306] Papa, G. et al. “The honey bee *Apis mellifera*: An insect at the interface between human and ecosystem health”. In: *Biology* 11.2 (2022), p. 233.
- [307] Pasquarella, V. J. et al. “From imagery to ecology: leveraging time series of all available Landsat observations to map and monitor ecosystem state and dynamics”. In: *Remote Sensing in Ecology and Conservation* 2.3 (2016), pp. 152–170.
- [308] Pastor-Satorras, R. and Vespignani, A. “Epidemic spreading in scale-free networks”. In: *Physical review letters* 86.14 (2001), p. 3200.
- [309] Pauly, D. “Why do fish reach first maturity when they do”. In: *J. Fish Biol* 101 (2022), pp. 333–341.
- [310] Pearson, K. “VII. Mathematical contributions to the theory of evolution.—III. Regression, heredity, and panmixia”. In: *Philosophical Transactions of the Royal Society of London. Series A, containing papers of a mathematical or physical character* 187 (1896), pp. 253–318.
- [311] Pearson, K. and Lee, A. “On the laws of inheritance in man: I. Inheritance of physical characters”. In: *Biometrika* 2.4 (1903), pp. 357–462.
- [312] Pecuchet, L., Törnroos, A., and Lindegren, M. “Patterns and drivers of fish community assembly in a large marine ecosystem”. In: *Marine Ecology Progress Series* 546 (2016), pp. 239–248.
- [313] Pecuchet, L. et al. “From traits to life-history strategies: Deconstructing fish community composition across European seas”. In: *Global Ecology and Biogeography* 26.7 (2017), pp. 812–822.
- [314] Peperzak, L. “An objective procedure to remove observer-bias from phytoplankton time-series”. In: *Journal of Sea Research* 63.2 (2010), pp. 152–156.
- [315] Perry, A. L. et al. “Climate change and distribution shifts in marine fishes”. In: *science* 308.5730 (2005), pp. 1912–1915.
- [316] Petchey, O. L. and Gaston, K. J. “Functional diversity (FD), species richness and community composition”. In: *Ecology letters* 5.3 (2002), pp. 402–411.
- [317] Petchey, O. L. and Gaston, K. J. “Functional diversity: back to basics and looking forward”. In: *Ecology letters* 9.6 (2006), pp. 741–758.
- [318] Petitgas, P. “Biomass-dependent dynamics of fish spatial distributions characterized by geostatistical aggregation curves”. In: *ICES Journal of Marine Science* 55.3 (1998), pp. 443–453.
- [319] Petitgas, P. et al. “Impacts of climate change on the complex life cycles of fish”. In: *Fisheries Oceanography* 22.2 (2013), pp. 121–139.
- [320] Pielou, E. C. “The measurement of diversity in different types of biological collections”. In: *Journal of theoretical biology* 13 (1966), pp. 131–144.
- [321] Pierce, D. J. and Mahmoudi, B. “Nearshore fish assemblages along the central west coast of Florida”. In: *Bulletin of Marine Science* 68.2 (2001), pp. 243–270.
- [322] Pimm, S. L. “The complexity and stability of ecosystems”. In: *Nature* 307.5949 (1984), pp. 321–326.
- [323] Pimm, S. L. et al. “The future of biodiversity”. In: *Science* 269.5222 (1995), pp. 347–350.
- [324] Pingree, R. “North Atlantic and North Sea climate change: curl up, shut down, NAO and ocean colour”. In: *Journal of the Marine Biological Association of the United Kingdom* 85.6 (2005), pp. 1301–1315.

- [325] Pires, M. A., Crokidakis, N., and Queirós, S. M. D. “Diffusion plays an unusual role in ecological quasi-neutral competition in metapopulations”. In: *Nonlinear Dynamics* 103 (2021), pp. 1219–1228.
- [326] Podani, J. and Schmera, D. “On dendrogram-based measures of functional diversity”. In: *Oikos* 115.1 (2006), pp. 179–185.
- [327] Pomati, F. et al. “Challenges and prospects for interpreting long-term phytoplankton diversity changes in Lake Zurich (Switzerland)”. In: *Freshwater Biology* 60.5 (2015), pp. 1052–1059.
- [328] Poos, M. S., Walker, S. C., and Jackson, D. A. “Functional-diversity indices can be driven by methodological choices and species richness”. In: *Ecology* 90.2 (2009), pp. 341–347.
- [329] Postma, H. “Introduction to the symposium on organic matter in the Wadden Sea”. In: *Neth Inst S* 10 (1984), pp. 15–22.
- [330] Potamias, M. et al. “Fast shortest path distance estimation in large networks”. In: *Proceedings of the 18th ACM conference on Information and knowledge management*. 2009, pp. 867–876.
- [331] Potts, S. et al. “Enhancing pollinator biodiversity in intensive grasslands”. In: *Journal of Applied Ecology* 46.2 (2009), pp. 369–379.
- [332] Power, M. E. et al. “Challenges in the quest for keystones: identifying keystone species is difficult—but essential to understanding how loss of species will affect ecosystems”. In: *BioScience* 46.8 (1996), pp. 609–620.
- [333] Price, P. W., Slobodchikoff, C. N., and Gaud, W. S. “A new ecology: Novel approaches to interactive systems”. In: (1984).
- [334] Priede, I. G. et al. “The absence of sharks from abyssal regions of the world’s oceans”. In: *Proceedings of the Royal Society B: Biological Sciences* 273.1592 (2006), pp. 1435–1441.
- [335] Proença, V. et al. “Global biodiversity monitoring: from data sources to essential biodiversity variables”. In: *Biological Conservation* 213 (2017), pp. 256–263.
- [336] Proulx, S. R., Promislow, D. E., and Phillips, P. C. “Network thinking in ecology and evolution”. In: *Trends in ecology & evolution* 20.6 (2005), pp. 345–353.
- [337] Pulliam, H. R. “On the relationship between niche and distribution”. In: *Ecology letters* 3.4 (2000), pp. 349–361.
- [338] Quammen, D. “The song of the dodo: island biogeography in an age of extinctions//Review”. In: *Queen’s Quarterly* 103.3 (1996), p. 594.
- [339] Raab, K. et al. “Influence of temperature and food availability on juvenile European anchovy *Engraulis encrasicolus* at its northern boundary”. In: *Marine Ecology Progress Series* 488 (2013), pp. 233–245.
- [340] Radlinski, M. K. et al. “Spatial and temporal distribution of Atlantic mackerel (*Scomber scombrus*) along the northeast coast of the United States, 1985–1999”. In: *ICES Journal of Marine Science* 70.6 (2013), pp. 1151–1161.
- [341] Ramond, P. et al. “Phytoplankton taxonomic and functional diversity patterns across a coastal tidal front”. In: *Scientific Reports* 11.1 (2021), p. 2682.
- [342] Rao, C. R. “Diversity and dissimilarity coefficients: a unified approach”. In: *Theoretical population biology* 21.1 (1982), pp. 24–43.
- [343] Rastetter, E. B. et al. “A revised assessment of species redundancy and ecosystem reliability”. In: *Conservation Biology* 13.2 (1999), pp. 440–443.

- [344] Reid, P. et al. “Impacts of fisheries on plankton community structure”. In: *ICES Journal of Marine Science* 57.3 (2000), pp. 495–502.
- [345] Reid, P. et al. “Phytoplankton of the North Sea and its dynamics: a review”. In: *Netherlands Journal of Sea Research* 26.2-4 (1990), pp. 295–331.
- [346] Reiss, H. et al. “Spatial patterns of infauna, epifauna, and demersal fish communities in the North Sea”. In: *ICES Journal of Marine Science* 67.2 (2010), pp. 278–293.
- [347] Reynolds, C. S. *The ecology of phytoplankton*. Cambridge University Press, 2006.
- [348] Richardson, A. et al. “Using continuous plankton recorder data”. In: *Progress in Oceanography* 68.1 (2006), pp. 27–74.
- [349] Ricotta, C. “A note on functional diversity measures”. In: *Basic and applied Ecology* 6.5 (2005), pp. 479–486.
- [350] Ricotta, C. and Moretti, M. “CWM and Rao’s quadratic diversity: a unified framework for functional ecology”. In: *Oecologia* 167.1 (2011), pp. 181–188.
- [351] Rijnsdorp, A. D. et al. “Resolving the effect of climate change on fish populations”. In: *ICES journal of marine science* 66.7 (2009), pp. 1570–1583.
- [352] Rivett, D. W. et al. “Resource-dependent attenuation of species interactions during bacterial succession”. In: *The ISME journal* 10.9 (2016), pp. 2259–2268.
- [353] Rooney, N., McCann, K. S., and Moore, J. C. “A landscape theory for food web architecture”. In: *Ecology letters* 11.8 (2008), pp. 867–881.
- [354] Roscher, C. et al. “The role of biodiversity for element cycling and trophic interactions: an experimental approach in a grassland community”. In: *Basic and Applied Ecology* 5.2 (2004), pp. 107–121.
- [355] Rosenberg, A. “How is biological explanation possible?” In: *The British Journal for the Philosophy of Science* 52.4 (2001), pp. 735–760.
- [356] Rosenfeld, J. S. “Functional redundancy in ecology and conservation”. In: *Oikos* 98.1 (2002), pp. 156–162.
- [357] Rowell, C. A. “The effects of fishing on the timing of maturity in North Sea cod (*Gadus morhua* L.)” In: *The Exploitation of Evolving Resources: Proceedings of an International Conference, held at Jülich, Germany, September 3–5, 1991*. Springer, 1993, pp. 44–61.
- [358] Russell, M. et al. “The effects of oil exploration and production in the Fladen Ground: Composition and concentration of hydrocarbons in sediment samples collected during 2001 and their comparison with sediment samples collected in 1989”. In: *Marine Pollution Bulletin* 50.6 (2005), pp. 638–651.
- [359] Ryabov, A. et al. “Estimation of functional diversity and species traits from ecological monitoring data”. In: *Proceedings of the National Academy of Sciences* 119.43 (2022), e2118156119.
- [360] Santos, F. C., Rodrigues, J., and Pacheco, J. M. “Graph topology plays a determinant role in the evolution of cooperation”. In: *Proceedings of the Royal Society B: Biological Sciences* 273.1582 (2006), pp. 51–55.
- [361] Sayama, H. *Introduction to the modeling and analysis of complex systems*. Open SUNY Textbooks, 2015.
- [362] Scheffer, M. et al. “Catastrophic shifts in ecosystems”. In: *Nature* 413.6856 (2001), pp. 591–596.

- [363] Scherer-Lorenzen, M. “Biodiversity and ecosystem functioning: basic principles”. In: *Biodiversity: Structure and Function. Encyclopedia of Life Support Systems*. EOLSS Publications, 2009.
- [364] Schleuter, D. et al. “A user’s guide to functional diversity indices”. In: *Ecological monographs* 80.3 (2010), pp. 469–484.
- [365] Schmera, D., Erős, T., and Podani, J. “A measure for assessing functional diversity in ecological communities”. In: *Aquatic Ecology* 43 (2009), pp. 157–167.
- [366] Schmidt, J. “The breeding places of the eel”. In: *Smithson Inst Ann Rep 1924* (1925), pp. 279–316.
- [367] Schulze, E.-D. and Mooney, H. A. *Biodiversity and Ecosystem function*. Springer-Verlag, New York, New York, USA, 1993.
- [368] Scott, J. and Heikkonen, R. “Estimating age at first maturity in fish from change-points in growth rate”. In: *Marine Ecology Progress Series* 450 (2012), pp. 147–157.
- [369] Setta, S. P. et al. “Oligotrophic waters of the Northwest Atlantic support taxonomically diverse diatom communities that are distinct from coastal waters”. In: *Journal of Phycology* 59.6 (2023), pp. 1202–1216.
- [370] Sguotti, C. et al. “Distribution of skates and sharks in the North Sea: 112 years of change”. In: *Global change biology* 22.8 (2016), pp. 2729–2743.
- [371] Shannon, C. E. “A mathematical theory of communication”. In: *The Bell system technical journal* 27.3 (1948), pp. 379–423.
- [372] Sharma, P. et al. “Diversity and evolution of nitrogen fixing bacteria”. In: *Sustainable Agriculture Reviews 60: Microbial Processes in Agriculture* (2023), pp. 95–120.
- [373] Siepielski, A. M. and McPeck, M. A. “On the evidence for species coexistence: a critique of the coexistence program”. In: *Ecology* 91.11 (2010), pp. 3153–3164.
- [374] Sifa, L. and Mathias, J. “The critical period of high mortality of larvae fish—a discussion based on current research”. In: *Chinese Journal of Oceanology and Limnology* 5 (1987), pp. 80–96.
- [375] Sillmann, J. et al. “Understanding, modeling and predicting weather and climate extremes: Challenges and opportunities”. In: *Weather and climate extremes* 18 (2017), pp. 65–74.
- [376] Simpson, E. H. “Measurement of diversity”. In: *nature* 163.4148 (1949), pp. 688–688.
- [377] Sjöberg, K. “Ecology of the European river lamprey (*Lampetra fluviatilis*) in northern Sweden”. In: *Canadian Journal of Fisheries and Aquatic Sciences* 37.11 (1980), pp. 1974–1980.
- [378] Smayda, T. J. and Reynolds, C. S. “Community assembly in marine phytoplankton: application of recent models to harmful dinoflagellate blooms”. In: *Journal of plankton research* 23.5 (2001), pp. 447–461.
- [379] Snoeijs, P. and Weckström, K. “Diatoms and environmental change in large brackish-water ecosystems”. In: *The Diatoms: Applications for the Environmental and Earth Sciences, edited by: Smol, JP and Stoermer, E F., Cambridge University Press, Cambridge* (2010), pp. 287–308.
- [380] Sokal, R. R. “Distance as a measure of taxonomic similarity”. In: *Systematic Zoology* 10.2 (1961), pp. 70–79.
- [381] Solé, R. and Goodwin, B. *Signs of life: How complexity pervades biology*. Basic Books, 2000.



- [382] Solé, R. V. and Bascompte, J. *Self-organization in complex ecosystems (Monographs in population biology)*. Princeton University Press, 2006.
- [383] Song, Y. et al. “Relationships between functional diversity and ecosystem functioning: A review”. In: *Acta Ecologica Sinica* 34.2 (2014), pp. 85–91.
- [384] Sorzano, C. O. S., Vargas, J., and Montano, A. P. “A survey of dimensionality reduction techniques”. In: *arXiv preprint arXiv:1403.2877* (2014).
- [385] Spearman, C. “The proof and measurement of association between two things”. In: *The American journal of psychology* 100.3/4 (1987), pp. 441–471.
- [386] Srivastava, D. S. and Vellend, M. “Biodiversity-ecosystem function research: is it relevant to conservation?” In: *Annu. Rev. Ecol. Evol. Syst.* 36.1 (2005), pp. 267–294.
- [387] Stokes, D. et al. *Irish Groundfish Survey Cruise Report, Sept. 24th–Dec. 17th, 2014*. Tech. rep. Marine Institute, 2015.
- [388] Straile, D., Jochimsen, M. C., and Kümmerlin, R. “The use of long-term monitoring data for studies of planktonic diversity: a cautionary tale from two Swiss lakes”. In: *Freshwater Biology* 58.6 (2013), pp. 1292–1301.
- [389] Stratoudakis, Y. et al. “Differences in life history features of long rough dab *Hippoglossoides platessoides* within Scottish waters”. In: *Marine Ecology Progress Series* 158 (1997), pp. 303–306.
- [390] Strogatz, S. H. “Exploring complex networks”. In: *nature* 410.6825 (2001), pp. 268–276.
- [391] Tandon, P., Abrol, Y. P., and Kumaria, S. *Biodiversity and its significance*. IK International Pvt Ltd, 2007.
- [392] Tansley, A. G. “The use and abuse of vegetational concepts and terms”. In: *Ecology* 16.3 (1935), pp. 284–307.
- [393] Ter Hofstede, R., Hiddink, J. G., and Rijnsdorp, A. D. “Regional warming changes fish species richness in the eastern North Atlantic Ocean”. In: *Marine Ecology Progress Series* 414 (2010), pp. 1–9.
- [394] Teschke, K. et al. “CRITTERBASE, a science-driven data warehouse for marine biota”. In: *Scientific Data* 9.1 (2022), p. 483.
- [395] Thomas, M. K. et al. “Temperature–nutrient interactions exacerbate sensitivity to warming in phytoplankton”. In: *Global change biology* 23.8 (2017), pp. 3269–3280.
- [396] Tillmann, U., Hesse, K.-J., and Colijn, F. “Planktonic primary production in the German Wadden Sea”. In: *Journal of Plankton Research* 22.7 (2000), pp. 1253–1276.
- [397] Tillmann, U. and Rick, H.-J. “North Sea phytoplankton: a review”. In: *Senckenbergiana maritima* 33.1 (2003), pp. 1–69.
- [398] Tilman, D. “Competition and biodiversity in spatially structured habitats”. In: *Ecology* 75.1 (1994), pp. 2–16.
- [399] Tilman, D. “Distinguishing between the effects of species diversity and species composition”. In: *Oikos* 80.1 (1997).
- [400] Tilman, D. “Functional diversity”. In: *Encyclopedia of Biodiversity*. Academic Press, 2001, pp. 109–120.
- [401] Tilman, D. et al. “The influence of functional diversity and composition on ecosystem processes”. In: *Science* 277.5330 (1997), pp. 1300–1302.
- [402] Tittensor, D. P. et al. “A mid-term analysis of progress toward international biodiversity targets”. In: *Science* 346.6206 (2014), pp. 241–244.

- [403] Tokeshi, M. *Species Coexistence: Ecological and Evolutionary Perspectives*. Blackwell Science, 1999.
- [404] Tokeshi, M. and Schmid, P. E. “Niche division and abundance: an evolutionary perspective”. In: *Population Ecology* 44.3 (2002), pp. 189–200.
- [405] Trippel, E. A. and Harvey, H. H. “Comparison of methods used to estimate age and length of fishes at sexual maturity using populations of white sucker (*Catostomus commersoni*)”. In: *Canadian Journal of Fisheries and Aquatic Sciences* 48.8 (1991), pp. 1446–1459.
- [406] Tylianakis, J. M. and Morris, R. J. “Ecological networks across environmental gradients”. In: *Annual Review of Ecology, Evolution, and Systematics* 48 (2017), pp. 25–48.
- [407] “UNEP/CBD/COP/DEC/X/2 2010”. In: 2010.
- [408] Urban, D. and Keitt, T. “Landscape connectivity: a graph-theoretic perspective”. In: *Ecology* 82.5 (2001), pp. 1205–1218.
- [409] Urlich, S. C. “What’s the end-game for biodiversity: is it time for conservation evolution?” In: *New Zealand Journal of Ecology* 39.1 (2015), pp. 133–142.
- [410] Usman, S., Muhammad, Y., and Chiroman, A. “Roles of soil biota and biodiversity in soil environment—A concise communication”. In: *Eurasian Journal of Soil Science* 5.4 (2016), pp. 255–265.
- [411] Valls, M., Rueda, L., and Quetglas, A. “Feeding strategies and resource partitioning among elasmobranchs and cephalopods in Mediterranean deep-sea ecosystems”. In: *Deep Sea Research Part I: Oceanographic Research Papers* 128 (2017), pp. 28–41.
- [412] Van Beusekom, J. and De Jonge, V. “Long-term changes in Wadden Sea nutrient cycles: importance of organic matter import from the North Sea”. In: *Nutrients and Eutrophication in Estuaries and Coastal Waters: Proceedings of the 31st Symposium of the Estuarine and Coastal Sciences Association (ECSA), held in Bilbao, Spain, 3–7 July 2000*. Springer. 2002, pp. 185–194.
- [413] Van Beusekom, J. et al. “The importance of sediments in the transformation and turnover of nutrients and organic matter in the Wadden Sea and German Bight”. In: *Deutsche Hydrographische Zeitschrift* 51.2-3 (1999), pp. 245–266.
- [414] Van Beusekom, J. E. et al. “Wadden Sea Eutrophication: long-term trends and regional differences”. In: *Frontiers in Marine Science* 6 (2019), p. 370.
- [415] Van Walraven, L. et al. “Long-term patterns in 50 years of scyphomedusae catches in the western Dutch Wadden Sea in relation to climate change and eutrophication”. In: *Journal of Plankton Research* 37.1 (2015), pp. 151–167.
- [416] Vellend, M. “Conceptual synthesis in community ecology”. In: *The Quarterly review of biology* 85.2 (2010), pp. 183–206.
- [417] Verleysen, M. and François, D. “The curse of dimensionality in data mining and time series prediction”. In: *International work-conference on artificial neural networks*. Springer. 2005, pp. 758–770.
- [418] Villéger, S. et al. “Contrasting changes in taxonomic vs. functional diversity of tropical fish communities after habitat degradation”. In: *Ecological applications* 20.6 (2010), pp. 1512–1522.
- [419] Violle, C. et al. “Let the concept of trait be functional!” In: *Oikos* 116.5 (2007), pp. 882–892.

- [420] Vogt, J. C., Albach, D. C., and Palinska, K. A. “Cyanobacteria of the Wadden Sea: seasonality and sediment influence on community composition”. In: *Hydrobiologia* 811 (2018), pp. 103–117.
- [421] Walker, B., Kinzig, A., and Langridge, J. “Plant attribute diversity, resilience, and ecosystem function: the nature and significance of dominant and minor species”. In: *Ecosystems* 2 (1999), pp. 95–113.
- [422] Wallace, A. R. *The Malay Archipelago: the land of the Orang-Utan, and the Bird of Paradise, a narrative of travel, with studies of man and nature*. London: Macmillan, 1869.
- [423] Walsh, S. J. “Life history traits and spawning characteristics in populations of long rough dab (American plaice) *Hippoglossoides platessoides* (Fabricius) in the North Atlantic”. In: *Netherlands Journal of Sea Research* 32.3-4 (1994), pp. 241–254.
- [424] Wang, L., Seelbach, P. W., and Hughes, R. M. “Introduction to landscape influences on stream habitats and biological assemblages”. In: *American Fisheries Society Symposium*. Vol. 48. American Fisheries Society. 2006, p. 1.
- [425] Watts, D. J. and Strogatz, S. H. “Collective dynamics of ‘small-world’ networks”. In: *nature* 393.6684 (1998), pp. 440–442.
- [426] Weber, M. G. and Strauss, S. Y. “Coexistence in close relatives: beyond competition and reproductive isolation in sister taxa”. In: *Annual Review of Ecology, Evolution, and Systematics* 47.1 (2016), pp. 359–381.
- [427] Weijerman, M., Lindeboom, H., and Zuur, A. F. “Regime shifts in marine ecosystems of the North Sea and Wadden Sea”. In: *Marine Ecology Progress Series* 298 (2005), pp. 21–39.
- [428] Westerhoff, H. V. et al. “Systems biology: the elements and principles of life”. In: *FEBS letters* 583.24 (2009), pp. 3882–3890.
- [429] Wheeler, A. *Key to the fishes of Northern Europe*. Frederick Warne. London, 1978.
- [430] Whitehorn, P. R. et al. “Mainstreaming biodiversity: A review of national strategies”. In: *Biological conservation* 235 (2019), pp. 157–163.
- [431] Wilson, E. O. *The diversity of life*. Harvard University Press, 1992.
- [432] Wiltshire, K. H. and Dürselen, C.-D. “Revision and quality analyses of the Helgoland Reede long-term phytoplankton data archive”. In: *Helgoland Marine Research* 58 (2004), pp. 252–268.
- [433] Winder, M. and Cloern, J. E. “The annual cycles of phytoplankton biomass”. In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 365.1555 (2010), pp. 3215–3226.
- [434] Winemiller, K. O. and Rose, K. A. “Patterns of life-history diversification in North American fishes: implications for population regulation”. In: *Canadian Journal of Fisheries and aquatic sciences* 49.10 (1992), pp. 2196–2218.
- [435] Wirth, T. and Bernatchez, L. “Genetic evidence against panmixia in the European eel”. In: *Nature* 409.6823 (2001), pp. 1037–1040.
- [436] Wolff, W. J. et al. “The Wadden Sea Quality Status Report-Synthesis Report 2010”. In: *The Wadden Sea 2010. Common Wadden Sea Secretariat (CWSS); Trilateral Monitoring and Assessment Group: Wilhelmshaven. (Wadden Sea Ecosystem; 29/editors, Harald Marencic and Jaap de Vlas)*. Vol. 25. CWSS, 2010.

- [437] Worm, B. et al. “Impacts of biodiversity loss on ocean ecosystem services”. In: *science* 314.5800 (2006), pp. 787–790.
- [438] Xu, M., Cai, H., and Liang, S. “Big data and industrial ecology”. In: *Journal of Industrial Ecology* 19.2 (2015), pp. 205–210.
- [439] Yachi, S. and Loreau, M. “Biodiversity and ecosystem productivity in a fluctuating environment: the insurance hypothesis”. In: *Proceedings of the National Academy of Sciences* 96.4 (1999), pp. 1463–1468.
- [440] Ye, L. et al. “Increasing zooplankton size diversity enhances the strength of top-down control on phytoplankton through diet niche partitioning”. In: *Journal of Animal Ecology* 82.5 (2013), pp. 1052–1061.
- [441] Zhang, H. et al. “Dispersal network heterogeneity promotes species coexistence in hierarchical competitive communities”. In: *Ecology Letters* 24.1 (2021), pp. 50–59.
- [442] Zhang, X. et al. “Environmental factors indirectly reduce phytoplankton community stability via functional diversity”. In: *Frontiers in Ecology and Evolution* 10 (2022), p. 990835.
- [443] Ziólkowska, E. et al. “Effects of different matrix representations and connectivity measures on habitat network assessments”. In: *Landscape ecology* 29 (2014), pp. 1551–1570.

# Eidesstattliche Erklärung

Ich versichere, dass ich die vorliegende Arbeit selbständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt und die allgemeinen Prinzipien wissenschaftlicher Arbeit und Veröffentlichungen, wie sie in den Leitlinien guter wissenschaftlicher Praxis der Carl von Ossietzky Universität Oldenburg festgelegt sind, befolgt habe.

Pedro Manuel Carrasco De La Cruz  
Oldenburg, den 27. September 2024