FACULTY OF ENGINEERING AND ARCHITECTURE

Marine Soundscapes in Shallow Water: Automated Tools for Characterization and Analysis

Clea Parcerisas Serrahima

Doctoral dissertation submitted to obtain the academic degree of
Doctor of Engineering

Supervisors

Prof. Dick Botteldooren, PhD* - Elisabeth Debusschere, PhD** - Prof. Paul Devos, PhD*

* Department of Information Technology
Faculty of Engineering and Architecture, Ghent University

** VLIZ

June 2024

GHENT UNIVERSITY

ISBN 978-94-6355-848-8
NUR 962, 969
Wettelijk depot: D/2024/10.500/53

Members of the Examination Board

Chair

Prof. Filip De Turck, PhD, Ghent University

Other members entitled to vote

Prof. Philippe Blondel, PhD, University of Bath, United Kingdom
Olaf Boebel, PhD, Alfred-Wegener-Institut, Germany
Prof. Lucia Di Iorio, PhD, Université de Perpignan Via Domitia, France
Prof. Nilesh Madhu, PhD, Ghent University
Prof. Timothy Van Renterghem, PhD, Ghent University

Supervisors

Prof. Dick Botteldooren, PhD, Ghent University
Elisabeth Debusschere, PhD, VLIZ
Prof. Paul Devos, PhD, Ghent University

Acknowledgment

> *"We are relearning and rediscovering what indigenous communities have long known about the importance of dialogue"*
> Karen Bakker

When I was faced with the question of "what-will-I-do-when-I-finish-studying?" I was puzzled to see that most of my options as an electromechanical engineer were to optimize and design production. Better, or more, or faster, production. And I did not want to do that. We live in a world where we produce too much, too fast and too careless. I thought, maybe research on something which can't be produced. But research in what? Trying to follow my father's advice, I thought it was a good idea to combine it with another of my passions, music. So I started to look for a master thesis which would somewhat be related to music. However, probably due to my bad searching skills, I ended up discovering something called bioacoustics. And, to my surprise, I discovered that there was a research lab at scarcely one hour from Barcelona called Laboratori d'Aplicacions Bioacustiques (LAB) where they did *marine* bioacoustics research. Once I heard this existed, I had no doubt; I wanted to do nothing else in my life. It combined my passion for music, environmental responsibility, my love for the sea and for programming. I therefore want to thank all the people from the LAB: Michel André, Marta Solé, Mike van de Schaar, Florence Erbs and Steffen De Vreese for introducing me to this fascinating topic.

After my master thesis I moved to Leuven for a one year internship at imec, where I got to play around with underwater sound waves propagation and acoustic holograms. I feel very grateful for having had this opportunity. During this period I was living with Bri, Gil and Niel, who became the main personal reason why I would leave my beloved Mediterranean for rainy Belgium a year later. I hold so dearly our time living together, and there was a before and after that year where I understood this was how I wanted to live. In the living room of that house I wrote a proposal for a PhD which I submitted to obtain individual funding in Spain, and got eventually rejected. However, once already back home in Ibiza after my internship I heard from a PhD position at the Flanders Marine Institute (VLIZ) which had

exactly the same topic. Literally, exactly the same. I would have written whatever it was necessary to get that position, but I just had to copy paste the cover letter which I had already written, change the name of the institute, and send my CV (I hope revealing this does not take my position back!). Eventually, I got this position and moved back to Belgium to start the path of this PhD.

As a result, this thesis is the outcome of 4.5 years of research conducted under the supervision of Dr. Elisabeth Debusschere, Prof. Paul Devos and Prof. Dick Botteldooren. Here I want to take the time to express my gratitude to the people who played an important role in my life during these years.

First, I would like to thank my supervisors for providing me the opportunity to conduct this PhD, and for helping me when I was stuck. The quality of this thesis improved thanks to the jury members who carefully read the text, provided me with valuable comments and asked me challenging question on the preliminary defense. I would like to thank Prof. Lucia Di Iorio, Dr. Olaf Boebel, Prof. Philippe Blondel, Prof. Nilesh Madhu and Prof. Timothy Van Renterghem. Your valuable feedback helped me improve this manuscript, and steered this research in the right direction for the future.

Despite the gender gap in STEM, during this professional path I met smart, strong and wonderful women from whom I learned a lot and who I deeply admire. Eventually, I think I can also call them friends, which makes me feel very honored. Thank you Elisabeth for your support and supervision, I would have never managed without you. Irene, thank you for being the best cabin mate one could wish for in the Polarstern and for encouraging me to go forward and give me the most valuable feedback. Elena, I am so grateful that I got to meet you and work together. Thank you for pointing the path every time I get side-tracked (happens way too often), and for teaching me so much. Working with you is fantastic and I hope we can continue doing so in the future.

At some point, I was a bit stuck and needed a change. Luckily, I was accepted to do a 6 months research stay at the Ocean Acoustics group at the Alfred Wegener Institute (AWI), which was financed by FWO. There I got to work with experts in marine bioacoustics such as Ilse van Opzeeland, Olaf Boebel, Stefanie Spiesecke and Elena Schall who helped me broaden my horizons. I want to thank all of them for welcoming me to their team. During this research stay I got the amazing opportunity to join their team on a 2-month Antarctic cruise with the RV Polarstern. During the very enjoyable (but sometimes rough) days onboard, I cherished the CTD shifts with Ole, Pedro and Irene. There I got to know as well David Barnes, who became for me an example of integrity, good science and humbleness.

Even though it took 3 years before I could attend my first conference because of the pandemic (and my slow progress), from the early beginning of my PhD I could attend the yearly small gathering called “bioacoustics day”. These days helped me build a network, learn a lot, and put my research into perspective. I will not do an extensive list of all the participants, but I thank all the people from Leiden University, WUR, TNO, Jasco, Kiel University, and Groningen University who regularly attend these meetings. For this reason, when I finally could attend the first “Effects of Noise on Aquatic Life” and “OCEANOISE” conferences, I was glad I

already knew some faces and did not need to be faced with socializing from scratch. I enjoyed the time we shared with Saskia, Karen, Kate and Emilie, who made me feel I was not alone. Another familiar face at conferences is Özkan, who I would also like to thank for the time he took to review the parts about sound propagation.

Despite my deep conviction in open-source publishing, I was (and still am) very surprised the day I received an email from John Ryan from MBARI asking if they could use my software *pypam* to process their data. This lead to a long collaboration where the software engineers of MBARI Carlos Rueda and Danelle Cline made some cloud magic where they used the package I had made to process sound into hybrid millidecade band levels in the cloud. This eventually led the participation of *pypam* in the SoundCoop project, and I was invited to Boulder (USA) to their annual workshop. I would like to take some time to thank Carrie Wall and Megan McKenna for organizing this and for such valuable feedback and information. John and Carlos, I am so grateful I have worked with you. I appreciate your clarity, enthusiasm, dedication and expertise, and how you manage to combine all of these into a relaxed and sweet personality. I want to especially thank John who took the time to read this dissertation and provide his always valuable feedback.

This PhD would not have been half as complete without the multiple field work days and resulting acquired data. I have been on board of the RV Simon Stevin, RIB Zeekat, RV Polarstern and the Sailboat Capoeira. No acoustic recording would have been possible without the crew members of Vloot, Polarstern, Jan Vermaut and the people from Sailawaywithme, so I would like to thank them for their constant work, support and help. Of course this fieldwork was not only with the crew members but also with the Marine Observation Centre (MOC) team, who have made from rough sea days an amazing experience. Thank you to Klaas, Jan, Elisabeth, Jonas, Anouk, Nick, Bram, Carlota, David, Lotte, Arienne, Wout, Hanneloor and Rune. Days at sea have been double enjoyable in your company. Bram, MOC (and myself) is so lucky with your incorporation, your arrival brought so much serenity to our team.

From the MOC, I would like to also mention my gratitude to everyone who has eventually participated in acoustics research and made from this path a bit of a less lonely journey. Thank you to Wout for the AIS acoustic analysis, Arienne for the clustering effort and Julia and Martin for deciding to come work with us. Furthermore, I had the chance of learning from the students Quentin Hamard, Simon Laurent and Svenja Woehle who came to do internships with us. And of course, this PhD would not have been possible without all the support services at VLIZ.

This PhD path has been full or work, but also full with fun moments which have led to beautiful friendships. My experience at VLIZ, and in Belgium in general would have never been the same without my beloved colleagues Marie, Sarah, Marine, Rune, Gizem, Nelle, Hisham, Roeland, Anouk, Jens, Michiel, Ruben, Salva, Claudia, and Peter. Dear friends, the lock-down would just not have been mentally bearable without your constant support. I cherish the relationships we have built, which I am proud to call close friendships. And I cherish the professional advice, support, hugs and laughs which has made this path an enjoyable one. Sarah, I want to thank you for our study conducted together and for your patience with my absolutely lack of lab skills. I really feel like working together was synergistic, and despite the long working hours, fun.

Even though I did not spend a lot of time at the UGent office, there I met Sergei and Sam, who have been my anchor to the waves group every time I have drifted, and eventually became good friends. And of course, there I got the fantastic support of Judith, who always found time to explain me the more practical things of academia.

During this PhD I have also had the privilege to collaborate with multiple people from different places and backgrounds. From one of these many cool collaborations I have fallen into, I met Randall, who I want to thank for helping me believe another Academia is possible. Other collaborations have often been with artists, who have made my work a lot richer and fulfilling. I would like to thank all of them for bringing this research to a more emotional level to which people can really connect.

After the tiring days at sea or long working ours, I came home to the Willem Tell. Living there was fun, warm and safe. Thank you Benoît, Ida, David and Julie for the delicious food, the amazing support and the quietness of our home. Living together converted Belgium into a home I cherish, and gave me amazing friends. I can't describe how grateful I feel for our time shared together. I look forward to the times to come.

The last couple of years I moved to Oostende to the Boldershof, where I have slowly adapted. I want to thank everyone there for teaching me other points of view, and for showing that we as a group can still be strong despite the differences. Our garden is the place where I find peace. A special thanks to Bram, Luca, Thomas and Ann-Sofie (and their amazing offspring) for pushing forward another concept of co-housing.

Not only I am lucky because I can work in an field which I find fascinating, but I also have been very lucky all my life with a loving and caring family who has supported me in all the decisions I have made. Tinc la sort d'haver crescut entre tres famílies fantàstiques, la Parcerisas, la Serrahima i la de Can Nebot. Marta i Jordi, gràcies per tot l'amor que m'heu donat, mai hagués après a caminar (ni a nadar) sense vosaltres. Segur que estar envoltades de dofins al Mediterrani acabada de néixer té alguna cosa a veure amb el camí professional que he triat. Marta, gràcies per ensenyar-me que vol dir ser forta, i també que ser vulnerable pot ser una altra fortalesa. Aprenc encara cada dia de tu. Jordi, el teu record m'acompanya cada dia quan surto al mar.

These last years my luck has increased (even more!) with the warm adoptive family I have fallen into. Bedankt Veronique en Dirk voor al jullie ondersteuning. Bij jullie voel ik me thuis, en jullie huis is een warme, gastvrij nest. Het maakt me gelukkig dat ik jullie, Frederik, Camillie, Hannes en Eva familie mag noemen.

Roeland, it was hard to choose in which section to thank you. If I would thank you as a colleague, as a friend, as a housemate or as family. I will not mention here anything here that you (or anyone else, for what matters) don't know already. Thank you, you inspire me.

And of course I would not be who I am if it was not for the constant love and support I have received throughout my life from the people who have always been there, seeing me grow, growing together, helping me broaden my view on life and holding me every time I fell. Gràcies a la Marina, la Gala, la Clara, l'Helena, la Iona, la Monica i el Dani per enriquir tant la meva vida. Us admiro i us estimo, la nostra amistat és el meu pilar per resistir i viure.

And finally I would like to thank my dear friend Helena Bosch who has managed to create the beautiful image on the cover you have in your hands.

Oostende, June; 2024
Clea Parcerisas Serrahima

Table of Contents

List of Figures

List of Tables

List of Acronyms

A

AC	Alternating Current
ACI	Acoustic Complexity Index
ADI	Acoustic Diversity Index
AEI	Acoustic Evenness Index
AI	Artificial Intelligence
AIS	Automatic Identification System
AUV	Autonomous Underwater Vehicle

B

BPNS	Belgian Part of the North Sea

C

CNN	Convolutional Neural Network

D

DBSCAN	Density-Based Spatial Clustering of Applications with Noise
DC	Direct Current
DCS	Detection and Classification System
DL	Deep Learning

E

EEZ	Exclusive Economic Zone
EOV	Essential Ocean Variable
ETN	European Tracking Network

F

FFT	Fast Fourier Transform

FSW	Filtered Seawater

G

GB	Gigabyte
GES	Good Environmental Status
GLUBS	Global Library of Underwater Biological Sounds
GMM	Gaussian Mixture Model

H

HMB	Hybrid Millidecade Bands
HMBL	Hybrid Millidecade Bands (sound pressure) Levels

I

IQOE	International Quiet Ocean Experiment
ISO	International Standardization Organization

L

LOBE	Level of Onset of Biologically adverse Effects
LTSA	Long Term Spectral Average

M

ML	Machine Learning
MMSI	Maritime Mobile Service Identity
MSFD	Marine Strategy Framework Directive

P

PAM	Passive Acoustic Monitoring
PB	Petabyte
PSD	Power Spectral Density

R

RF	Random Forest
RV	Research Vessel
RMS	Root Mean Squared value

S

SHAP	SHapely Additive exPlanations
SNR	Signal to Noise Ratio
SL	Source Level (ISO 18405:2017)
SPD	Spectral Probability Density
SPL	Sound Pressure Level (ISO 18405:2017)
SVM	Support Vector Machine

T

TB	Terabyte

U

UMAP	Uniform Manifold Approximation and Projection
USV	Uncrewed Surface Vehicle

V

VLIZ	Vlaams Instituut voor de Zee (Flanders Marine Institute)

X

XAI	Explainable Artificial Intelligence

Y

YOLO	You Only Look Once

W

WoRMS	World Register of Marine Species

Nederlandse samenvatting

In tegenstelling tot licht, reist geluid verder onder water dan in de lucht. Hoewel geluid belangrijk is voor de meeste dieren op deze planeet, speelt het een bijzondere rol voor onder water fauna. Zo gebruiken mariene fauna geluid om te communiceren, te paren, te foerageren, zich te oriënteren en roofdieren te vermijden. Mariene soundscapes zijn de samenstelling van alle sonifere informatie uit een ecosysteem, zowel van biotische als van abiotische en antropogeen oorsprong. Daarom kan het beluisteren van natuurlijke soundscapes worden gebruikt als een niet-invasief instrument om mariene habitats te bestuderen. Met de toenemende menselijke druk op de oceaanen verdwijnen ongerepte soundscapes in een ongekend tempo. Een deel van deze verandering wordt toegeschreven aan het stiller worden van ecosystemen in verval, een andere reden is de door de mens veroorzaakte geluidsvervuiling op zee die natuurlijke geluiden maskeert.

De impact van akoestische vervuiling op zee is de afgelopen drie decennia zorgwekkend geworden, met name in ondiepe, kustgebieden waar verschillende menselijke activiteiten samenkomen. Daarom is het begrijpen van de veerkracht en de dynamiek van een ecosysteem essentieel voor behoud en duurzame exploitatie. Het karakteriseren en analyseren van de soundscape is cruciaal om menselijke impact te kwantificeren.

Om de langetermijndynamiek van soundscapes te bestuderen, zijn langetermijnopnames nodig. Echter, het handmatig analyseren van maanden en jaren aan gegevens is zeer tijdintensief en tijdrovend, onderzoekers kunnen de snelheid waarmee grote volumes aan data verzameld wordt niet bijhouden. Daarom zijn automatische methoden nodig om deze soundscape data te verwerken en te analyseren. Dit proefschrift richt zich op het ontwikkelen van automatische methoden om soundscapes van zeer ondiepe en drukke kustgebieden te analyseren, specifiek gericht op de soundscape van het Belgische deel van de Noordzee (BPNS). We analyseren welke van de huidige methoden toegepast kunnen worden onder welke omstandigheden en we ontwikkelen nieuwe methodologieën die beter geschikt zijn voor het doel van de studie in de BPNS. Alle voorgestelde benaderingen zijn complementaire manieren om soundscapes te beschrijven en te begrijpen.

Het beschrijven van mariene soundscapes alleen op basis van geluidsdrukniveaus leidt tot maar gedeeltelijke representaties van de data en beperkt het begrip van de akoestisch kenmerkende eigenschappen. Het is nogal noodzakelijk om soundscapesanalyse te combineren met contextuele informatie om zinvolle conclusies te kunnen trekken. Voordat we nieuwe methoden ontwikkelden om het geluidslandschap van de BPNS te bestuderen, hebben we daarom de biologische en milieukenmerken van het gebied bestudeerd die geluid produceren of beïnvloeden (Hoofdstuk 3). Deze eerste verkenning leidt tot de conclusie dat in ondiepe en sterk geëxploiteerde mariene gebieden verschillende uitdagingen samenkomen voor het meten en analyseren van onderwatergeluid. Biofouling, stromingsgeluid, onbekende geluidsbronnen en geluidsmaskering ondermijnen de voortplanting of ontvangst van geluid. Met uitzondering van sommige zeezoogdieren zijn de sonifere kenmerken van de meeste dieren die aanwezig zijn in de BPNS onbekend. Bovendien zijn er voor de overgrote meerderheid van akoestisch bestudeerde dieren geen gedetailleerde beschrijvingen en grote databases met geluidsopnames.

Soundscapekarakterisering kan aangepakt worden vanuit een holistisch oogpunt of door de verschillende elementen te ontwarren. Holistische analyse richt zich meer op de geïntegreerde geluidskarakteristieken, terwijl bij het focussen op de verschillende elementen afzonderlijk nauwkeurigere informatie kan worden verkregen.

Het tweede deel van dit proefschrift verkent de soundscapes van de BPNS vanuit een holistische benadering. In hoofdstuk 4 beschrijven we de BPNS-soundscape en de belangrijkste elementen ervan, waarbij we gebruik maken van de meest voorkomende benaderingen in de literatuur. Deze beschrijving gebeurt met behulp van 1-minuut hybride millidecadale band niveaus berekend uit de langetermijngegevens verzameld in het kader van LifeWatch. Deze analyse brengt ons tot de conclusie dat het BPNS-soundscape sterk afhankelijk is van de locatie. De belangrijkste drijfveren aan de soundscape zijn stroming, scheepvaart en wind. De meest invloedrijke parameters zijn de nabijheid van een scheepvaartroute en de diepte van de opnamelocatie. Luide en korte impulsieve geluidsevenementen, zoals die van vissen, kunnen de niveaus van 1-minuut gemiddelde hybride millidecade band niveaus verhogen. Dit verhoogt echter niet de hybride millidecade band niveaus wanneer de mediaan wordt overwogen. Bijgevolg kunnen grote verschillen tussen gemiddelde en mediaan hybride millidecade band niveaus worden gebruikt als een indicatie voor de aanwezigheid van dergelijke korte, luide geluidsevenementen. In dit proefschrift hebben we aangetoond dat dergelijke geluidsevenementen sterk gecorreleerd zijn met de aanwezigheid van visgeluiden.

Om deze beschrijving aan te vullen, stellen we in hoofdstuk 5 een nieuwe methode voor semi-supervised categorisatie van ondiepe mariene soundscapes voor, met verdere interpretatie van deze categorieën volgens gelijktijdig gemeten milieuparameters met behulp van uitlegbare SHAP. De geluidscategorieën zijn gebaseerd op continu geluid of geluiden die vaak herhaald worden in de opnames. We testen deze methodologie op de twee datasets die voor dit proefschrift zijn verzameld, met de focus op verschillende temporele en ruimtelijke schalen. Eerst leggen we de focus op geluiden van enkele locaties maar op lange termijn schalen met een lage temporele resolutie (1 minuut). Vervolgens kijken we naar gegevens van meerdere locaties maar op korte termijn met een hoge temporele resolutie (5 seconden). Beide analyses resulteerden in duidelijk identificeerbare categorieën van soundscapes die kunnen worden verklaard door ruimtelijke en temporele milieuparameters, zoals afstand tot de kust, bathymetrie, getij of seizoen. De classificatie van soundscapes in betekenisvolle en begrijpbare categorieën vergemakkelijkt hun identificatie en interpretatie. Deze informatie kan nuttig zijn voor beleidsvorming of conservatieprogramma's.

Het derde deel van dit proefschrift richt zich op het analyseren van soundscapes door de verschillende akoestische elementen te ontwarren met behulp van akoestische geluidsevenementendetectie. Het detecteren van deze geluidsevenementen in real-life langetermijnakoestische datasets is geen triviaal taak, en hoewel sommige studies goede resultaten hebben getoond, is er momenteel geen geautomatiseerd hulpmiddel bruikbaar op alle locaties. We stellen voor om een state-of-the-art computer vision model (YOLOv8) te gebruiken om geluidgebeurtenissen te detecteren in een mariene omgeving in een overdrachtsleren paradigma. Dit wordt bereikt door het audio om te zetten naar spectrogrammen met behulp van schuiframen die langer zijn dan de verwachte geluidsevenementen van interesse. Voordat we echter duiken in het complexe soundscape van de BPNS, testen we eerst de toepasbaarheid van ons model uit op een open-access dataset van een gebied dat akoestisch goed bestudeerd is. In hoofdstuk 6 wordt het YOLOv8-model met succes getraind en getest op een dataset met handmatige annotaties van Antarctische baleinenwalvisoproepen. De verkregen prestaties zijn vergelijkbaar met andere open-source gepubliceerde modellen en menselijke annotaties.

Vervolgens gaan we in hoofdstuk 7 verder met de analyse van de BPNS-soundscape met behulp van geluidsevenementendetectie. De bijzonderheid van deze akoestische gegevens is dat de geluidsbronnen van interesse niet gekend zijn. Het verkennen van onderwatergeluidsgegevens om mogelijke geluidsevenementen van interesse te vinden en te identificeren kan zeer tijdsintensief zijn voor menselijke analisten. Om dit proces te versnellen, stellen we een nieuwe methodologie voor die eerst alle potentieel relevante onderwaterakoestische gebeurtenissen detecteert en ze op een onbegeleide manier clustert voordat ze handmatig worden gecontroleerd. Deze methode kan worden toegepast op nieuwe opnames als een hulpmiddel om bioakoestici te ondersteunen bij het identificeren van terugkerende geluiden bij het bestuderen van ruimtelijke en temporele patronen. Dit vermindert de tijd die onderzoekers nodig hebben om lange akoestische opnames manueel te verwerken en maakt een gerichtere analyse mogelijk. Het biedt ook een kader om soundscapes te monitoren, ongeacht of de geluidsbronnen gekend zijn of niet.

Om de geluidsevenementen te detecteren, passen we dezelfde techniek toe als gepresenteerd in het hoofdstuk 6 op de langetermijngegevens verzameld uit de BPNS. De verkregen detectieprestaties zijn vergelijkbaar met die van menselijke validators en robuust over verschillende locaties, als ook op onafhankelijke zoetwaterakoestische gegevens. Om de noodzaak van menselijke betrokkenheid bij het genereren van annotaties voor het trainen van het gebeurtenisdetectiemodel te minimaliseren, stellen we voor om een actieve leerstrategie (active learning) aan te nemen om de meest informatieve audiobestanden voor handmatige annotatie te identificeren en te selecteren. De selectie van bestanden met behulp van de actieve leerbenadering lijkt het model sneller te verbeteren dan wanneer ze willekeurig worden gekozen. Alle gedetecteerde gebeurtenissen met behulp van deze aanpak worden vervolgens op een onbegeleide manier geclusterd, waarbij verschillende geluidsklassen worden verkregen. Deze klassen worden handmatig herzien en hun ruimtelijke en temporele patronen worden geanalyseerd om mogelijke bronnen te identificeren, waardoor de tijd die nodig is voor het vinden van verschillende geluidsklassen aanzienlijk wordt verminderd. Uit de geanalyseerde gegevens vonden we verschillende geluidstypes die kunnen worden toegeschreven aan biologische bronnen.

Tenslotte, in hoofdstuk 8 duiken we in op een verder categoriseringsstrategie van gedetecteerde geluiden, in dit geval van een handmatig geannoteerde subset van geluidsevenementen in de BPNS. Handmatige annotaties en labels van niet-géidentificeerde geluiden zijn zeer inconsistent, omdat deze geluiden niet goed zijn gekarakteriseerd in de literatuur en het daarom moeilijk is om te beslissen of twee geluiden afkomstig zijn van dezelfde bron. Daarom is een strategie nodig om te beslissen welke klassen zinvol en clusterbaar zijn. Om dit te doen, is het kiezen van de juiste akoestische kenmerkrepresentatie belangrijk, evenals de geselecteerde hyper-parameters tijdens het clusteringsproces. Een laatste belangrijke opmerking is dat dimensiereductie noodzakelijk is bij het gebruik van deep learning kenmerken om zinvolle clusters te verkrijgen. De verkregen klassen met behulp van beide benaderingen uit hoofdstuk 7 en hoofdstuk 8 zijn vergelijkbaar, met twee duidelijke visgeluiden en enkele andere niet-géidenticeerde geluiden die mogelijks ook afkomstig kunnen zijn van biologische bronnen.

Delen II en III richten zich op verschillende methodologieën om soundscapes te analyseren en informatie te extraheren. Maar hoe kan deze informatie gebruikt worden? In deel IV van dit proefschrift geven we een voorbeeld van de noodzaak om soundscapes te begrijpen, te onderscheiden en te karakteriseren, en hoe deze informatie kan worden gebruikt. Steeds meer bewijs suggereert dat mariene ongewervelde dieren informatie uit oceaan-soundscapes gebruiken om vestigingsbeslissingen te nemen. Er wordt is gesteld dat sommige ongewervelde dieren distincte geluiden uit de soundscape gebruiken om naar geschikte habitats te navigeren. Daarom hebben we, als voorbeeld om te zien of BPNS-soundscapes een belangrijk rol kunnen spelen in de ontwikkeling van de lokale fauna, in een laboratoriumexperiment larven van de echte oester Magallana gigas blootgesteld aan verschillende waarheidsgetrouwe soundscapes, waaronder rif-, buitenrif- en bootopnames. Een vierde behandeling werd toegevoegd waar bootgeluiden kunstmatig werden toegevoegd aan de rifgeluiden. De resultaten tonen aan dat de soundscape waaraan larven werden blootgesteld invloed heeft op hun vestigingspercentage. Rif opnames vertoonden een hoger vestigingspercentage dan buitenrif, boten en geen geluid. De combinatie van rif en bootgeluiden vertoonde een lager vestigingspercentage dan rif geluiden alleen (niet-significant trend). Door de akoestische karakteristieken van de gewenste rifgeluiden te onderzoeken, veronderstellen we dat spectro-temporeale patronen in rif geluiden deze soort aanzet tot vestiging.

Over de verschillende aspecten van dit onderzoek willen we benadrukken dat biologische geluiden bijdragen aan het BPNS-soundscape ondanks de lawaaiige omgeving, en dat deze bijdragen op verschillende manieren kunnen worden gekwantificeerd en geëvalueerd. Bovendien willen we benadrukken dat ondiepe watersoundscapes anders zijn dan diepwatersoundscapes, en dus in overeenstemming hiermee moeten worden bestudeerd. Ondiepe watersoundscapes veranderen op kleine ruimtelijke en temporele schalen, en het bepalen van de locaties en het tijdsbestek en de resolutie van de studie zijn cruciaal om om de variabiliteit vast te leggen.

Tijdens het verloop van dit proefschrift laten we zien dat Machine Learning een geschikt instrument is om mariene soundscapes te bestuderen, en dat het noodzakelijk wordt bij het analyseren van grote hoeveelheden onderwaterakoestische gegevens. Onze bijdragen bevatten zowel begeleide als onbegeleide modellen en pipelines voor analyse van onderwater-soundscapes en clustering van geluidgebeurtenissen. State-of-the-art Machine Learning-technieken kunnen worden gebruikt om onderwater-soundscapes en hun individuele geluidselementen te begrijpen, verkennen, voorspellen en kwantificeren, taken die uitdagend en tijdrovend zijn om handmatig uit te voeren.

Note

De Nederlandse samenvatting was met de hulp van ChatGPT 3.5* geschreven. De tekst werd zorgvuldig gelezen, gecorrigeerd en goedgekeurd door de auteur en een moedertaalspreker.

*OpenAI. (2023). ChatGPT (Mar 14 version) [Large language model]. https://chat.openai.com/chat

English Summary

Contrary to light, sound travels further underwater than in air. While sound is important for most animals on this planet, it plays a special role for those underwater. Marine fauna use sound to communicate, mate, forage, orient, and avoid predators. Marine soundscapes are the composite of all the soniferous information from one ecosystem, including both biotic, abiotic and anthropogenic sounds. Therefore, eavesdropping into nature's soundscapes can be used as a non-invasive tool to study marine habitats. With the increasing human pressure applied to the ocean, pristine soundscapes are disappearing at an unprecedented rate. Part of this change is attributed to the silencing of ecosystems in decline, yet another reason is human-made noise pollution at sea, which masks natural sounds.

The impact of acoustic pollution at sea has been growing concern during the last three decades, particularly in shallow, coastal, areas where multiple human activities converge. For this reason, understanding the ecosystem's resilience and dynamics is essential for their preservation and sustainable exploitation. Characterizing and analyzing the soundscape is crucial to quantify this impact.

To study long-term dynamics of soundscapes, long-term recordings are necessary. However, manually analyzing months and years of data is highly time intensive, and researchers cannot keep up with the streaming of data collected. For this reason automatic methods are necessary to process and analyze these soundscapes. This dissertation focuses on developing automatic methods to analyze soundscapes of very shallow and busy coastal areas, specifically the soundscape of the Belgian Part of the North Sea (BPNS). We analyze which of the existing methods work under which circumstances and we develop new methodologies more suited to the purpose of study. All the proposed approaches are complementary ways of describing and understanding soundscapes.

Describing marine soundscapes simply based on sound pressure levels leads to incomplete representations of the data and limits the understanding of distinctive acoustic characteristics. It is rather necessary to combine soundscape analysis with contextual information to be able to extract meaningful conclusions. Prior to developing new methods to study the soundscape of the BPNS, we hence studied the biological and environmental characteristics of the area known to produce or influence sound (Chapter 3). This first exploration leads us to the conclusion that in shallow and heavily exploited marine areas several challenges converge for underwater sound measurement and analysis. Bio-fouling, flow-noise, unknown sound sources, shallow bathymetry, and masking compromise propagation or sound reception. Except for some marine mammals, the soniferous characteristics of most

of the animals known to be present in the BPNS remain unknown. Additionally, the vast majority of those which have been studied lack detailed descriptions and large databases with examples.

Soundscape characterization can be tackled from a holistic point of view or by disentangling its different elements. Holistic analysis focus more on the integrated sound characteristics, while when focusing on the different elements separately, more precise information can be obtained.

The second part of this dissertation explores the BPNS soundscapes from a holistic approach. In Chapter 4, using the most common approaches found in literature, we describe the BPNS soundscape and its main contributors. This description is done using 1-minute hybrid millidecade band levels computed from the long-term data collected in the framework of LifeWatch. This analysis brings us to the conclusions that the BPNS soundscape is strongly location-dependent. The main contributors to the soundscape are current, shipping, and wind. The most influential parameters are proximity to a shipping lane and depth of the recording location. Loud and short impulsive sound events, such as those from fish can raise the levels of 1-minute mean hybrid millidecade band levels. However, it does not raise the hybrid millidecade band levels when the median is considered. Consequently, large differences between mean and median hybrid millidecade band levels can be used as an indication for the presence of such short, loud events. In this thesis, we showed that such events strongly correlated with the presence of fish sounds.

To complement this description, in Chapter 5 we propose a new method for semi-supervised categorization of shallow marine soundscapes, with further interpretation of these categories according to concurrent environmental conditions using SHAP. The sound categories are based on continuous sound or sounds that are repeated frequently in the recordings. We test this methodology on the two datasets collected for this dissertation, focusing on different temporal and spatial scales. First, the focus was on sounds from few locations but long-term, while using a low temporal resolution (1 minute). Second, we looked into data from multiple locations but short-term, while using a high temporal resolution (5 seconds). Both analysis resulted in clearly identifiable categories of soundscapes that could be explained by spatial and temporal environmental parameters, such as distance to the shore, bathymetry, tide or season. The classification of soundscapes in meaningful and understandable categories facilitates their identification and interpretation. This information can be useful for policy making or conservation programs.

The third part of this thesis focuses on analyzing soundscapes by disentangling different acoustic elements, namely by acoustic event detection. Detecting these events in real-world long-term acoustic datasets is not a trivial task, and even though some studies have shown good results, there is currently no automated tool usable across locations. We propose to use a state-of-the-art computer vision model (YOLOv8) to detect sound events in a marine environment in a transfer-learning paradigm. This is achieved by converting the audio to spectrograms using sliding windows longer than expected sound events of interest. However, before diving into the complex soundscape of the BPNS, we first test the applicability of our proposal on an open dataset from an area well-studied acoustically. In Chapter 6 the YOLOv8 model is successfully trained and tested on a dataset containing manual annotated Antarctic baleen whale calls. The obtained performance is comparable to other open-source published models and human annotations.

In Chapter 7 we then proceed to analyze the BPNS soundscape using sound event detection. The particularity of these acoustic data is that the sound sources of interest are not known. The exploration of underwater sound data to find and identify possible sound events of interest can be highly time-intensive for human analysts. To speed up this process, we propose a novel methodology which first detects all the potentially relevant underwater acoustic events and clusters them in an unsupervised way prior to manual revision. This method can be applied to new collected data as a tool to help bioacousticians identify recurrent sounds while at the same time studying their spatio-temporal patterns. This reduces the time researchers need to go through long acoustic recordings and allows for a more targeted analysis. It also provides a framework to monitor soundscapes regardless of whether the sound sources are known or not.

To detect the events, we apply the same technique presented in Chapter 6 to the long-term data collected from the BPNS. The obtained detection performance is similar to that of human annotators and robust across locations, including a completely independent freshwater acoustic data. To minimize the need for human involvement in generating annotations for training the event detection model, we suggest adopting an active learning strategy to identify and choose the most informative audio files for manual annotation. The selection of files using the active learning approach seems to improve the model's performance faster than when randomly selecting them. All the detected events using this approach are then clustered in an unsupervised way, obtaining different sound classes. These classes are manually revised, and their spatio-temporal patterns are analyzed to identify possible sources, overall reducing the time investment needed to find different sound classes. For the analyzed data we found several sound types which could be attributed to biological sources.

Finally, in Chapter 8 we dive further into the categorization strategy of detected sounds, in this case from a manually annotated subset of events in the BPNS. Manual annotations and labels from unidentified sounds are highly inconsistent, due to the fact that these sounds have not been properly characterized in literature and therefore it is difficult to decide whether two sounds are from the same source. Therefore, a strategy is necessary to decide which classes are meaningful and clusterable. To do so, choosing the right acoustic feature representation is important, and so are the selected hyper-parameters during the clustering process. A last important note is that dimension reduction is necessary when using deep learning features to obtain meaningful clusters. The obtained classes using both approaches from Chapter 7 and Chapter 8 are similar, featuring two clear fish sounds and some other unidentified sounds which could also come from biological sources.

Parts II and III focus on different methodologies to analyze soundscapes and obtain information from them. But, how can this information be used? In part IV of this dissertation we provide an example to showcase the need to understand soundscapes, distinguish them and characterize them, and how can this information be used. Increasing evidence suggests that marine invertebrates use information from ocean soundscapes to decide on settlement. It has been hypothesized that some invertebrates use distinct soundscape sounds to navigate to suitable habitats. Therefore, as an example to see if BPNS soundscapes could play an important role in the development of local fauna, in a laboratory experiment we exposed larvae of true oyster *Magallana gigas* to different real-world soundscapes including reef, off-reef and boat recordings. A fourth treatment was added where boat sounds were artificially added to the reef sounds. The results show that the soundscape to which they were exposed to did have an effect on their settlement rate. Reef recordings had a higher settlement compared to off-reefs, boats and no sound. The mixture of reefs and boat sounds had a lower settlement rate than that of reefs (non-significant). Examining the acoustic characteristics of the preferred reef sounds, we hypothesize that spectro-temporal patterns are the driving attractive quality in reef sounds for this species.

Across the different aspects of this research, we want to highlight the fact that biological sounds do contribute to the BPNS soundscape despite the noisy environment, and that these contributions can be quantified and assessed in several different ways. Furthermore, we want to emphasize that shallow water soundscapes are different than deep water ones, and thus need to be studied accordingly. They change at small spatio-temporal scales, and deciding the locations and time frame and resolution of the study is crucial to capture the variability.

During the course of this thesis we show that Machine Learning is a suitable tool to study marine soundscapes, and that it becomes necessary when analyzing extensive amounts of underwater acoustic data. Our contributions contain supervised and unsupervised models, and pipelines for underwater soundscape analysis and sound event clustering. State-of-the-art Machine Learning techniques can be used to understand, explore, predict, and quantify underwater soundscapes and their individual sound components, tasks that would be challenging to accomplish manually.

Publications

International Journals

(publications in journals listed in the ISI Web of Science)

As first author

[CP1] **Parcerisas, C.**, Roca, I.T., Botteldooren, D., Devos, P. & Debusschere, E. (2023). Categorizing Shallow Marine Soundscapes Using Explained Clusters. *J. Mar. Sci. Eng.*, 11, 550. https://doi.org/10.3390/jmse11030550
(Impact Factor: 2.9, Q1, Engineering, Marine)

[CP2] **Parcerisas, C.**. Schall, E., Te Velde, K., Boteldooren, D., Devos, P. & Debusschere, E. (2024). Machine learning for efficient segregation and labeling of potential biological sounds in long-term underwater recordings. *Front. Remote Sens.* 5:1390687. https://doi.org/10.3389/frsen.2024.1390687

[CP3] Schmidlin, S., & **Parcerisas, C.**., Hubert, J., Watson, M.S., Mees, J., Botteldooren, D., Devos, P., Debusschere, E., & Hablützel, P.I. (2024). Comparison of the effects of reef and anthropogenic soundscapes on oyster larvae settlement. Scientific Reports, 14(1), 12580. https://doi.org/10.1038/s41598-024-63322-2
(Impact Factor: 4.6, Q1, Multidisciplinary Sciences)

As co-author

[CP4] Schall, E., & **Parcerisas, C.** (2022). A Robust Method to Automatically Detect Fin Whale Acoustic Presence in Large and Diverse Passive Acoustic Datasets. *J. Mar. Sci. Eng.* 10(12), Article 12, https://doi.org/10.3390/jmse10121831.

[CP5] Rubbens, P., Brodie, S., Cordier, T., Destro Barcellos, D., Devos, P., Fernandes-Salvador, J. A., Fincham, J., Gomes, A., Handegard, N. O., Howell, K., Jamet, C., Kartveit, K. H., Moustahfid, H., **Parcerisas, C.**, Politikos, D., Sauzède, R., Sokolova, M., Uusitalo, L., Van den Bulcke, L., van Helmond, A. T. M., Watson, Jordan T., Welch, H., Beltran-Perez, O., Chaffron, S., Greenberg, D.S., Kühn, B., Kiko, R., Lo, M., Lopes, R.

M., Möller, K. O., Michaels, W., Pala, A., Romagnan, J. B., Schuchert, P., Seydi, V., Villasante, S., Malde & K, Irisson, J.O. (2023). Machine learning in marine ecology: an overview of techniques and applications. *ICES Journal of Marine Science*, Volume 80, Issue 7, Pages 1829–1853, https://doi.org/10.1093/icesjms/fsad100.

[CP6] Schall, E., Kaya, I.I., D., Debusschere, E., Devos, P., & **Parcerisas, C.** (2024). Deep learning in marine bioacoustics: A benchmark for baleen whale detection. *Remote Sensing in Ecology and Conservation* https://doi.org/10.1002/rse2.392.

[CP7] Calonge, A., **Parcerisas, C.**, Schall, E. & Debusschere, E. (2024) Revised clusters of annotated unknown sounds in the Belgian part of the North Sea. Frontiers in Remote Sensing. https://doi.org/10.3389/frsen.2024.1384562

[CP8] Cosentino, M., **Parcerisas, C.**, Brundiers, K., Sarnocinska-Kot, J., Botteldooren, D., Devos, P., Dähne, M., Tougaard, J. & Wahlberg, M., Debusschere, E. (2024). Passive acoustic monitoring of odontocetes: the devil is in the details. *In prep.*

International conferences

As first author

[CP9] **Parcerisas, C.**, Botteldooren, D., Devos, P., Hamard, Q., Debusschere, E. (2023). Studying the Soundscape of Shallow and Heavy Used Marine Areas: Belgian Part of the North Sea. In: Popper, A.N., Sisneros, J., Hawkins, A.D., Thomsen, F. (eds) The Effects of Noise on Aquatic Life. Springer, Cham. https://doi.org/10.1007/978-3-031-10417-6_122-1

[CP10] **Parcerisas, C.**, lgaz-Kaya, İ., Botteldooren, D., Devos, P., Debusschere & E., Schall, E.. Using CNN classifiers as underwater sound source detectors: learning about noise. *OCEANOISE*, May 2023, Vilanova i la Geltru, Spain.

[CP11] **Parcerisas, C.**, Roca, I.T., Botteldooren, D., Devos, P., & Debusschere, E. Clustering, categorization, and mapping of shallow coastal water soundscapes. *Forum Acusticum*, Sept 2023, Turin, Italy. https://www.doi.org/10.61782/fa.2023.1070

[CP12] **Parcerisas, C.**, Kaya, I.I., Debusschere, E., Devos, P., Schall, E. Tuning an object detector on a benchmark for the detection of Antarctic baleen whale calls. *DCLDE*, June 2024, Rotterdam, The Netherlands. *Co-chair in session "Baleen Whale Detection and Classification"*

[CP13] **Parcerisas, C.**, Devos, P., Botteldooren, D., Debusschere, E. Detecting and clustering unknown sound events using transfer learning for marine soundscape analysis. *ICUA*, June 2024, Bath, UK.

Datasets

[CP14] **Parcerisas, C.**, Schall, E., Calonge, A. & Debusschere, E., Flanders Marine Institute; Alfred Wegener Institute; (2024): Annotated unknown underwater sounds in the Belgian part of the North Sea. Marine Data Archive. https://doi.org/10.14284/659

[CP15] **Parcerisas C.**, Schall, E., Aubach, J., Te Velde, K., Debusschere, E.: Flanders Marine Institute; Alfred Wegener Institute; (2024): Acoustic salient event annotations. Marine Data Archive. https://doi.org/10.14284/667

[CP16] **Parcerisas C.**, Debusschere E., Flanders Marine Institute (VLIZ): Belgium; (2022): Acoustic data from drifts in the Belgian Part of the North Sea 2020-2021. Marine Data Archive. https://doi.org/10.14284/586

[CP17] **Parcerisas C.**, Botteldooren, D., Devos, P. & Debusschere, E., Flanders Marine Institute (VLIZ); 2021; Broadband Acoustic Network dataset. https://marineinfo.org/id/dataset/7879

Software

[CP18] **Parcerisas, C.** (2023). lifewatch/bpnsdata: v0.1.10 (v0.1.10). Zenodo. https://doi.org/10.5281/zenodo.7620700

[CP19] **Parcerisas, C.** (2023). PyPAM: a package to process long-term underwater acoustics data in chunks (0.3.0). Zenodo. https://doi.org/10.5281/zenodo.10037826

[CP20] **Parcerisas, C.** (2023). lifewatch/pyhydrophone: A python package to ease the calibration and metadata handling for hydrophones (v0.1.3). Zenodo. https://doi.org/10.5281/zenodo.7588428

[CP21] **Parcerisas, C.** (2022). lifewatch/pyporcc: Version 0.1.3 (0.1.3). Zenodo. https://doi.org/10.5281/zenodo.7310801

Other

[CP22] Rey Baquero, M. P. & **Parcerisas, C.**, Seger, K., Perazio, C., Botero Acosta, N., Mesa, F., Acosta, A. L., Botteldooren, D., & Debusschere, E. (2021). Comparison of Two Soundscapes: An Opportunity to Assess the Dominance of Biophony Versus Anthropophony. Oceanography, 62–65.

[CP23] Boebel, O., Allerholt, J., Engicht, C., Hoppema, M., Llanillo, P., **Parcerisas, C.**, Pinner, O., Roca, I.T., Spiesecke, S., Tippenhauer, S. (2023). 2.1. Physical Oceanography, in: Hoppema, M. The Expedition PS129 of the Research Vessel POLARSTERN to the Weddell Sea in 2022. Berichte zur

Polar- und Meeresforschung = Reports on Polar and Marine Research, 776: pp. 18-52

[CP24] Spiesecke, S., **Parcerisas, C.**, Roca, I.T., Boebel, O., Burkhardt, E., Thomisch, K., van Opzeeland, I. (2023). 2.2. Ocean Acoustics, in: Hoppema, M. The Expedition PS129 of the Research Vessel POLARSTERN to the Weddell Sea in 2022. Berichte zur Polar- und Meeresforschung = Reports on Polar and Marine Research, 776: pp. 53-98

1 Introduction

1.1 Context and Motivation

Passive acoustic monitoring (PAM) refers to the monitoring of environments by recording and analyzing their sounds, and it has become widely used in the past decades. It is a field of special importance in marine environments, as many marine species rely on sound as a means to obtain information from their surroundings. For this reason, the study of underwater sound can provide valuable ecological information about the marine world. PAM data can be obtained from boats, autonomous devices – either fixed or moving ones –, cabled stations, and acoustic tags (with incorporated hydrophones and recorders), enabling data collection in a variety of situations that would otherwise be unreachable [1]. This makes PAM a flexible, non-invasive, and cost-effective solution to acquire ecological information and disturbing anthropogenic pressures from remote areas or busy shallow coastal areas. However, because of the inaccessibility of most marine locations and the consequent elevated costs of deployment and retrieval of instruments, data are often collected for long periods. This together with the high sampling rate required to study (most) biophony quickly generates large datasets that require automation to extract information because analyzing them manually is very time- and resource-consuming [2].

When recording any environment using a broadband acoustic recorder, we will always obtain recordings of all the sounds occurring at that moment (with limitations from the recorder's capabilities). When speaking about this complex polyphony of sounds occurring at a certain habitat, we refer to it as soundscape [3, 4, 5]. In ecological applications, the different sounds present in a soundscape are subdivided into three categories: biophony (biotic sound events); geophony (abiotic sound events); and anthrophony (man-made sounds) [6]. In a marine environment, biological sound sources include sounds actively produced by marine animals to communicate, navigate, mate, and forage: these signals can then be used to detect species occurrence, behavior, and conspecific interactions [2]. In addition to these actively produced sounds, some animals also produce sound as a byproduct of their behavior, for example, the sound of moving or feeding se urchins [7]. Abiotic sounds are mainly produced by physical events, such as rain, waves, wind, sediment transport, earthquakes, or currents [8]. A soundscape can also comprise sounds originating from human activities, such as shipping, sonar, pile driving, seismic exploration, or sand and oil extraction, increasingly present during recent decades [9]. Furthermore, a soundscape is also influenced by passive elements, such as the vegetation and landform of a particular location because of its interaction with the current and its influence on propagation patterns [10, 11]. Propagation patterns depend on sediment type and shape of the seafloor, and the presence of sound absorbers, scatterers, and reflectors (e.g., aquatic fauna, bubble clouds, or suspended sediment). In addition to local acoustic events, distant sounds also contribute to the local soundscape. The propagation of these distant sound sources can ultimately be shaped by depth, topography, salinity, and temperature, among others. All these elements influence how these sounds are received.

Even though the use of the term ”soundscape” has exponentially grown in literature during the last decades, there are multiple understandings currently in use of what soundscape analysis means. It is thus necessary to specify what we are referring to when speaking about soundscape analysis and characterization. For clarification, we classify the methods for soundscape analysis into three different groups:

(1) Methods analyzing soundscapes holistically by incorporating multiple elements. Some of these methods are used in terrestrial acoustics to describe a human-focused perceptual experience, and try to analyze how a soundscape is perceived and understood [12]. Adding a human perceptual component in the analysis does not apply to underwater soundscapes because humans do not spend large amounts of their lifetime underwater, and thus have little information about the underwater acoustic experience. Therefore, when applied to ecology, this collection of methods can be more similar to acoustic environment analysis. In this context of holistic characterization of soundscapes, ecoacoustics is defined as a “theoretical and applied discipline that studies sound along a broad range of spatial and temporal scales in order to tackle biodiversity and other ecological questions. The use of sound as a material from which to infer ecological information enables ecoacoustics to investigate the ecology of populations, communities and landscapes”

[13, 14, 15, 16]. Ecoacoustics is not to be confused with bioacoustics. Ecoacoustics and bioacoustics are related but distinct fields. Ecoacoustics views sound as a key component and indicator of ecological processes, studying soundscapes to understand ecosystem health and dynamics. Bioacoustics, on the other hand, focuses on sound in animal behavior, examining how animals use sound to communicate information between individuals [17, 13]. Within this holistic approach, we include the definition and classification of “acoustic scenes”, which refers to the recognition of an environment from recorded acoustic signals. It is a well-described research area for human environments but can also be applied to ecological monitoring and understanding [18, 19]. We will refer to this collection of methods as Holistic Soundscape Characterization.

(2) Methods isolating acoustic events, namely individually identifiable sounds. The definition of an acoustic event depends on the temporal scale one is interested in. For example, sounds lasting several hours can be identified as background when one is analyzing several seconds, but can be identified as salient events when one analyzes several days. When isolating acoustics events, the soundscape is then analyzed by reporting the proportion and spatio-temporal distribution of these events [20, 21]. We will refer to this method as Acoustic Event Detection. In marine bioacoustics, one of the biggest challenges is the sparsity of the occurring sounds of interest, and the variability of all the other sounds classified as ”noise” [22]. This, together with the lack of public sound libraries with annotated sets and reference clips, makes accurately detecting and classifying biological underwater sounds challenging.

(3) The study of models predicting the acoustic characteristics of a certain environment, often focused on sound propagation models [23, 24]. We will refer to this method as mapping and predicting. This is often done by modeling sound sources and wave propagation in large marine areas to get a general idea of the areas most affected by certain anthropogenic sound sources.

All these soundscape analysis techniques offer the possibility of surveying entire underwater habitats and their acoustic environments simultaneously, from different perspectives, which is critical to understanding their fundamental interaction and long-term dynamics [25, 26]. Different approaches can be complementary, as they analyze the same habitat from different perspectives, bringing new insights for interpretation and understanding.

Despite the increased interest in soundscape analysis, characterizing marine soundscapes remains a challenging task. There are currently no standards by which to do so, as mentioned in the International Quiet Ocean Experiment (IQOE) report published by the Marine Bioacoustical Standardization Working Group [27]. Regarding the data acquisition, there are still no standards on how to record underwater sound to perform soundscape analysis, and methodologies from study to study present a great variability in mooring systems, recorder settings and specifications, and sampling settings such as duty cycle, among others. In the Marine Strategy Framework Directive (MSFD) [28], it is stated that they note "that there are no international standards for the measurement, modelling or storage of data related to underwater ambient noise, and recommends that such international standards be developed, including the measurement of radiated sound from sources such as airgun arrays and underwater explosions (standards for the measurement of sound radiated from ships and pile driving equipment are already being developed by ISO)". The working group proposed a monitoring guidance document to support the implementation of Descriptor 11 [29]. Other initiatives have also proposed their own guidelines, such as the one proposed by the IQOE Essential Ocean Variable (EOV) working group [30], the IQOE standarization group [27], or the recommendations proposed by the JOMOPANS project [31]. Even though the monitoring programs are shifting towards a unification, these guidelines are still not fully adopted by the broader community. Regarding the data processing, a soundscape analysis can comprise one or multiple steps and approaches such as (but not limited to) listening to recordings, visual inspection of spectrograms, automated detection of signals of interest, computation of several acoustic metrics, or statistical analysis [25]. In this process, understanding the natural variability of a soundscape is important for using soundscapes as an ecological and management tool to determine the potential effects of human activity [32] and to increase our understanding of ecosystems.

So far soundscape characterization has mostly been conducted locally, not centering on a global application of methods but rather on particular ecosystems and environmental conditions. In general, there is a bias in literature towards describing natural or pristine areas, areas with high rates of endemism, or areas of concern [26, 33]. Furthermore, most of the described soundscapes are from temperate coastal zones and tropical coral reef habitats, even though other areas from sub-tropical or polar environments have also been described [33].

When speaking about sound events, there is a big bias towards studying the sounds produced by marine mammals, even though recently there has been raised interest in fish [34] and invertebrates [35] acoustics. Furthermore, most of the studies are based on the fact that we know which species makes which sound. This is not a trivial statement, as associating sounds with their sources can be challenging at times. This is especially the case when relying on visual surveys underwater, which are often impractical or severely limited due to poor visibility caused by high turbidity [36], light condition, depth, or general accessibility. Recording animals in captivity provides a straightforward ground truthing approach, but many animals may present different behaviors in the wild than in captivity due to different environmental conditions [34]. When recording in the wild it is necessary to localize the source of a sound to be able to unambiguously attribute it to an origin. This has been addressed for sessile animals by placing sensors on them while recording in the wild [37]. Recently, Mouy et al. (2023) [38] have proposed a promising system to localize sound producers with an array of hydrophones and video cameras, but the effort required for these deployments remains high and is not applicable in high turbidity areas.

1.1.1 Shallow waters and the Belgian Part of the North Sea

PAM presents certain specific challenges when applied to shallow waters. To begin with, certain technological challenges in recording underwater sound are particularly prevalent in shallow waters. First, shallow environments often present high currents. The current around the hydrophones generates flow-noise when recording in a static position. Flow-noise is the turbulence around the hydrophone captured by the hydrophone, corrupting the data. The recorded signal is not sound but pseudo-sound, as these pressure fluctuations are not propagating acoustic waves. There is currently no agreement on how to collect and process acoustic data with a high flow-noise rate [39, 40, 41]. Some existing options are to discard frequencies highly correlated with current [42], only consider slack tide periods when reporting underwater sound [43], measure with an array of hydrophones to be able to remove the non-correlating part of the signal [44], or use protection cases or foam around the hydrophone to reduce the pressure fluctuations [40]. Current also can generate sediment transport, which also generates pseudo-noise at high frequencies when hitting the acoustic recorder. Second, because of the high productivity of coastal shelf systems [45], high rates of colonization occur on artificial substrates [46] such as the recording equipment, limiting the recorder's capabilities.

Moreover, shallow environments are quite particular when studied acoustically [47]. Lower frequencies have a wavelength greater than the water column, which creates lower cut-off frequencies. In shallow waters, seabed sediment type plays an important role in propagation [48], and sound generated at the surface is received at the bottom, and vice versa. Moreover, sound waves bounce between the water surface and seabed, creating multiple paths that complicate signal reception due to interference, fading, and transmission losses [49]. This contributes to creating a complex soundscape, which can vary over small spatial scales (some meters [50]). Other factors affecting the high variation of sound at a small scale are the diversity in the occurrence and proximity of human activities, sound propagation conditions, and localized biological activity. This variation in received levels is not only happening horizontally but also throughout the water column, and therefore the received sound depends on the hydrophone depth.

Considering all the above, shallow waters are different than open waters when studied acoustically. For example, the Joint Research Center of the European Commission proposed root mean squared (RMS) sound levels in 1/3 octave bands centered at 63 and 125 Hz as indicators of shipping noise for the MSFD [28]. In very shallow areas these frequencies might be above the theoretical cut-off frequency, so they do not propagate. Focusing exclusively on these frequency bands might then result in an inaccurate estimation of shipping noise [51], as low sound levels within these bands may not necessarily correspond to low levels in higher frequency bands. Likewise, because the hydrophone's position on the water column influences the received sound levels, it is complicated to estimate the overall noise pollution levels without using complex sound propagation models.

To create sound maps or predict noise pollution, knowledge of the present sound sources is necessary. In this context, some human-made sound signatures have been well studied in the last decades in open water [52], and so have some of the geophonic sound sources [53, 54]. However, the received spectrum levels of these sound sources are very different in shallow water than in open water [55]. Only recently some studies have focused on characterizing shallow water shipping sound [56]. In these types of studies, it is common to use Automatic Identification System (AIS) data to create shipping noise maps [57]. However, most of these noise maps only consider the sound produced by the vessels' underway engine. Other maritime operations common in coastal regions, like trawling and dredging, produce distinct sound patterns different from those of vessels underway. Consequently, their radiated sound levels present a different frequency distribution. This should be taken into account when creating and evaluating human-made noise maps.

In conclusion, very little research has been done on the soundscape of an area such as the Belgian Part of the North Sea (BPNS), which is a small Exclusive Economic Zone (EEZ) located in the southern part of the North Sea. It is a coastal and highly exploited area with a unique sandbank system. This environment entails strong tidal currents [58], very shallow bathymetry (maximum depth of 45 m), and high turbidity [59] and sediment transport. Furthermore, as a result of the First and Second World Wars and the rough maritime conditions, the BPNS contains more than 300 wrecks [60] which have grown into artificial hard substrates, becoming a *hotspot* for biodiversity. This unique sandbank system together with the shipwrecks and some gravel bed areas are home to a high invertebrate biodiversity, creating several benthic communities. At the same time, in an area as small as the BPNS a lot of human activities converge. The BPNS contains shipping routes to the major European ports and offshore wind farms. Activities such as dredging, sand and gravel extraction, military exercises, fishing, and trawling, among others, cover its entire area.

So far, the few studies carried out on the BPNS concerning underwater sound have focused solely on the quantification of underwater sound levels [61, 62]. Therefore, there is a lack of reference data and understanding regarding the biological sound sources – except for a few mammals – within the BPNS. Additionally, due to the high turbidity, deploying cameras to track recorded sounds to their respective producing species is not feasible. Therefore, with the current knowledge of biological sounds from the BPNS it is not possible to analyze soundscapes by detecting and quantifying known sound events. Nevertheless, soundscape analysis can still be performed even if the sound events are not known, as sounds unidentified can still provide meaningful acoustic information that can be used to characterize soundscape components [63, 64]. This is however more challenging than the traditional detection and quantification of known events, as it is harder to find sound events in long-term recordings when one does not know which events to expect. Furthermore, the multiple human activities occurring in the BPNS generate significant levels of sound [61, 65]. These sounds have the potential to mask quieter biological sounds occurring simultaneously within the same frequency band. This complicates the detection and characterization of (unidentified) biological sounds.

Coastal, highly exploited marine areas such as the BPNS have important economic value but are threatened by overexploitation of resources. With the blue economy growing to unprecedented numbers and the strategic plan of the EU for offshore development for renewable energy [66], there is concern about the growing human pressure applied to the ocean and our coast [67, 68, 69]. Consequently, it is necessary to understand the ecosystem and its limits to protect it so that use can continue sustainably. Among other pressures, it is critical to evaluate the effect of anthropogenic sound on marine fauna. These effects include physical and behavioral effects but also masking of sounds, both for sounds produced for communication and sounds used as cues to obtain information crucial for survival [9]. However, to quantify the pressure that anthropogenic sound exerts on marine fauna and estimate its effects, it is first necessary to understand the sound sources in the soundscape and determine the contributions of human-made sound. Soundscapes holistic characterization might be more important for general assessment and long-term analysis, while Acoustic Event Detection might be more effective in identifying whether communication is disturbed by anthropogenic sounds and biodiversity changes. For this reason, it is necessary to develop new techniques to analyze soundscapes focused on shallow water highly-used marine areas such as the BPNS. There has been some focus on developing more integrated approaches to characterizing soundscapes, for example, adding contextual information to understand possible sound sources [32, 70], or performing unsupervised source separation and quantification [71] but there is still a knowledge gap.

1.2 Machine learning in PAM

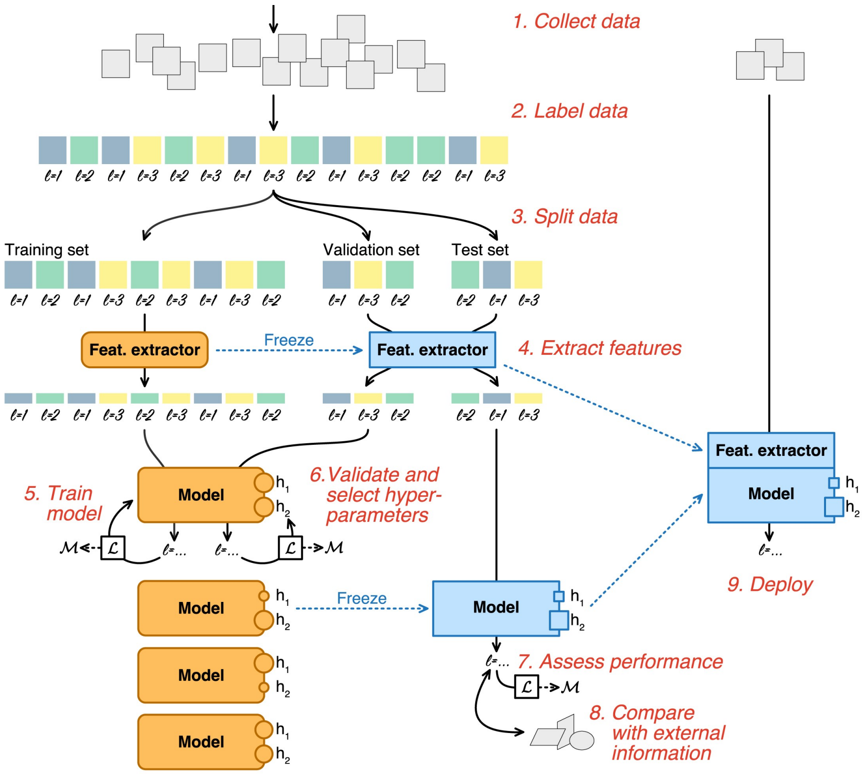
As a result of the mentioned vast volumes (hundreds of TB) of acquired acoustic data, and because analyzing acoustic data manually is labor-intensive by nature, there has been growing interest in addressing soundscape analysis using automated tools [2]. Before the popularization of ML, algorithms based on predefined rules (rule-based) were used to automatically detect sound events in long-term recordings

[72, 73, 74]. Although ML methods were introduced to analyze marine PAM data as early as the 90s, their application has significantly increased in the recent decades [75].

To date, the most common approach to processing marine acoustic data for soundscape analysis using ML is to perform Acoustic Event Detection and Classification in a supervised manner. The terms *classification* and *detection* are sometimes used interchangeably in literature [22]. Throughout this dissertation we refer to classification to the problem of classifying short snippets of sounds, already preselected manually or by another algorithm. We refer to detection as the process of extracting these snippets from long-term recordings. Detection and classification can be performed in one step (joint detection and classification), or as two steps, where detection is applied to extract possible snippets which are afterward fed into a classification model.

When developing supervised machine learning models, several steps are necessary (see Figure 1.1). First data need to be collected. For supervised machine learning, these data need to be manually annotated. A very important step for evaluating ML models is to split these data into training, validation, and test data sets. The training data are used to train the model. The validation data are used to validate the model during training successively to select hyperparameters. The test data are used to provide an unbiased evaluation of a final model fit on the training data set. This assessment is critical to get an idea of how this model will perform on new data. The selection of both the test and the validation sets is critical when assessing ML models applied to PAM [76], as depending on what is selected, it might not be a representative set for real-case scenarios. If a model is to be used in new locations, testing it on one or several locations never used for training is good practice to assess the generalization capacity of the model.

Figure 1.1 The general process of (supervised) machine learning. After being collected (1), data need to be annotated/labeled (2). The data are then split into training, validation, and test datasets (3). Each input in the training set can be summarized into features (4). The (transformed) training set is used to train the model (5), by minimizing a loss function (L) that computes the value of one or several performance metrics (M). The validation set undergoes the same transformation as the training set, if any, and is then used to evaluate the predictive performance of the model, ideally with the same metric(s) (6). Several versions of the model can be trained with different hyperparameters (i.e. settings, noted h*) of the machine learning system, and the one with the best performance on the validation set is retained. At this point, the model is frozen and its final performance is assessed on the test set (7). If external information, different from the original data, is available, it should be used to ensure that model predictions are reasonable, in addition to achieving a given performance (8). Finally, the model is ready to be deployed and used with newly collected data (9). Extracted from Rubbens et. al (2023) [75]

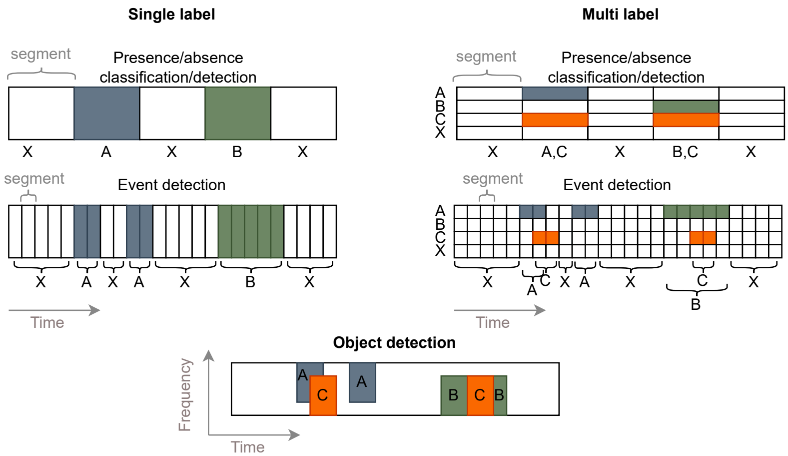


When a long-term recording is segmented into equally sized windows, a classification algorithm can be applied to detect the presence or absence of a certain sound in that time window, but can't provide the exact start and end time of that event. We will refer to this as ”classification applied to detection”. On the other hand, if these windows are a lot shorter than the sound of interest, sequences of positive detections can be merged into predicted event regions, providing the start and end of an event [22]. When using a “classifier” as a “detector” with the windowing approach in underwater bioacoustics data in long-term data, the imbalance between noise and calls of interest has to be taken into account. A common practice in ML to deal with these datasets with highly imbalanced classes is to sub-sample the most common class to artificially balance the datasets. However, in Schall et. al (2024) [76] it is shown that for underwater bioacoustics, if the classifier is to be used in a long-term recording with a high imbalance in noise (sparse-occurring calls), training and testing it on an artificially balanced dataset will most likely not provide good performance on a real-world dataset. This could be due to the fact that this artificially selected “noise” examples might not be representative enough of the “noise” class [76]. The detection or classification can be binary, where only one specific call or species is tackled, or multiclass, where different species or sound types are identified. The output can also be a single label per segment of multiple labels. A single label model does not allow for more than one class to be detected simultaneously on the same segment, while multiple label models do. The segmentation approach can be framed into different ML problems, depending on the research question at hand and the available data. A graphical summary of all the possibilities is presented in Figure 1.2, representing the 3 main concepts:

- Single-label (binary or multi-class): the model assigns one class from a subset of classes per segment. If the subset is only wo classes, then it is a binary approach.
- Multi-label: the model assigns one or several classes from a subset per segment
- Object detection: the model assigns multiple regions per segment

The detected sound events can be anthropogenic, from physical phenomena or biological. Nonetheless, the effort on source classification has been put mainly on the study of mammals' vocalizations (as examples: [77, 78, 79, 80, 81]) and shipping sounds [82, 83]. More recently some classifiers for fish species have been developed [84]. Detection and classification algorithms have been used to identify species [85], specific calls [86] or even dialects and individuals [87, 88]. In some cases, the target is not the classification of a certain species but the quantification of similarities and differences between acoustic signals. This quantification can then be used to understand differences between calls, for example in Deecke et al. (1999) the classification algorithm was used to quantify and understand differences between whale dialects [89].

Figure 1.2 Summary of the different approaches to segmenting continuous recordings and different possible model types to apply to the segments data for Acoustic Event Detection and Classification. Left: single label model approaches. Right: multi-label model approaches. Bottom: object detection models. The single label approaches can be stated as binary classification problems if only one class is used. x-axis represents time, and y-axis represents frequency for the object detection example. Figure inspired by Stowell (2022). [22].



When using ML for underwater Acoustic Event Detection and Classification, a common approach is to extract hand-picked acoustic features from the sound and use them as input for an ML algorithm. These features can be derived from the time, frequency, or cepstral domain (transformation of the data to highlight periodic signals), or based on the full image of the spectrogram, a visual representation of sound intensity per frequency as a function of time [90]. ML applied to PAM benefits from the advances in both image and speech-recognition algorithms, as sound can also be treated as an image after being converted to a spectrogram. In DL approaches, using spectrograms as input is the most common approach, as they can be used as input for convolutional neural networks (CNN) [91]. However, recently some models have been developed that are applied directly on the waveform [92, 93]. In general, ML approaches have improved performance and generalization [94] compared to previous rule-based models. DL models reduce necessary human input when choosing acoustic features, and these features can be more tailored to the task. This is of special importance for sounds which have not yet been described, as one would have to test first which of the traditional features discriminate better these sounds.

Some recent publications have focused on developing models transferable across different bioacoustics datasets (both terrestrial and marine), such as Hagiwara (2023) [93], Best et al. (2023) [95] or Robinson et al. (2024) [96]. However, the developed models perform significantly worse on marine datasets, while presenting relatively high performance on terrestrial datasets. This can occur because some of these models are pre-trained on large datasets containing only terrestrial recordings, such as AudioSet [97]. It might also be because the amount of annotated data from marine environments is significantly less than that from terrestrial ones. For marine datasets it is harder to increase the annotated data volume using citizen science because of the lack of knowledge in the common population about underwater sounds - opposite to bird songs. Another reason is that in the marine context, sounds of interest can be very sparsely occurring and datasets can comprise long periods of time where the signals of interest are not present. This leads to highly imbalanced datasets, which has a great influence on performance [76]. This final challenge is occasionally addressed by initially using a binary output for detection, followed by the classification of the identified snippets [22] (two-step detection and classification). When using the two-step approach of detection and classification, often the detection step is a rule-based signal-processing algorithm and the classification step is a DL approach [98, 99]. The two steps combined are commonly referred to as automatic Detection and Classification Systems (DCS), and the performance is usually computed considering the entire system and not only the classifier. The most common algorithms used for the classification step are and several types of Neural Networks (NN) [76], but other algorithms such as Support Vector Machines (SVM) [100], Random Forests (RF) [101], Gaussian Mixture Model (GMM) [102] k-means [87], or a combination of them, are also used.

There seems to be a greater interest in finding which acoustic features are representative and therefore sufficient for the detection, classification, and characterization of sound events than which algorithm performs the best. Often more than one classification algorithm is tested using different features or pre-processing to check which features are more robust (for birds: [103]). The difficulty in obtaining real-time data is a current limitation to the full exploitation of marine PAM data using autonomous recording platforms for long periods, as data are generally be accessed once the recorders are recovered. The primary issue is due to the large amounts of data produced, which are too extensive to transmit without a physical connection. However, if a set of features proves to be sufficiently informative, it could be processed directly on the device and transmitted in place of the raw waveform data [104]. This approach requires the recording systems to have enough onboard computing capacity for embedded processing. The resulting data (either features or detections) could then be transmitted via the internet (through satellite, 4G, 5G, etc.), enabling real-time monitoring from otherwise inaccessible locations [105].

Except for some marine mammals, one of the biggest limitations of machine learning applied to PAM for marine Acoustic Event Detection is the limited amount of available annotated data, which makes supervised machine learning a challenging task. Some recent international initiatives address this gap, such as FishSounds database [106] or GLUBS [107], which are working on building public online databases with (labeled) underwater sounds. Some other research is focused on helping to speed up this annotating process [108, 109, 110]. Furthermore, to solve this problem there has been some effort made for few-shot learning in bioacoustics, such as the challenge proposed by DCASE 2023 Task 5, where one marine dataset is included in the evaluation set. In few-shot learning tasks, the algorithm must make predictions after being given only a few instances of each class [111].

Even though ML in PAM has been most widely used for Acoustic Event Detection, there has also been advancement in other applications. Another common application is to localize sound sources underwater [112, 113]. This can be a goal on its own or can be used for ground truthing sound sources if the sound is recorded together with video images [38]. From array recordings a model can be trained to localize the position of a certain source without the need to model the sound propagation, outperforming conventional matched field processing methods [114].

Another usage of ML applied to soundscape characterization has been to create regression models to find a relationship between specific characteristics of the source and the sound itself. Among these characteristics, some examples are the size of male sperm whales [115], wind speed [54] or fish abundance [116].

In addition, ML has also been used for acoustic source separation. This has been relatively understudied compared to the detection of events, but it has shown promising results both for individual signal isolation [117, 118, 119, 120, 121] and as an enhancement tool to compute ecoacoustic metrics [122, 123, 71, 124]. Source separation of simultaneous same-type vocalizations cannot be solved by traditional signal processing techniques, and only recently a model has been developed capable of disentangling simultaneous dolphin whistles [125].

Clustering is also commonly applied as a tool to analyze underwater sounds which does not rely on ground-truth data. Some approaches characterize entire marine habitats from their soundscape using clustering [126, 70, 124, 127, 123, 128], enhancing the understanding on the long-term dynamics of ecosystems. Clustering can also be applied to already selected sounds and can be used to unravel similarities, differences, and latent groups between calls [95, 129]. The latest published clustering applications apply a feature reduction using Uniform Manifold Approximation and Projection for Dimension Reduction (UMAP) [130] to the original features before applying a clustering algorithm. This pipeline seems to result in better clusters and improve the agreement between ground-truth labels and obtained clusters, and it provides a framework for manually exploring and visualizing the data. A common algorithm used for clustering is HDBSCAN [131], as it can also discard samples as ”noise” when they are not included in a dense cluster.

In the last years – during the course of this dissertation – different software has been released to ease the creation of ML models for bioacoustics detection. Some of them come with already pre-trained models for certain species and some of them with an interface where the users can train the model with their own dataset. Recently, a database with available software tailored to bioacoustics analysis has been created [132]. It can be found at `https://rhine3.github.io/bioacoustics-software`. This list includes several packages that can include - or can assist in creating - models to detect, classify, or localize acoustic events in bioacoustic data.

1.3 Objectives and research questions

Natural sound environments are disappearing at an unprecedented pace due to human influence [133, 134]. For this reason, understanding and preserving aquatic soundscapes is a goal in itself, as a part of preserving sonic heritage of millions of years of evolution [135, 136, 137]. Documenting these natural soundscapes before they disappear can also be used as a tool to increase chances of acoustic habitat restoration and to raise awareness for citizens about the importance of biodiversity. Characterizing soundscapes can provide ecologically meaningful information, which can be used for multiple reasons. This information can comprise animal distribution, abundance, and behavior; species diversity; and changes of all of these over time. Among others, it can be a useful tool to monitor biodiversity, ecosystem health, assess ecological impact, preserve habitats, and speed up habitat restoration [25]. Once information is obtained and decoded, it can be used to inform conservation management and to assess the effectiveness of management and conservation efforts.

Obtaining this information is however not a straightforward task. Consequently, this dissertation aims to explore how we can monitor and characterize the underwater soundscape components in the BPNS. In this particular area, all the following characteristics are present: shallow water, high turbidity, high flow-noise, high human activity, and unknown sound sources. Because of these particular challenges, it is necessary to investigate new methodologies. Some other challenges common to underwater bioacoustics but not particular to this area are also considered. These include the need for reliable automatic detectors when there is a high sparsity of sound events of interest and the lack of big annotated datasets.

Considering the mentioned challenges and focusing on the ultimate goals, we explore the following research questions:

1. Can we holistically describe and characterize the soundscape of the BPNS, and identify the external environmental factors (such as wind, rain, and currents) that influence it?

2. Is it possible to disentangle previously undocumented biological sound events from long-term recordings, and can machine learning assist in detecting and classifying such events?

3. Are there examples where the North Sea soundscape plays a crucial role in the development of local fauna? Specifically, does the natural soundscape from the North Sea impact the settlement behavior of oyster larvae *Magallana gigas*?

To answer these questions we propose new methodologies and tools for soundscape analysis. To achieve this goal, we propose the use of different state-of-the-art ML techniques. However, to build these methods some annotation efforts are necessary, but the annotation process is effort-intensive. Therefore, the proposed methods focus on having as little and efficient human input as possible. First, in Chapter 2 we list the datasets acquired during the course of this PhD. In Chapter 3 we describe the technological challenges of recording and analyzing PAM data in shallow and heavily exploited areas, and we list all the known sound-making species present in the BPNS. Part II focuses on a holistic approach, where soundscapes are not characterized using specific events but with a general approach. Particularly, in Chapter 4 we analyze the soundscape in the BPNS using traditional methods that focus on statistical analysis to identify the primary components. InChapter 5 we propose to categorize soundscapes in an unsupervised way by grouping the different acoustically similar soundscapes. The obtained categories are then correlated to environmental parameters using explainable ML (SHAP), in a way that we can use to understand the spatiotemporal patterns at which the different categories occur. Part III focuses on Acoustic Event Detection, and we first showcase how object detection models can be used to detect sound events in Chapter 6. Then in Chapters 7 and 8, we focus on unknown sound sources, showing how clustering can be used to understand sound types and proposing a model to detect any sound event which could possibly be of interest. In Chapter 9 we present a real application of the importance of understanding and analyzing soundscapes by studying how oyster larvae react when exposed to different real-world soundscapes. Finally, in Chapter 10 we present a general discussion on the work done. We highlight the importance of our work for the advancement of the research field, we contextualize the limitations of the presented strategies, and we list the remaining challenges and future work.

References

[1] K. A. Kowarski and H. Moors-Murphy. *A Review of Big Data Analysis Methods for Baleen Whale Passive Acoustic Monitoring*. Marine Mammal Science, n/a(n/a), 2020.

[2] R. Gibb, E. Browning, P. Glover-Kapfer, and K. E. Jones. *Emerging Opportunities and Challenges for Passive Acoustics in Ecological Assessment and Monitoring*. Methods in Ecology and Evolution, 10(2):169–185, 2019.

[3] B. Krause and A. Farina. *Using Ecoacoustic Methods to Survey the Impacts of Climate Change on Biodiversity*. Biological Conservation, 195:245–254, March 2016.

[4] B. C. Pijanowski, L. J. Villanueva-Rivera, S. L. Dumyahn, A. Farina, B. L. Krause, B. M. Napoletano, S. H. Gage, and N. Pieretti. *Soundscape Ecology: The Science of Sound in the Landscape*. BioScience, 61(3):203–216, March 2011.

[5] International Organization for Standardization. *ISO 18405:2017*. Technical report, International Organization for Standardization, 2017.

[6] B. Krause. *Bioacoustics: Habitat Ambience in Ecological Balance*. Whole Earth Review, No 57, December 1987.

[7] L. Coquereau, J. Grall, L. Chauvaud, C. Gervaise, J. Clavier, A. Jolivet, and L. Di Iorio. *Sound Production and Associated Behaviours of Benthic Invertebrates from a Coastal Habitat in the North-East Atlantic*. Marine Biology, 163(5):127, May 2016.

[8] G. M. Wenz. *Acoustic Ambient Noise in the Ocean: Spectra and Sources*. The Journal of the Acoustical Society of America, 34(12):1936–1956, December 1962.

[9] C. M. Duarte, L. Chapuis, S. P. Collin, D. P. Costa, R. P. Devassy, V. M. Eguiluz, C. Erbe, T. A. C. Gordon, B. S. Halpern, H. R. Harding, M. N. Havlik, M. Meekan, N. D. Merchant, J. L. Miksis-Olds, M. Parsons, M. Predragovic, A. N. Radford, C. A. Radford, S. D. Simpson, H. Slabbekoorn, E. Staaterman, I. C. Van Opzeeland, J. Winderen, X. Zhang, and F. Juanes. *The Soundscape of the Anthropocene Ocean*. Science, 371(6529):eaba4658, February 2021.

[10] L. A. Do Nascimento, M. Campos-Cerqueira, and K. H. Beard. *Acoustic Metrics Predict Habitat Type and Vegetation Structure in the Amazon*. Ecological Indicators, 117:106679, October 2020.

[11] S. Dröge, D. A. Martin, R. Andriafanomezantsoa, Z. Burivalova, T. R. Fulgence, K. Osen, E. Rakotomalala, D. Schwab, A. Wurz, T. Richter, and H. Kreft. *Listening to a Changing Landscape: Acoustic Indices Reflect Bird Species Richness and Plot-Scale Vegetation Structure across Different Land-Use Types in North-Eastern Madagascar*. Ecological Indicators, 120:106929, January 2021.

[12] International Organization for Standardization (ISO). *Acoustics - Soundscape - Part 1: Definition and Conceptual Framework*. Technical report, International Organization for Standardization, Geneva, 2014.

[13] J. Sueur and A. Farina. *Ecoacoustics: The Ecological Investigation and Interpretation of Environmental Sound*. Biosemiotics, 8(3):493–502, December 2015.

[14] A. Farina. *Ecoacoustic Codes and Ecological Complexity*. Biosystems, 164:147–154, February 2018.

[15] A. Farina. *Ecoacoustics: A Quantitative Approach to Investigate the Ecological Role of Environmental Sounds*. Mathematics, 7(1), 2018.

[16] A. Farina. *Soundscape Ecology: Principles, Patterns, Methods and Applications*. Springer Netherlands, Dordrecht, 2014.

[17] N. Fletcher. *Animal Bioacoustics*. In Handbook of Acoustics. New York, Springer, 2007.

[18] J. Abeßer. *A Review of Deep Learning Based Methods for Acoustic Scene Classification*. Applied Sciences, 10(6):2020, January 2020.

[19] Z. Hao, H. Zhan, C. Zhang, N. Pei, B. Sun, J. He, R. Wu, X. Xu, and C. Wang. *Assessing the Effect of Human Activities on Biophony in Urban Forests Using an Automated Acoustic Scene Classification Model*. Ecological Indicators, 144:109437, November 2022.

[20] R. P. Schoeman, C. Erbe, and S. Plön. *Underwater Chatter for the Win: A First Assessment of Underwater Soundscapes in Two Bays along the Eastern Cape Coast of South Africa*. Journal of Marine Science and Engineering, 10(6):746, June 2022.

[21] P. Nguyen Hong Duc, D. Cazau, P. R. White, O. Gérard, J. Detcheverry, F. Urtizberea, and O. Adam. *Use of Ecoacoustics to Characterize the Marine Acoustic Environment off the North Atlantic French Saint-Pierre-et-Miquelon Archipelago*. Journal of Marine Science and Engineering, 9(2):177, February 2021.

[22] D. Stowell. *Computational Bioacoustics with Deep Learning: A Review and Roadmap*. PeerJ, 10:e13152, March 2022.

[23] R. L. Putland, C. A. F. de Jong, B. Binnerts, A. Farcas, and N. D. Merchant. *Multi-Site Validation of Shipping Noise Maps Using Field Measurements*. Marine Pollution Bulletin, 179:113733, June 2022.

[24] Interreg North Sea Region JOMOPANS. *10 Years of North Sea Soundscape Monitoring. Looking Back on a Four-Year Interreg NSR Project and Looking Forward to the Six-Year Monitoring Cycle*. Technical report, June 2021.

[25] R. P. Schoeman, C. Erbe, G. Pavan, R. Righini, and J. A. Thomas. *Analysis of Soundscapes as an Ecological Tool*. In C. Erbe and J. A. Thomas, editors, Exploring Animal Behavior Through Sound: Volume 1: Methods, pages 217–267. Springer International Publishing, Cham, 2022.

[26] M.-N. Havlik, M. Predragovic, and C. M. Duarte. *State of Play in Marine Soundscape Assessments*. Frontiers in Marine Science, 9, 2022.

[27] IQOE – Marine Bioacoustical Standardization Working Group. *IQOE – Inventory of Existing Standards and Guidelines Relevant to Marine Bioacoustics*. Technical report, IQOE, 2021.

[28] M. Tasker, M. Amundin, M. Andre, A. Hawkins, W. Lang, T. Merck, A. Scholik-Schlomer, J.Teilmann, F. Thomsen, S. Werner, and M. Zakharia. *Marine Strategy Framework Directive : Task Group 11 Report : Underwater Noise and Other Forms of Energy, April 2010*. Publications Office, LU, 2010.

[29] European Commission. Joint Research Centre. Institute for Environment and Sustainability. *Monitoring Guidance for Underwater Noise in European Seas: A Guidance Document within the Common Implementation Strategy for the Marine Strategy Framework Directive. Part II, Monitoring Guidance Specifications*. Publications Office, LU, 2014.

[30] P. Tyack, T. Akamatsu, O. Boebel, L. Chapuis, E. Debusschere, C. de Jong, C. Erbe, K. Evans, J. Gedamke, T. Gridley, G. Haralabus, R. Jenkins, J. Miksis-Olds, H. Sagen, F. Thomsen, K. Thomisch, and E. Urban. *Ocean Sound Essential Ocean Variable Implementation Plan*. Technical report, Zenodo, November 2023.

[31] J.-G. Fischer, D. Kühnel, and F. Basan. *JOMOPANS Measurement Guidelines*. Technical report, EU INTERREG Joint Monitoring Programme for Ambient Noise North Sea (JOMOPANS), 2021.

[32] M. F. McKenna, S. Baumann-Pickering, A. C. M. Kok, W. K. Oestreich, J. D. Adams, J. Barkowski, K. M. Fristrup, J. A. Goldbogen, J. Joseph, E. B. Kim, A. Kügler, M. O. Lammers, T. Margolina, L. E. Peavey Reeves, T. J. Rowell, J. A. Stanley, A. K. Stimpert, E. J. Zang, B. L. Southall, C. C. Wall, S. V. Parijs, and L. T. Hatch. *Advancing the Interpretation of Shallow Water Marine Soundscapes*. Frontiers in Marine Science, 8, 2021.

[33] T. A. Mooney, L. Di Iorio, M. Lammers, T.-H. Lin, S. L. Nedelec, M. Parsons, C. Radford, E. Urban, and J. Stanley. *Listening Forward: Approaching Marine Biodiversity Assessments Using Acoustic Methods*. Royal Society Open Science, 7(8):201287, 2020.

[34] A. Looby, K. Cox, S. Bravo, R. Rountree, F. Juanes, L. K. Reynolds, and C. W. Martin. *A Quantitative Inventory of Global Soniferous Fish Diversity*. Reviews in Fish Biology and Fisheries, 32(2):581–595, June 2022.

[35] M. Solé, K. Kaifu, T. A. Mooney, S. L. Nedelec, F. Olivier, A. N. Radford, M. Vazzana, M. A. Wale, J. M. Semmens, S. D. Simpson, G. Buscaino, A. Hawkins, N. Aguilar de Soto, T. Akamatsu, L. Chauvaud, R. D. Day, Q. Fitzgibbon, R. D. McCauley, and M. André. *Marine Invertebrates and Noise*. Frontiers in Marine Science, 10, 2023.

[36] C. Wall, C. Lembke, and D. Mann. *Shelf-Scale Mapping of Sound Production by Fishes in the Eastern Gulf of Mexico, Using Autonomous Glider Technology*. Marine Ecology Progress Series, 449:55–64, March 2012.

[37] L. Di Iorio, C. Gervaise, V. Jaud, A. A. Robson, and L. Chauvaud. *Hydrophone Detects Cracking Sounds: Non-intrusive Monitoring of Bivalve Movement*. Journal of Experimental Marine Biology and Ecology, 432–433:9–16, November 2012.

[38] X. Mouy, M. Black, K. Cox, J. Qualley, S. Dosso, and F. Juanes. *Identification of Fish Sounds in the Wild Using a Set of Portable Audio-Video Arrays*. Methods in Ecology and Evolution, 14(8):2165–2186, 2023.

[39] R. Dekeling, M. Tasker, A. Van der Graaf, M. Ainslie, M. Andersson, M. André, J. Borsani, K. Brensing, M. Castellote, D. Cronin, J. Dalen, T. Folegot, R. Leaper, J. Pajala, P. Redman, S. Robinson, P. Sigray, G. Sutton, F. Thomsen, S. Werner, D. Wittekind, and J. Young. *Monitoring Guidance for Underwater Noise in European Seas, Part II: Monitoring Guidance Specifications, JRC Scientific and Policy Report EUR 26555 EN, Publications Office of the European Union, Luxembourg, 2014, Doi: 10.2788/27158*. Publications Office of the European Union, Luxembourg, 2014.

[40] Y. Wei, M. Li, J. Liu, Y. Chen, F. Qu, and M. Zhang. *Flow Noise and Its Suppression in Exposed Hydrophone*. Applied Acoustics, 182:108190, November 2021.

[41] C. Bassett, J. Thomson, P. H. Dahl, and B. Polagye. *Flow-Noise and Turbulence in Two Tidal Channels*. The Journal of the Acoustical Society of America, 135(4):1764–1774, April 2014.

[42] A. M. von Benda-Beckmann, P. J. Wensveen, F. I. P. Samarra, S. P. Beerens, and P. J. O. Miller. *Separating Underwater Ambient Noise from Flow Noise Recorded on Stereo Acoustic Tags Attached to Marine Mammals*. Journal of Experimental Biology, 219(15):2271–2275, August 2016.

[43] N. C. F. van Geel, N. D. Merchant, R. M. Culloch, E. W. J. Edwards, I. M. Davies, R. B. O'Hara Murray, and K. L. Brookes. *Exclusion of Tidal Influence on Ambient Sound Measurements*. The Journal of the Acoustical Society of America, 148(2):701–712, August 2020.

[44] J. Y. Chung. *Rejection of Flow Noise Using a Coherence Function Method*. Journal of the Acoustical Society of America, 62(2):388–395, 1977.

[45] V. Otero, S. Pint, K. Deneudt, M. De Rijcke, J. Mortelmans, L. Schepers, P. Cabrera, K. Sabbe, W. Vyverman, M. Vandegehuchte, and G. Everaert. *Pronounced Seasonal and Spatial Variability in Determinants of Phytoplankton Biomass Dynamics along a Near–Offshore Gradient in the Southern North Sea*. Biogeosciences Discussions, pages 1–45, January 2022.

[46] S. Degraer, R. Brabant, and B. Rumes. *Offshore Wind Farms in the Belgian Part of the North Sea: Heading for an Understanding of Environmental Impacts.* Technical report, Royal Belgian Institute of Natural Sciences (RBINS), September 2012.

[47] J. L. Miksis-Olds, B. Martin, and P. L. Tyack. *Exploring the Ocean Through Soundscapes*. Acoustics Today, 14(1):9, 2018.

[48] C. W. Holland. *Propagation in a Waveguide with Range-dependent Seabed Properties. I. Theory and Implications.* The Journal of the Acoustical Society of America, 129(4_Supplement):2425, April 2011.

[49] Y. Durgachandramouli, A. Sailaja, P. J. Josephson, T. N. Prasad, K. E. Prasad, M. R. Sankar, and C. V. Rao. *Analysis of Acoustic Channel Characteristics in Shallow Water Based on Multipath Model*. In P. Pareek, N. Gupta, and M. J. C. S. Reis, editors, Cognitive Computing and Cyber Physical Systems, Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, pages 288–297, Cham, 2024. Springer Nature Switzerland.

[50] C. Mueller, A. Monczak, J. Soueidan, B. McKinney, S. Smott, T. Mills, Y. Ji, and EW. Montie. *Sound Characterization and Fine-Scale Spatial Mapping of an Estuarine Soundscape in the Southeastern USA*. Marine Ecology Progress Series, 645:1–23, 2020.

[51] J. K. Garrett, Ph. Blondel, B. J. Godley, S. K. Pikesley, M. J. Witt, and L. Johanning. *Long-Term Underwater Sound Measurements in the Shipping Noise Indicator Bands 63Hz and 125Hz from the Port of Falmouth Bay, UK*. Marine Pollution Bulletin, 110(1):438–448, September 2016.

[52] A. MacGillivray and C. de Jong. *A Reference Spectrum Model for Estimating Source Levels of Marine Shipping Based on Automated Identification System Data*. Journal of Marine Science and Engineering, 9(4):369, April 2021.

[53] M. A. Ainslie. *Effect of Wind-Generated Bubbles on Fixed Range Acoustic Attenuation in Shallow Water at 1 – 4 kHz*. The Journal of the Acoustical Society of America, 3513(2005), 2011.

[54] D. Cazau, J. Bonnel, and M. Baumgartner. *Wind Speed Estimation Using Acoustic Underwater Glider in a Near-Shore Marine Environment*. IEEE Transactions on Geoscience and Remote Sensing, 57(4):2097–2106, April 2019.

[55] H. Ö. Sertlek, B. Binnerts, and M. A. Ainslie. *The Effect of Sound Speed Profile on Shallow Water Shipping Sound Maps*. The Journal of the Acoustical Society of America, 140(1):EL84–EL88, 2016.

[56] A. O. MacGillivray, S. B. Martin, M. A. Ainslie, J. N. Dolman, Z. Li, and G. A. Warner. *Measuring Vessel Underwater Radiated Noise in Shallow Water*. The Journal of the Acoustical Society of America, 153(3):1506–1524, March 2023.

[57] A. Farcas, C. F. Powell, K. L. Brookes, and N. D. Merchant. *Validated Shipping Noise Maps of the Northeast Atlantic*. Science of The Total Environment, 735:139509, September 2020.

[58] E. Ivanov, A. Capet, A. Barth, E. J. M. Delhez, K. Soetaert, and M. Grégoire. *Hydrodynamic Variability in the Southern Bight of the North Sea in Response to Typical Atmospheric and Tidal Regimes. Benefit of Using a High Resolution Model*. Ocean Modelling, 154:101682, October 2020.

[59] M. Fettweis and D. Van den Eynde. *The Mud Deposits and the High Turbidity in the Belgian–Dutch Coastal Zone, Southern Bight of the North Sea*. Continental Shelf Research, 23(7):669–691, May 2003.

[60] D. Termote. *Schatten en scheepswrakken : boeiende onderwaterarcheologie in de Noordzee - Dirk Termote | Vlaamse catalogus*. Davidsfonds, 2009.

[61] F. Basan, J. G. Fischer, R. Putland, J. Brinkkemper, C. A. F. de Jong, B. Binnerts, A. Norro, D. Kühnel, L. A. Ødegaard, M. Andersson, E. Lalander, J. Tougaard, E. T. Griffiths, M. Kosecka, E. Edwards, N. D. Merchant, K. de Jong, S. Robinson, L. Wang, and N. Kinneging. *The Underwater Soundscape of the North Sea*. Marine Pollution Bulletin, 198:115891, January 2024.

[62] J. Tougaard, L. Hermannsen, and P. T. Madsen. *How Loud Is the Underwater Noise from Operating Offshore Wind Turbines?* The Journal of the Acoustical Society of America, 148(5):2885–2893, November 2020.

[63] C. Parcerisas, E. Schall, K. te Velde, D. Botteldooren, P. Devos, and E. Debusschere. *Machine Learning for Efficient Segregation and Labeling of Potential Biological Sounds in Long-Term Underwater Recordings*. Frontiers in Remote Sensing, 5, April 2024.

[64] S. D. Jarriel, N. Formel, S. R. Ferguson, F. H. Jensen, A. Apprill, and T. A. Mooney. *Unidentified Fish Sounds as Indicators of Coral Reef Health and Comparison to Other Acoustic Methods*. Frontiers in Remote Sensing, 5, March 2024.

[65] C. Parcerisas, D. Botteldooren, P. Devos, Q. Hamard, and E. Debusschere. *Studying the Soundscape of Shallow and Heavy Used Marine Areas: Belgian Part of the North Sea*. In A. N. Popper, J. Sisneros, A. D. Hawkins, and F. Thomsen, editors, The Effects of Noise on Aquatic Life, pages 1–27. Springer International Publishing, Cham, 2023.

[66] European Comission. *COMMUNICATION FROM THE COMMISSION TO THE EUROPEAN PARLIAMENT, THE COUNCIL, THE EUROPEAN ECONOMIC AND SOCIAL COMMITTEE AND THE COMMITTEE OF THE REGIONS An EU Strategy to Harness the Potential of Offshore Renewable Energy for a Climate Neutral Future*, 2020.

[67] F. Thomsen, S. Mendes, F. Bertucci, M. Breitzke, E. Ciappi, A. Cresci, E. Debusschere, C. Ducatel, T. Folegot, C. Juretzek, Frans-Peter Lam, J. O'Brien, and M. E. D. Santos. *Addressing Underwater Noise in Europe: Current State of Knowledge and Future Priorities*. Technical report, Zenodo, October 2021.

[68] P. A. k, M. M, S. Rajamanickam, S. Sivarethinamohan, M. K. R. Gaddam, P. Velusamy, G. R, G. Ravindiran, T. R. Gurugubelli, and S. K. Muniasamy. *Impact of Climate Change and Anthropogenic Activities on Aquatic Ecosystem – A Review*. Environmental Research, 238:117233, December 2023.

[69] Commission Directive (EU). *Directive 2008/56/EC of the European Parliament and of the Council of 17 June 2008 Establishing a Framework for Community Action in the Field of Marine Environmental Policy (Marine Strategy Framework Directive) (Text with EEA Relevance)*, June 2017.

[70] C. Parcerisas, I. T. Roca, D. Botteldooren, P. Devos, and E. Debusschere. *Categorizing Shallow Marine Soundscapes Using Explained Clusters*. Journal of Marine Science and Engineering, 11(3):550, March 2023.

[71] T.-H. Lin and Y. Tsao. *Source Separation in Ecoacoustics: A Roadmap towards Versatile Soundscape Information Retrieval*. Remote Sensing in Ecology and Conservation, 6(3):236–247, 2020.

[72] O. Adam. *Segmentation of Killer Whale Vocalizations Using the Hilbert-Huang Transform*. EURASIP Journal on Advances in Signal Processing, 2008(1):1–10, December 2008.

[73] F. Dadouchi, C. Gervaise, C. Ioana, J. Huillery, and J. I. Mars. *Automated Segmentation of Linear Time-Frequency Representations of Marine-Mammal Sounds*. The Journal of the Acoustical Society of America, 134(3):2546–2555, August 2013.

[74] H. Ou, W. W. L. Au, and J. N. Oswald. *A Non-Spectrogram-Correlation Method of Automatically Detecting Minke Whale Boings*. The Journal of the Acoustical Society of America, 132(4):EL317–EL322, September 2012.

[75] P. Rubbens, S. Brodie, T. Cordier, D. Destro Barcellos, P. Devos, J. A. Fernandes-Salvador, J. I. Fincham, A. Gomes, N. O. Handegard, K. Howell, C. Jamet, K. H. Kartveit, H. Moustahfid, C. Parcerisas, D. Politikos, R. Sauzède, M. Sokolova, L. Uusitalo, L. Van den Bulcke, A. T. M. van Helmond, J. T. Watson, H. Welch, O. Beltran-Perez, S. Chaffron, D. S. Greenberg, B. Kühn, R. Kiko, M. Lo, R. M. Lopes, K. O. Möller, W. Michaels, A. Pala, J.-B. Romagnan, P. Schuchert, V. Seydi, S. Villasante, K. Malde, and J.-O. Irisson. *Machine Learning in Marine Ecology: An Overview of Techniques and Applications*. ICES Journal of Marine Science, page fsad100, August 2023.

[76] E. Schall, I. I. Kaya, E. Debusschere, P. Devos, and C. Parcerisas. *Deep Learning in Marine Bioacoustics: A Benchmark for Baleen Whale Detection*. Remote Sensing in Ecology and Conservation, April 2024.

[77] D. K. Mellinger. *A Comparison of Methods for Detecting Right Whale Calls*. Canadian Acoustics, 32(2):55–65, June 2004.

[78] M. Bittle and A. Duncan. *A Review of Current Marine Mammal Detection and Classification Algorithms for Use in Automated Passive Acoustic Monitoring*. In Proceedings of Acoustics 2013, Australia, November 2013.

[79] M. A. Roch, M. S. Soldevilla, R. Hoenigman, S. M. Wiggins, and J. A. Hildebrand. *Comparison of Machine Learning Techniques for the Classification of Echolocation Clicks from Three Species of Odontocetes*. Canadian Acoustics, 36(1):41–47, March 2008.

[80] C. Bergler, H. Schröter, R. X. Cheng, V. Barth, M. Weber, E. Nöth, H. Hofer, and A. Maier. *ORCA-SPOT: An Automatic Killer Whale Sound Detection Toolkit Using Deep Learning*. Scientific Reports, 9(1):1–17, 2019.

[81] F. Merchan, A. Guerra, H. Poveda, H. M. Guzmán, and J. E. Sanchez-Galan. *Bioacoustic Classification of Antillean Manatee Vocalization Spectrograms Using Deep Convolutional Neural Networks*. Applied Sciences, 10(9):3286, January 2020.

[82] S. Zaugg, M. van der Schaar, L. Houégnigan, C. Gervaise, and M. André. *Real-Time Acoustic Classification of Sperm Whale Clicks and Shipping Impulses from Deep-Sea Observatories*. Applied Acoustics, 71(11):1011–1019, November 2010.

[83] C. Soares, F. Zabel, and S. M. Jesus. *A Shipping Noise Prediction Tool*. In OCEANS 2015 - Genova, pages 1–7, Genova, Italy, May 2015. IEEE.

[84] V. R. Barroso, F. C. Xavier, and C. E. L. Ferreira. *Applications of Machine Learning to Identify and Characterize the Sounds Produced by Fish*. ICES Journal of Marine Science, 80(7):1854–1867, September 2023.

[85] P. C. Bermant, M. M. Bronstein, R. J. Wood, S. Gero, and D. F. Gruber. *Deep Machine Learning Techniques for the Detection and Classification of Sperm Whale Bioacoustics*. Scientific Reports, 9(1):12588, August 2019.

[86] C. Bergler, M. Schmitt, R. X. Cheng, H. Schröter, A. Maier, V. Barth, M. Weber, and E. Nöth. *Deep Representation Learning for Orca Call Type Classification*. In K. Ekštein, editor, Text, Speech, and Dialogue, Lecture Notes in Computer Science, pages 274–286, Cham, 2019. Springer International Publishing.

[87] L. Weilgart and H. Whitehead. *Group-Specific Dialects and Geographical Variation in Coda Repertoire in South Pacific Sperm Whales*. Behavioral Ecology and Sociobiology, 40(5):277–285, May 1997.

[88] J. C. Brown, P. Smaragdis, and A. Nousek-McGregor. *Automatic Identification of Individual Killer Whales*. The Journal of the Acoustical Society of America, 128(3):EL93–EL98, August 2010.

[89] V. B. Deecke, J. K. B. Ford, and P. Spong. *Quantifying Complex Patterns of Bioacoustic Variation: Use of a Neural Network to Compare Killer Whale (Orcinus Orca) Dialects*. The Journal of the Acoustical Society of America, 105(4):2499–2507, April 1999.

[90] G. Sharma, K. Umapathy, and S. Krishnan. *Trends in Audio Signal Feature Extraction Methods*. Applied Acoustics, 158:107020, January 2020.

[91] B. Sadaiappan, P. Balakrishnan, V. C.R., N. T. Vijayan, M. Subramanian, and M. U. Gauns. *Applications of Machine Learning in Chemical and Biological Oceanography*. ACS Omega, 8(18):15831–15853, May 2023.

[92] M. A. Roch, S. Lindeneau, G. S. Aurora, K. E. Frasier, J. A. Hildebrand, H. Glotin, and S. Baumann-Pickering. *Using Context to Train Time-Domain Echolocation Click Detectors*. The Journal of the Acoustical Society of America, 149(5):3301–3310, May 2021.

[93] M. Hagiwara. *AVES: Animal Vocalization Encoder Based on Self-Supervision*. In ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pages 1–5, June 2023.

[94] H. Schröter, E. Nöth, A. Maier, R. Cheng, V. Barth, and C. Bergler. *Segmentation, Classification, and Visualization of Orca Calls Using Deep Learning*. In ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pages 8231–8235, May 2019.

[95] P. Best, S. Paris, H. Glotin, and R. Marxer. *Deep Audio Embeddings for Vocalisation Clustering*. PLOS ONE, 18(7):e0283396, July 2023.

[96] D. Robinson, A. Robinson, and L. Akrapongpisak. *Transferable Models for Bioacoustics with Human Language Supervision*, August 2023.

[97] J. F. Gemmeke, D. P. W. Ellis, D. Freedman, A. Jansen, W. Lawrence, R. C. Moore, M. Plakal, and M. Ritter. *Audio Set: An Ontology and Human-Labeled Dataset for Audio Events*. In 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pages 776–780, New Orleans, LA, March 2017. IEEE.

[98] R. Harakawa, T. Ogawa, M. Haseyama, and T. Akamatsu. *Automatic Detection of Fish Sounds Based on Multi-Stage Classification Including Logistic Regression via Adaptive Feature Weighting*. The Journal of the Acoustical Society of America, 144(5):2709–2718, November 2018.

[99] I. R. Urazghildiiev and S. M. Van Parijs. *Automatic Grunt Detector and Recognizer for Atlantic Cod (* Gadus Morhua *)*. The Journal of the Acoustical Society of America, 139(5):2532–2540, May 2016.

[100] S. Jarvis, N. DiMarzio, R. Morrissey, and D. Moretti. *A Novel Multi-Class Support Vector Machine Classifier for Automated Classification of Beaked Whales and Other Small Odontocetes*. Canadian Acoustics, 36(1):34–40, March 2008.

[101] M. Malfante, J. I. Mars, M. Dalla Mura, and C. Gervaise. *Automatic Fish Sounds Classification*. The Journal of the Acoustical Society of America, 143(5):2834–2846, May 2018.

[102] M. A. Roch, H. Klinck, S. Baumann-Pickering, D. K. Mellinger, S. Qui, M. S. Soldevilla, and J. A. Hildebrand. *Classification of Echolocation Clicks from Odontocetes in the Southern California Bight*. The Journal of the Acoustical Society of America, 129(1):467–475, January 2011.

[103] H. Ghaffari and P. Devos. *On the Role of Audio Frontends in Bird Species Recognition*. Ecological Informatics, 81:102573, July 2024.

[104] M. M. Hotkani, J.-F. Bousquet, B. Martin, L. Bolt, and B. MacDonald. *A Real-Time Acoustic Spectrum Analyzer on FPGA for the Internet of Underwater Things*. In 2022 20th IEEE Interregional NEWCAS Conference (NEWCAS), pages 426–430, June 2022.

[105] B. M. Howe, J. Miksis-Olds, E. Rehm, H. Sagen, P. F. Worcester, and G. Haralabus. *Observing the Oceans Acoustically*. Frontiers in Marine Science, 0, 2019.

[106] A. Looby, S. Vela, K. Cox, A. Riera, S. Bravo, H. L. Davies, R. Rountree, L. K. Reynolds, C. W. Martin, S. Matwin, and F. Juanes. *FishSounds Version 1.0: A Website for the Compilation of Fish Sound Production Information and Recordings*. Ecological Informatics, 74:101953, May 2023.

[107] M. J. G. Parsons, T.-H. Lin, T. A. Mooney, C. Erbe, F. Juanes, M. Lammers, S. Li, S. Linke, A. Looby, S. L. Nedelec, I. Van Opzeeland, C. Radford, A. N. Rice, L. Sayigh, J. Stanley, E. Urban, and L. Di Iorio. *Sounding the Call for a Global Library of Underwater Biological Sounds*. Frontiers in Ecology and Evolution, 10, 2022.

[108] S. M. Balvanera, O. Mac Aodha, M. J. Weldy, H. Pringle, E. Browning, and K. E. Jones. *Whombat: An Open-Source Annotation Tool for Machine Learning Development in Bioacoustics*, November 2023.

[109] A. Solsona-Berga, K. E. Frasier, S. Baumann-Pickering, S. M. Wiggins, and J. A. Hildebrand. *DetEdit: A Graphical User Interface for Annotating and Editing Events Detected in Long-Term Acoustic Monitoring Data*. PLOS Computational Biology, 16(1):e1007598, 13 de gen. 2020.

[110] H. Kath, T. S. Gouvêa, and D. Sonntag. *A Human-in-the-Loop Tool for Annotating Passive Acoustic Monitoring Datasets*. In Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence, pages 7140–7144, Macau, SAR China, August 2023. International Joint Conferences on Artificial Intelligence Organization.

[111] I. Nolasco, S. Singh, V. Morfi, V. Lostanlen, A. Strandburg-Peshkin, E. Vidaña-Vila, L. Gill, H. Pamuła, H. Whitehead, I. Kiskin, F. H. Jensen, J. Morford, M. G. Emmerson, E. Versace, E. Grout, H. Liu, B. Ghani, and D. Stowell. *Learning to Detect an Animal Sound from Five Examples*. Ecological Informatics, 77:102258, November 2023.

[112] A. M. Thode, K. H. Kim, S. B. Blackwell, C. R. Greene, C. S. Nations, T. L. McDonald, and A. M. Macrander. *Automated Detection and Localization of Bowhead Whale Sounds in the Presence of Seismic Airgun Surveys*. The Journal of the Acoustical Society of America, 131(5):3726–3747, May 2012.

[113] S. F. Woodward, D. Reiss, and M. O. Magnasco. *Learning to Localize Sounds in a Highly Reverberant Environment: Machine-learning Tracking of Dolphin Whistle-like Sounds in a Pool*. PLOS ONE, 15(6):e0235155, June 2020.

[114] H. Niu, E. Reeves, and P. Gerstoft. *Source Localization in an Ocean Waveguide Using Supervised Machine Learning*. The Journal of the Acoustical Society of America, 142(3):1176–1188, September 2017.

[115] W. A. M. Beslin, H. Whitehead, and S. Gero. *Automatic Acoustic Estimation of Sperm Whale Size Distributions Achieved through Machine Recognition of On-Axis Clicks*. The Journal of the Acoustical Society of America, 144(6):3485–3495, December 2018.

[116] T. J. Rowell, D. A. Demer, O. Aburto-Oropeza, J. J. Cota-Nieto, J. R. Hyde, and B. E. Erisman. *Estimating Fish Abundance at Spawning Aggregations from Courtship Sound Levels*. Scientific Reports, 7(1):1–14, 2017.

[117] S. Innami and H. Kasai. *NMF-based Environmental Sound Source Separation Using Time-Variant Gain Features*. Computers & Mathematics with Applications, 64(5):1333–1342, September 2012.

[118] Y.-J. Sun, S.-C. Yen, and T.-H. Lin. *soundscape_IR: A Source Separation Toolbox for Exploring Acoustic Diversity in Soundscapes*. Methods in Ecology and Evolution, 13(11):2347–2355, 2022.

[119] M. Mancusi, N. Zonca, E. Rodolà, and S. Zuffi. *Fish Sounds: Towards the Evaluation of Marine Acoustic Biodiversity through Data-Driven Audio Source Separation*, January 2022.

[120] Z. Zhang and P. R. White. *A Blind Source Separation Approach for Humpback Whale Song Separation*. The Journal of the Acoustical Society of America, 141(4):2705–2714, April 2017.

[121] X. Deng, Y. Tao, X. Tu, and X. Xu. *The Separation of Overlapped Dolphin Signature Whistle Based on Blind Source Separation*. In 2017 IEEE International Conference on Signal Processing, Communications and Computing (ICSPCC), pages 1–5, October 2017.

[122] T.-H. Lin and Y. Tsao. *Listening to the Deep: Exploring Marine Soundscape Variability by Information Retrieval Techniques*. In 2018 OCEANS - MTS/IEEE Kobe Techno-Oceans (OTO), pages 1–6, May 2018.

[123] T.-H. Lin, T. Akamatsu, and Y. Tsao. *Sensing Ecosystem Dynamics via Audio Source Separation: A Case Study of Marine Soundscapes off Northeastern Taiwan*. PLOS Computational Biology, 17(2):e1008698, 18 de febr. 2021.

[124] E. B. Kim, K. E. Frasier, M. F. McKenna, A. C. M. Kok, L. E. Peavey Reeves, W. K. Oestreich, G. Arrieta, S. Wiggins, and S. Baumann-Pickering. *SoundScape Learning: An Automatic Method for Separating Fish Chorus in Marine Soundscapes*. The Journal of the Acoustical Society of America, 153(3):1710–1722, March 2023.

[125] P. C. Bermant. *BioCPPNet: Automatic Bioacoustic Source Separation with Deep Neural Networks*. Scientific Reports, 11(1):23502, December 2021.

[126] B. Williams, T. A. C. Lamont, L. Chapuis, H. R. Harding, E. B. May, M. E. Prasetya, M. J. Seraphim, J. Jompa, D. J. Smith, N. Janetski, A. N. Radford, and S. D. Simpson. *Enhancing Automated Analysis of Marine Soundscapes Using Ecoacoustic Indices and Machine Learning*. Ecological Indicators, 140:108986, July 2022.

[127] B. Beckler, A. Pfau, M. Orescanin, S. Atchley, N. Villemez, J. E. Joseph, C. W. Miller, and T. Margolina. *Multilabel Classification of Heterogeneous Underwater Soundscapes With Bayesian Deep Learning*. Ieee Journal of Oceanic Engineering, 47(4):1143–1154, October 2022.

[128] E. Ozanich, A. Thode, P. Gerstoft, L. A. Freeman, and S. Freeman. *Deep Embedded Clustering of Coral Reef Bioacoustics*. The Journal of the Acoustical Society of America, 149(4):2587–2601, April 2021.

[129] T. Sainburg, M. Thielk, and T. Q. Gentner. *Finding, Visualizing, and Quantifying Latent Structure across Diverse Animal Vocal Repertoires*. PLOS Computational Biology, 16(10):e1008228, October 2020.

[130] L. McInnes, J. Healy, and J. Melville. *UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction*, September 2020.

[131] L. McInnes, J. Healy, and S. Astels. *Hdbscan: Hierarchical Density Based Clustering*. Journal of Open Source Software, 2(11):205, March 2017.

[132] T. Rhinehart, D. Nicholson, L. Freeland-Haynes, and J. Kitzes. *Bioacoustics Software Database*, 2024.

[133] G. Pavan. *Fundamentals of Soundscape Conservation*. In Ecoacoustics, chapter 14, pages 235–258. John Wiley & Sons, Ltd, 2017.

[134] R. Carson. *Silent Spring*. Houghton Mifflin Harcourt, January 2024.

[135] A. McDermott. *Artists and Scientists Come Together to Explore the Meaning of Natural Sound*. Proceedings of the National Academy of Sciences, 116(26):12580–12583, June 2019.

[136] D. Monacchi. Fragments of Extinction: *Acoustic Biodiversity of Primary Rainforest Ecosystems*. Leonardo Music Journal, 23:23–25, December 2013.

[137] R. M. Schafer. *The Tuning of the World*. New York : A. A. Knopf, 1977.

Part I

The Data

2

The datasets

Overview

In this chapter we provide an overview of the underwater recordings collected during this dissertation, separated in two datasets,. These datasets have been published and stored according to the FAIR (Findable, Accessible, Interoperable, and Reusable) principle. Subsections from these datasets have been selected and used for Chapter 7 and Chapter 8. This chapter has not been published and has been written specifically for this thesis.

2.1 Introduction

Data presented and analyzed in this PhD dissertation were (and still are) collected in the BPNS, which is a small part of the Southern part of the North Sea (for more information see Chapter 3). The collected data are part of the LifeWatch Broadband Acoustic Network [1]. This network is part of the LifeWatch project, which is an European Strategy Forum on Research Infrastructures (ESFRI) project, and it is part of the marine, freshwater and terrestrial observatory in Belgium. Recording in this environment presented certain challenges, some of which are more detailed in Chapter 3. For clarity, we define a deployment as a measuring interval corresponding to the time when a hydrophone is in the water, without changing any recording parameters.

2.2 Stationary dataset: LifeWatch Broadband Acoustic Network

The network was set up with the aim to collect (semi-)continuous acoustic data at several locations to provide a broad temporal coverage which would capture long term trend and seasonal, lunar and diel patterns as well as some spatial ones. With this purpose 2 locations were defined as ”fixed”, so there was always a recorder deployed on that location, and then other 5 rotating stations from which only 2 would be recording at a time. The long-term audio data were collected in the framework of the LifeWatch Broadband Acoustic Network [1] using four RESEA 320 recorders (RTSys, France) together with one or two Colmar GP1190M-LP hydrophones (Colmar, Italy, nominal sensitivity: -180 dB/V re $1\mu Pa$, frequency range -3 dB: 10 Hz to 170 kHz). The recorders were set to record at 48 kS/s or 96 kS/s, and they would record either continuously or with a 24h-on , 24h-off duty cycle.

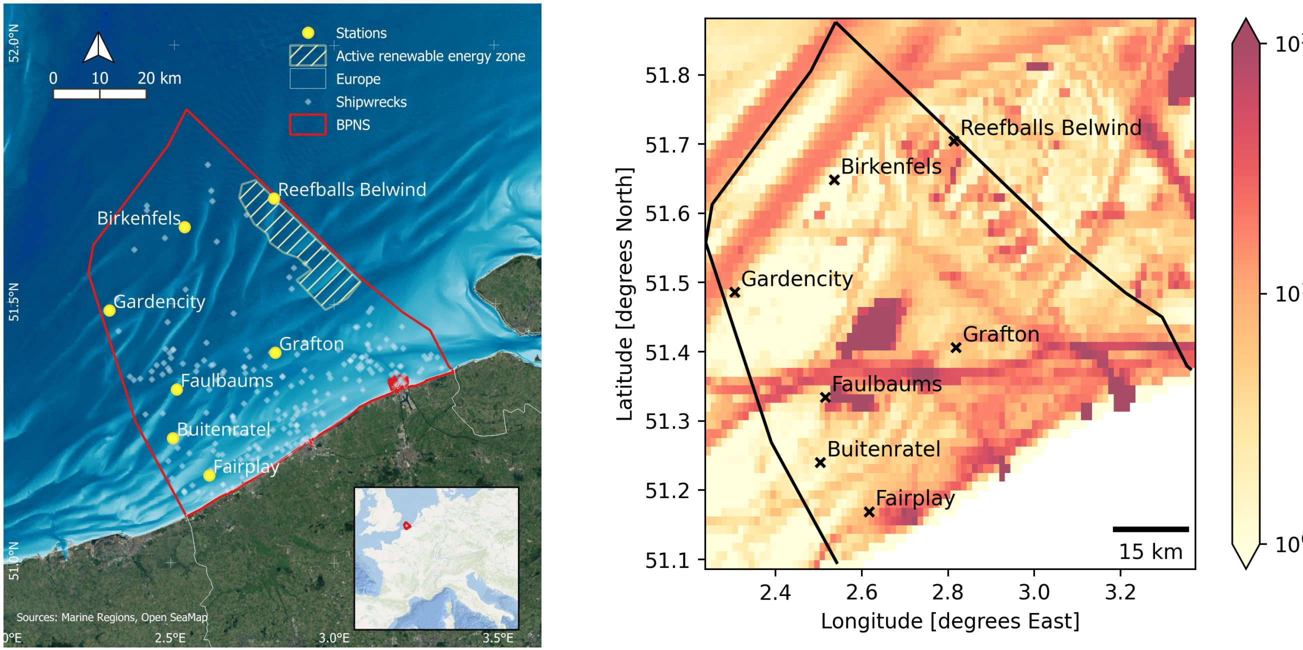
While, with the recent release of the SoundTrap 600 (Oceaninstruments, New Zealand) and other newer technologies it is now possible to record autonomously for long periods (months), when this thesis started it was necessary to have big battery packs with Lithium D-cells to record for longer periods. For this reason it was necessary to design a system to moor these instruments which was big enough for a 1.2 m battery pack. Inspired by the tripod from Goossens et al. (2020) [2], a modular steel frame was designed, tested and used to deploy the acoustic recorders. The instruments were attached to the mooring frame at 1 m above the sea floor using pipe clamps, with no moving parts (see Figure 2.1). Attached to each mooring, apart form the acoustic recorder, there was always a C-POD or F-POD (Chelonia Ltd, Penzance, UK) and a VR2AR (Innovasea Systems Inc, Boston, Massachusetts, USA) or a TBR (Thelma Biotel, Trondheim, Norway) Acoustic Receiver and Release. The C-POD and the F-POD allowed to simultaneously detect harbor porpoise and dolphins clicks, which were too high for the sampling rate of the recorders when set up to record for longer periods of time. The acoustic release and receiver had a double function: to detect tagged fish in the framework of the European Tracking Network (ETN, `https://www.lifewatch.be/etn`) and to allow to recover the entire mooring without leaving anything behind on the seafloor.

The recording sites were fixed during an entire deployment, and they are plotted in Figures 2.2 and 2.3, with different background information. The stations were selected according to their ecological relevance based on previous knowledge obtained from the LifeWatch Cetacean network [3] and the Acoustic Receiver Network [4]. Faulbaums and Grafton are situated mid-shore but still represent the coastal waters. Birkenfels and GardenCity are two offshore stations, respectively on the east and west side of the BPNS. It is known that harbor porpoises prefer the western offshore zone of the BPNS [5], therefore it is interesting to monitor the soundscape between those two zones. Buitenratel is a highly productive and extremely dynamic shallow area. Harbor porpoises are only present in low densities over summer throughout the BPNS except in Buitenratel, where they are abundant. This reflects high densities of prey availability. Fairplay is a coastal station, shielded from the shipping lanes and close to a Natura 2000 area. Reefballs Belwind is situated within the offshore wind farm Belwind and monitors the operational underwater sound and the ecosystem created by the artificial reefs resulting from the installation of the turbines.

Because of the heavy fishing and traffic activities, all the stations were always located next to a shipwreck, as shipwrecks are usually avoided by trawling fishing boats. Shipwrecks are usually *hotspots* for biodiversity in the BPNS, and therefore provide more chances of monitoring biological sounds. However, only recording close to shipwrecks also introduces a bias in the dataset which should be accounted for when doing biological and ecological analysis.

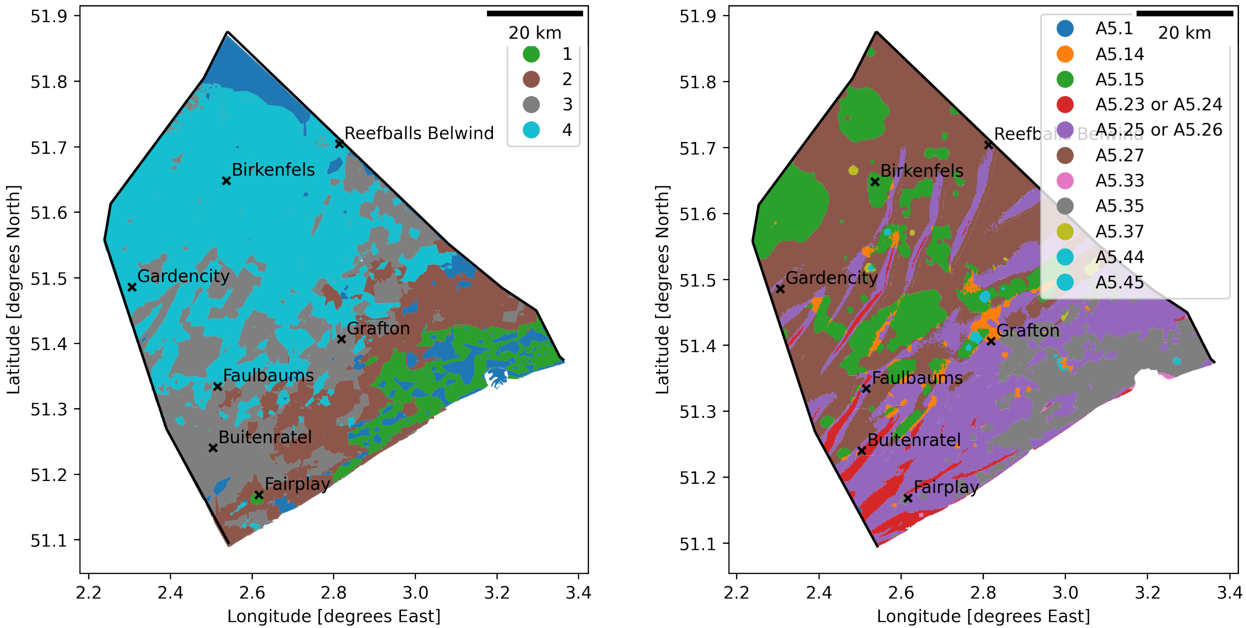
Figure 2.1 Mooring with the acoustic recorder and the hydrophone, the C-POD and the acoustic receiver attached ready to be deployed from the RV Simon Stevin

Figure 2.2 The stations of the LifeWatch Broadband Acoustic Network [1] in the Belgian part of the North Sea (BPNS), plotted on different backgrounds. **(a)** Vessel density from EMODnet [6] and **(b)** Active renewable energy zone and shipwrecks.



(a) **(b)**

Figure 2.3 The stations of the LifeWatch Broadband Acoustic Network [1] in the Belgian part of the North Sea (BPNS), plotted on different backgrounds. **(a)** The benthic habitat type modeled by Derous et al. (2007) [7] labels represent (1) *Macoma balthica* community; (2) *Abra alba* community; (3) *Nephtys cirrosa* community and (4) *Ophelia limacina* community. **(b)** The seabed sediment type by EMODnet Seabed Habitats [8].



(a) **(b)**

A list of all the deployments which have been successfully completed during this PhD is shown in Table 2.1, and summarized in Figure 2.4. Finally, we provide a table summarizing the basic characteristics of each station for future reference (see Table 2.2).

Table 2.1: Summary of all the successful deployments performed for this dissertation in the framework of the LifeWatch Broadband Acoustic Network [1] by 7th April 2024. HMBL stands for hybrid millidecade band (sound pressure) levels, and a value of 1 is assigned to the deployments which have been processed.

Acoustic Receiver	**Start**	**End**	**Station**	**HMBL**	**SR [kHz]**
RESEA320 1	09/03/2021	19/05/2021	Grafton	1	48
RESEA320 2	09/03/2021	27/03/2021	Birkenfels	1	96
RESEA320 3	09/03/2021	06/05/2021	Gardencity	1	96
RESEA320 4	09/03/2021	02/07/2021	Faulbaums	1	48
RESEA320 3	11/05/2021	11/06/2021	Fairplay	1	96
RESEA320 2	13/07/2021	05/08/2021	Grafton	1	48
RESEA320 3	19/01/2022	22/04/2022	Grafton	1	48
RESEA320 2	20/01/2022	04/05/2022	Buitenratel	1	48
RESEA320 4	20/01/2022	23/05/2022	Gardencity	1	48
RESEA320 1	27/04/2022	19/06/2022	Faulbaums	1	48
RESEA320 3	22/06/2022	01/07/2022	Gardencity	1	48
RESEA320 2	24/06/2022	22/07/2022	Belwind	1	48
RESEA320 1	20/08/2022	27/10/2022	Grafton	1	48
RESEA320 1	28/10/2022	08/11/2022	Grafton	1	48
ST300-5118	27/06/2023	10/10/2023	Grafton	0	48
ST600HF-6714	29/06/2023	07/09/2023	Gardencity	0	48
RESEA320 2	29/06/2023	06/09/2023	Faulbaums	0	48
ST600HF-7560	15/07/2023	19/10/2023	Grafton	0	48
ST600HF-7564	09/09/2023	12/12/2023	Faulbaums	0	48
ST600HF-7575	09/09/2023	ongoing	Gardencity	0	48
ST600HF-7573	04/10/2023	06/12/2023	Grafton	0	48
ST600HF-7565	09/11/2023	12/12/2023	Faulbaums	0	48
ST600HF-7560	08/12/2023	ongoing	Faulbaums	0	96
ST600HF-7565	12/12/2023	ongoing	Grafton	0	96
ST600HF-7564	23/02/2024	ongoing	Faulbaums	0	96
ST600HF-7573	23/02/2024	ongoing	Grafton	0	96

Figure 2.4 Summary of all the collected data in the framework of the LifeWatch Broadband Acoustic Network [1] by 7th April 2024. Data were considered invalid due to recorder or hydrophone malfunctioning.

Table 2.2 Summary of the different stations and their characteristics. See Figure 2.2 for a map of the stations location.

Station name	Description	Depth [m]	Lat [°N]	Lon [°E]
Buitenratel	Close to shore	7	51.2403	2.504
Fairplay	Close to shore	14.5	51.1689	2.6167
Grafton	Mid-shore, close to shipping lane	19	51.4065	2.8185
Faulbaums	Mid-shore	19	51.3346	2.6151
Gardencity	Off-shore, close to shipping lane	28	51.4865	2.3048
Birkenfels	Off-shore, close to shipping line	37.5	51.6483	2.5362
Reefballs Belwind	Windpark	26	51.7044	2.8132

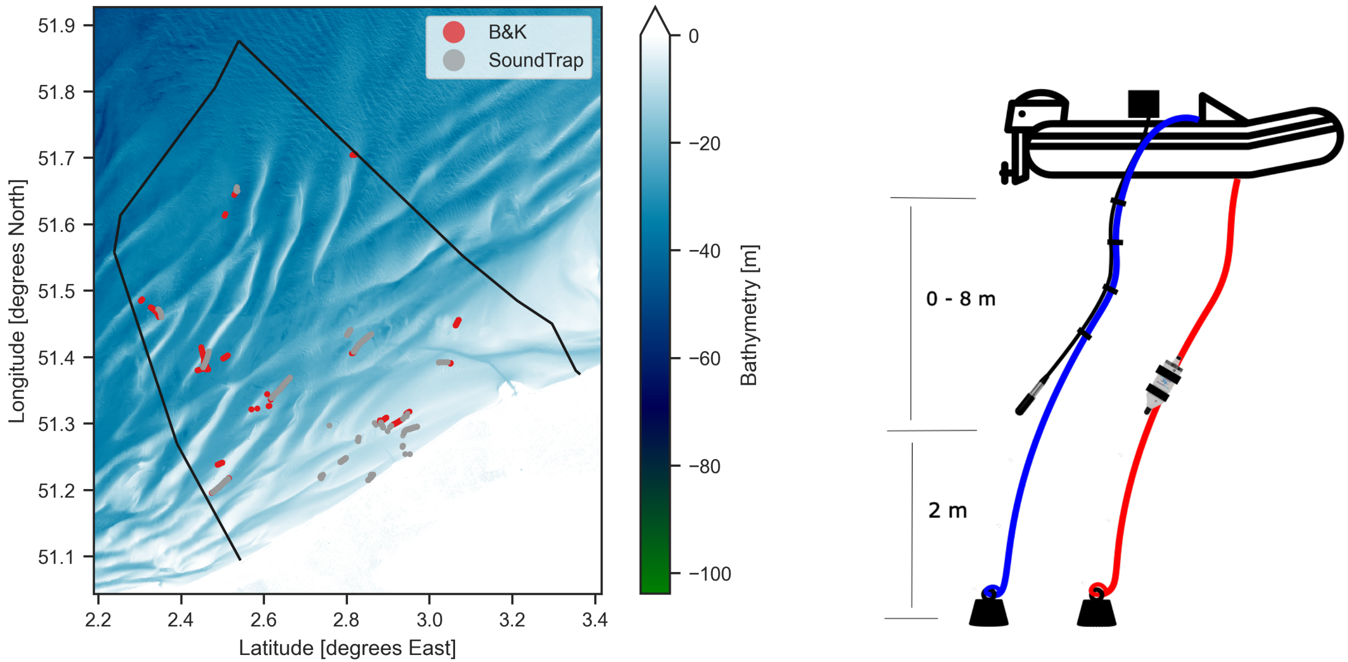
2.3 Drifts dataset

Between April 2020 and October 2020, and in June 2021, we recorded underwater sound at 40 strategic sites of the BPNS (see Figure 2.5a). All the recordings were acquired from three different small boats, adrift with a hydrophone attached to a rope with weights (see Table 2.3 and Figure 2.5). Drifting was chosen as an ecologically meaningful approach to acoustically record the different coastal benthic habitats [9]. Considering the available ship time and equipment, spatial coverage was preferred over temporal coverage, meaning that recording more sites was prioritized over acquiring long recordings at one site. Furthermore, recording while drifting was expected to diminish the possible flow noise due to the tidal current. Locations were chosen to cover the four benthic habitat types defined in Derous et al. (2007): (1) *Macoma balthica* community; (2) *Abra alba* community; (3) *Nephtys cirrosa* community and (4) *Ophelia limacina* community [7], and some shipwreck areas, to capture their soundscapes (see Figure 2.3 for a map of these communities).

Each of the deployments consisted of 30 to 60 min of continuous recording, following the current by drifting, with the engines and the navigation plotter turned off, to reduce the noise created by the vessel itself. The typical drifting speed was 1 ms^{-1}, and the rope length was kept constant throughout one deployment. During each deployment, a GPS Garmin with a time resolution of 1s stored the location during the entire deployment, and was synchronized with the instrument clock. The length of the rope was chosen according to the depth, so that the hydrophone would be, on average, between the 1/2 and 1/3 of the water column from the surface (see Figure 2.5b). The rope length was kept constant during each entire deployment, and it was considered the instrument depth. The instruments used were a SoundTrap ST300HF (Ocean Instruments, Auckland, New Zealand) (sensitivity: -172.8 dB re 1 V/μPa—from now on, SoundTrap) and a Bruel and Kjaer Nexus 2690, with a hydrophone type 8104 (Bruel and Kjaer, Virum, Denmark) (sensitivity: -205 dB re1V/μPa—from now on, B&K) together with a DR-680 TASCAM recorder. The amplification in the Nexus was set to 10 mV, 3.16 mV or 1 mV, depending on the sound pressure level of the recording location. The SoundTrap was set to sample at 560 ksps, and the B&K at 192 ksps.

The recordings were made on 10 different days, yielding a total of 40 h 47 min of acoustic data. Of the 40 independent tracks recorded, 14 were recorded simultaneously with the two different devices (B&K and SoundTrap 300 HF), to test the similarity of the obtained recordings: this made for a total of 54 deployments (Figure 2.5). The deployment metadata (e.g. sampling rate, start and end time, rope depth...) were stored in the online repository of the European Tracking Network, under the Underwater Acoustics component (ETN, `https://www.lifewatch.be/etn` (accessed on 19 January 2023)).

Figure 2.5 **(a)** Data collected, colored according to instrument used. The background represents the bathymetry from the EMODnet bathymetry [10]. The solid black line represents the borders of the Belgian Exclusive Economic Zone; **(b)** Deployment scheme for two instruments simultaneously. The left instrument setup (blue rope) is the B&K hydrophone. The black squared box represents the Nexus amplifier and the DR-680 TASCAM recorder attached to the B&K system. The right instrument (red rope) is the SoundTrap.



(a) **(b)**

Table 2.3 Summary of all the field work days and the equipment used. RIB SS refers to the working boat from the RV Simon Stevin. Capoeira is the name of a sailing boat. ST stands for SoundTrap

Date	Vessel	Shipwrecks	ST [kSs^{-1}]	B&K [kSs^{-1}]
27/04/2020	RIB Zeekat	Coast, Loreley, Nautica Ena, Noordster, Paragon, Renilde	576	NA
29/04/2020	RIB Zeekat	Heinkel 111, HMS Colsay, Lola, Westerbroek	576	192
05/05/2020	RIB SS	Buitenratel, Killmore, Westhinder	NA	192
13/07/2020	RIB SS	Faulbaums, Grafton	NA	192
06/05/2020	RIB SS	Belwind, CPower	NA	192
07/05/2020	RIB SS	Garden City	NA	192
08/05/2020	RIB SS	G88, Loreley, Nautica Ena, Paragon	NA	192
1-3/06/2020	Capoeira	Birkenfels, Buitenratel, Gardencity, Grafton, Nautica Ena, VG, Westhinder, WK8	NA	192
12-13/10/2020	Capoeira	Birkenfels, Buitenratel, Faulbaums, Garden city, Grafton, Nautica Ena, VG, Westhinder, WK8	NA	192
9-10/06/2021	Capoeira	Birkenfels, Buitenratel, Faulbaums, Garden City, Grafton, Nautica Ena, Noordster, VG2, Westhinder, WK8	576	192

2.4 From wav to μPa

Converting wav files to μPa should be a straightforward operation. However, due to the different nature of the recorders which were used, each of the recorders was processed in a different way. Some manufacturers provide single frequency calibration values while others provide frequency-dependent calibrations. Sometimes the sensitivity is given for the full system (end-to-end calibration) and sometimes separately for each component. Furthermore, a pistonphone can be used to check any calibration offset before and after each deployment. For this reason, we provide a small guide on how to convert the data from WAV files to absolute dB re 1 μPa^2.

For the RESEA recorders, the conversion is listed in equation 2.1, and for SoundTrap and B&K in equation 2.2.

$$A = 2.5x \frac{1}{G} \frac{1}{G_c} \tag{2.1}$$

$$A = x \tag{2.2}$$

Where,
A is the analogue value in Volts (V),
x is the value in the WAV file normalized to [-1, 1] (done automatically in R, Python[1] or MATLAB when loading a WAV file),
G the gain (for RESEA, depends on the recorder), and
G_c is the gain correction factor (for RESEA, depends on the recorder and the specific channel).

For the B&K system, the Nexus system offers different amplification values to be selected. These values provide the end to end sensitivity of the system. The available values are 30 mV, 10 mV, 3.16 mV, 1 mV, 316 μV, 100 μV, 31.6 μV, or 10 μV.

These values can be converted to sensitivity as specified in equation 2.3.

$$s = 10 \log_{10} \left(\left(\frac{g}{1E6} \right)^2 \right) \tag{2.3}$$

Where,
s is the sensitivity of the hydrophone, in dB/V re 1μPa^2 (negative value), and
g is the selected amplification in nexus, in Volts (V).

The analogue value can then be converted to μPa (see equation 2.5). However, to do that the sensitivity first needs to be converted to linear values instead of dB, which is usually the value given by the manufacturer (see equation 2.4).

$$s_{lin} = 10^{s/20} \tag{2.4}$$

[1] Only when using specific audio packages such as librosa or soundfile

Where, s_{lin} is the sensitivity of the hydrophone, in μPa/V.

$$p = \frac{A}{s_{lin}} \tag{2.5}$$

Where, p is the pressure in μPa.

Once the data are converted to μPa, they are commonly converted to dB (see equation 2.6).

$$SPL = 20\log_{10}\left(\frac{p}{p_0}\right) \tag{2.6}$$

Where,
SPL is the sound pressure level in dB re 1 μPa^2, and
p_0 is 1 μPa.

Alternatively, if one is not interested in the pressure values in linear scale (μPa) but only in dB, it is also possible to compute the sound pressure levels in db without needing to convert the sensitivity to linear scale first, as specified in equation 2.7.

$$SPL = 20\log_{10}(A) - s \tag{2.7}$$

2.5 Conclusions

Two datasets have been collected, of which the LifeWatch Broadband Acoustic Network (stationary dataset) is still ongoing. Currently obtained data spans more than 3 years. During this dissertation we present further analysis done on the collected data, and provide insights on the soundscape of the BPNS. In Chapter 3 we further expand on the particularities of the BPNS when studying it acoustically.

Collecting the sound data involved the deployment of specialized equipment in the field, which led to the start of a semi-continuous recording network in the BPNS [1]. All the data collected during this PhD thesis are open and available on demand (due to the impossibility of storing it all due the size). Data products resulting from the following chapters of this dissertation have been published as separate open datasets so they can be reused in the future. This contributes to the global effort made by the IQOE initiative, and all the performed deployments have been added to the IQOE database of Existing Observing Systems.

During this thesis, a multi-purpose mooring system was developed at the Flanders Marine Institute (VLIZ) to place sensors (including acoustic recorders and hydrophones) on the sea floor, without leaving anything behind once the measurement is done. The mooring was designed to be modular so it could be used for multiple purposes and different sensors. This objective has been achieved, as it has been used in several national and international projects such as Apelafico, NWO, LifeWatch, Bar-Pelfish, Marco-Bolo (EU Horizon), DTO-bioflow, Coastbusters, Coastbusters 2.0, IG-Waves, and Outflow.

References

[1] C. Parcerisas, D. Botteldooren, P. Devos, and E. Debusschere. *Broadband Acoustic Network Dataset*, 2021.

[2] J. Goossens, M. T'Jampens, K. Deneudt, and J. Reubens. *Mooring Scientific Instruments on the Seabed—Design, Deployment Protocol and Performance of a Recoverable Frame for Acoustic Receivers*. Methods in Ecology and Evolution, 11(8):974–979, 2020.

[3] E. Debusschere, F. Hernandez, and K. Deneudt. *Cpod-Lifewatch: Permanent Cetacean Passive Acoustic Sensor Network in the Belgian Part of the North Sea*, 2015.

[4] J. Reubens, F. Hernandez, and K. Deneudt. *LifeWatch Observatory - Permanent Acoustic Receiver Network in the Belgian Part of the North Sea*, 2017.

[5] J. Haelters and F. Kerckhof. *Zeezoogdieren En Zeeschildpadden in België in 2023 [Marine Mammals and Sea Turtles in Belgium in 2023]*. Technical report, Instituut voor Natuurwetenschappen (KBIN), Brussels, 2024.

[6] EMODnet Human Activities. *Vessel Density*, 2022.

[7] S. Derous, M. Vincx, S. Degraer, K. Deneudt, P. Deckers, D. Cuvelier, J. Mees, W. Courtens, E. W. M. Stienen, H. Hillewaert, K. Hostens, I. Moulaert, V. V. Lancker, and E. Verfaillie. *A Biological Valuation Map for the Belgian Part of the North Sea*. Global Change, page 162, 2007.

[8] EMODnet Seabed Habitats. *Seabed Habitats*, 2018.

[9] A. Lillis, F. Caruso, T. A. Mooney, J. Llopiz, D. Bohnenstiehl, and D. B. Eggleston. *Drifting Hydrophones as an Ecologically Meaningful Approach to Underwater Soundscape Measurement in Coastal Benthic Habitats*. Journal of Ecoacoustics, 2(February):STBDH1, 2018.

[10] EMODnet Bathymetry Consortium. *EMODnet Digital Bathymetry (DTM 2018)*, 2018.

3

The data and its particularities

This chapter is based on the contents of the publication:

- *Parcerisas, C., Botteldooren, D., Devos, P., Hamard, Q., Debusschere, E. (2023). Studying the Soundscape of Shallow and Heavy Used Marine Areas: Belgian Part of the North Sea. In: Popper, A.N., Sisneros, J., Hawkins, A.D., Thomsen, F. (eds) The Effects of Noise on Aquatic Life. Springer, Cham.* https://doi.org/10.1007/978-3-031-10417-6_122-1

This publication has been reproduced with permission from Springer Nature. The introduction of this chapter has been adapted to improve the flow of this dissertation. The original section "Common Soundscape Analysis Techniques and Their Applicability in Shallow Anthropogenic Marine Area" has been moved to Chapter 4. Table 3.1 has been updated with the current state of knowledge about biological sounds in the BPNS.

3.1 Overview

Analyzing soundscapes in shallow and heavily exploited marine areas poses several challenges and particularities. For this reason, we made the implicit knowledge of the challenges more formal and explicit to set the scope for further developments. Bio-fouling, flow-noise, unknown sound sources, and masking compromise propagation and sound reception. Section 3.3 provides a list of known (possible) underwater sound sources in the BPNS. Section 3.4 is an overview of the inherent challenges to measure and analyze the soundscape in the BPNS. This chapter is a ground work to be able to answer all the research questions presented in this dissertation.

3.2 Introduction

The BPNS (Figure 3.1) is one of the most exploited marine areas in the world. In a relatively small area, activities like shipping, green energy, fishery and sand extraction converge. Every year there are more than 300,000 shipping movements, ranging from commercial ships to pleasure boats. Belgium is the country with the largest percentage of offshore renewable space in the world [1]. The BPNS only reaches up to 45 nautical miles offshore, and it is very shallow, with an average depth of 20 m, and a maximum of 45 m. The BPNS is characterized by a unique subtidal sandbank system, which is very dynamic and has strong tidal currents [2]. The sandbank system is subdivided in shallow sandbanks, sand mason worm banks and gravel beds. These unique shallow sandbanks exhibit high biodiversity, like the hotspots formed by gravel beds and sand mason worm banks [3]. There are approximately 140 species of fish, and more than 60 species of seabirds in the BPNS' coastal waters. Marine mammals, such as porpoises (*Phocoena phocoena*) and seals (*Phoca vitulina*), are becoming increasingly common. Moreover, there are more than 300 shipwrecks. Shipwrecks, which are artificial hard substrates in these sandbank systems, attract various species [4]. Wrecks have great ecological, cultural and economic value; they serve as important shelters and nurseries for many different marine species, and are very popular with wreck divers and anglers [5].

Because of the small spatial variation of the soundscape in the BPNS due to propagation patterns, the fauna occupying the different habitats is expected to contribute to the soundscape, and so is the fauna living in the multiple shipwrecks present in the BPNS. The different types of substrate influence the sound propagation, and thereby the soundscape signature as well. In addition to the spatial soundscape variation, seasonal, latitudinal and celestial factors will have an effect on the presence of particular sounds [6, 7]. Some sounds present cyclical patterns, or repeat at regular intervals, while others occur at random times. These patterns can be found in biophony, anthropophony or geophony. This time-dependent occurrence can be either short transient signals that occur over seconds or minutes, or a continuous presence in the soundscape over hours or days, resulting in chronic contributions to the soundscape [8].

Scientific research is important for monitoring the state of the environment, and that is one of the reasons why the BPNS is one of the most studied sea areas in the world. Derous et al. (2007) [9] created a marine biological valuation map, in order to support policymakers in objectively allocating the various user functions at the

Figure 3.1 Location of the BPNS in Europe (marked in red). Map made by maps@vliz.

BPNS, which classifies the different benthic communities into several groups. Even though the BPNS is a well-studied area, very little research about its soundscapes has been done.

3.3 Sound sources in the BPNS

3.3.1 Biophony

Biophony comprises all sounds produced by non-human organisms. Sound is an important source of information for many marine animals. For this reason, all vertebrates and many invertebrates are sensitive to sound [10].

The BPNS can be characterized by a positive east-west gradient of species richness and densities in coastal waters [11]. Furthermore, a clear coastal-offshore pattern is visible in the plankton communities [12]. In the BPNS, there are four main benthic communities linked to different seabed habitats. From these habitats, the sandbanks cover more than 80% of the area, and they consist of gullies, flanks and a sandbank top. Each part is home to one of the benthic communities. The highest biodiversity is found on the flanks and in the gullies. Further out at sea, there are also gravel beds, which consist of coarse sand and gravel bigger than 2mm. They have an important function as a breeding ground and nursery for some species, but the current species composition of the gravel beds has been impoverished by frequent fishing [13]. There used to be banks of the native flat oyster, which create their own micro-habitat just like the sand mason worm. However, these were fished away in just a few years from the end of the nineteenth century. Additionally, shipwrecks are artificial hard structures that are hotspots for biodiversity. During a study of ten shipwrecks in the BPNS, more than 200 animal species were observed [5].

There are 140 different species of fish listed in the BPNS, and the largest group is formed by species living close to the sea bottom (demersal species). The many sandbanks provide shelter and breeding ground for many species. All studied fish have been shown to be sensitive to sound [14], although not all species use sound for communication. Nevertheless, sound production is important for territorial defense, reproduction, and feeding for many species of fish. In a recent review about fish sound production, [15] identified 989 actively soniferous fish species, evenly distributed among all the geographical areas of the world. However, this represents a substantial underestimate of global actively soniferous fish diversity as over 96% of fish species lack published examination. The active sounds they produce are usually in the lower frequencies (<1 kHz), and are often described as ”grunts”, ”growls”, ”thumps”, ”knocks”, among others [16]. In addition to active sounds, it is probable that all fish produce passive sounds, such as feeding sounds, but it still remains unknown which ones do and a recognizable signature of such sounds is rare to find. Due to the importance of sound for many species of fish, masking of their signals can have a large impact on their fitness [17]. Despite the importance of fish sounds, the field of fish bioacoustics has historically been less developed than for taxa higher in the food chain [15]. Moreover, the studies on fish sound production are biased towards fish species from areas where visual ground-truthing is possible. For these reasons the FishSounds database [18] was recently published; an online open database where all the published fish sounds have been collected.

Other source of biophony in the BPNS are marine invertebrates, which can produce sounds for a variety of behaviors. They can generate the sound unintentionally or deliberately for communication. The capacity to produce sounds is known in only three groups of marine invertebrates: bivalves, echinoderms and crustaceans [19]. From the crustacean group, only barnacles (Cirripeda) and decapods (Decapoda) have been observed to produce sound. However, many species have been described to be sensitive to sound. They are the group of animals living in the BPNS less studied in terms of sound production and reception, and there exist very few automated detectors focused on sound produced by marine invertebrates.

Finally, marine mammals also contribute to the BPNS soundscape. The most common marine mammal in the BPNS is the harbor porpoise (*Phocoena phocoena*), which is using echolocation to forage, navigate and communicate. These echolocation clicks are very stereotypical and concentrated around 132 kHz, making them a perfect target to study using passive acoustics [20, 21, 22]. Harbor porpoises are present throughout the year, but are found in greater abundance between January and April. Other vocalizing cetaceans are the bottlenose dolphin (*Tursiops truncatus*) and the white-beaked dolphin *Lagenorbynchus albirostris*, which are also residents in Belgian waters. Furthermore, less vocal marine mammals are the grey and common seal (*Halichoerus grypus* and *Phoca vitulina*, respectively).

In Table 3.1 a non-exhaustive list of the species of the BPNS that have been reported to produce sound is provided, together with the available information about the sound they produce.

Table 3.1: Non-exhaustive list of all the known soniferous species of the BPNS. Soniferous behavior extracted from Looby et al. (2023) [23] and WoRMS [24] (excluding species only seen very sporadically). Asterisks (*) indicate unknown.

			Sound Production			
Group	**Scientific name**	**Common name**	**Type**	**Bandwidth**	**Frequency peak**	**Comment**
Mammals [25, 26]	Halichoerus grypus	Gray seal	Calls	0.1-5kHz	0.1-3 kHz	6 different call types (breeding)
			Click	0-30 kHz	**	**
			Hiss	0-40 kHz	**	**
			Knock	≤16 kHz	≤10 kHz	**
	Lagenorbynchus albirostris	White-beaked dolphin	Click	≤325 kHz	**	**
			Squeals	**	8-12 kHz	**
	Phoca vitulina	Harbor seal	Growl	<100-400 Hz	<100-250 Hz	Bubbly
			Click		12-40 kHz	**
			Creak	0.7-4 kHz	0.7-2 kHz	**
			Grunt	<0.1-4 kHz	**	**
			Groan			
			Social	0.5-3.5 kHz	**	Pulse duration 0.019-0.4 s
			Roar	0.4-4 kHz	0.4-0.8 kHz	**
	Phocoena phocoena	Harbour porpoise	Click	2 kHz	**	Not echolocation
			Click	110-150 kHz	∼132 kHz	**
	Tursiops truncatus	Bottlenose dolphin	Click	110-140 kHz	**	Echolocation
			Bark			
			Mew			
			Grate	**	**	**
			Rasp			
			Yelp			
			Narrowband	<2 kHz	0.3-0.9 kHz	**
			Whistle	0.8-24 kHz	3.5-14.5 kHz	**

Table 3.1: (continued) Non-exhaustive list of all the known soniferous species of the BPNS. Soniferous behavior extracted from Looby et al. (2023) [23] and WoRMS [24] (excluding species only seen very sporadically). Asterisks (*) indicate unknown.

Group	Scientific name	Common name	Sound Production			
			Type	**Bandwidth**	**Frequency peak**	**Comment**
Fish [15]	Anguilla anguilla	European eel	Click	<10 kHz	**	**
			Knock	**	∼200 Hz	**
			Thump			
	Atherina boyeri	Big-scale sand smelt	Passive	**	**	Feeding noises
	Chelidonichthys cuculus	Red gurnard	Growl	<1500 Hz	**	**
	Chelidonichthys lucerna	Tub gurnard	Growl	<1500 Hz	**	**
	Chelon auratus	Golden grey mullet	Passive	**	**	Feeding noises
	Ciliata mustela	Five bearded rockling	Passive	**	**	Feeding noises
	Clupea harengus	Atlantic herring	Grunt	**	∼60 Hz	**
			Knock			
			Thump			
	Conger conger	European conger	Passive	**	**	Feeding noises
	Dasyatis pastinaca	Common stingray	Undescribed	**	**	**
	Engraulis encrasicolus	Anchovy	Undescribed	**	**	**
	Eutrigla gurnardus	Grey gurnard	Growl	100 -2000 Hz	∼100 Hz	**
			Grunt	100 -1500 Hz		
			Knock	100 -1000 Hz		
	Gadus morhua	Atlantic cod	Grunt	20-500 Hz	∼60 Hz	**
	Hippocampus guttulatus	Long-snouted seahorse	Click	<1500 Hz	**	**

Table 3.1: (continued) Non-exhaustive list of all the known soniferous species of the BPNS. Soniferous behavior extracted from Looby et al. (2023) [23] and WoRMS [24] (excluding species only seen very sporadically). Asterisks (*) indicate unknown.

Group	**Scientific name**	**Common name**	**Sound Production**			
			Type	**Bandwidth**	**Frequency peak**	**Comment**
Fish [15]	Hippocampus hippocampus	Short snouted seahorse	Click	<1500 Hz	**	**
	Melanogrammus aeglefinus	Haddock	Hum	<600 Hz	**	Multiple haddock will produce a rumble from 30 to 400 Hz.
			Knock			
	Merlangius merlangus	Whiting	Passive	**	**	Feeding noises
	Molva molva	Ling	Undescribed	**	**	**
	Myoxocephalus scorpius	Shorthorn sculpin	Groan	**	~76 Hz	Dull
			Growl			Sustained
	Pegusa lascaris	Sand sole	Passive	**	**	Feeding noises
	Petromyzon marinus	Sea lamprey	Undescribed	**	**	**
	Pollachius pollachius	Pollack	Grunt	<700 Hz	~100 Hz	**
	Pollachius virens	Saithe	Thump	<400 Hz	~10-60 Hz	**
	Pomatoschistus	Gobies	Drum	**	~150-190 Hz	**
			Thump	**	~83	**
	Scomber scombrus	Atlantic mackerel	Grunt	**	**	harsh
	Scophthalmus maximus	Turbot	Thump	**	~60 Hz	**
	Trachurus trachurus	Atlantic horse mackerel	Grunt	300-5000 Hz	**	**
	Trisopterus minutus	Poor cod	Passive	**	**	Feeding noises
	Zeus faber	John Dory	Bark	200-600 Hz	312 ± 10 Hz	**

Table 3.1: (continued) Non-exhaustive list of all the known soniferous species of the BPNS. Soniferous behavior extracted from Looby et al. (2023) [23] and WoRMS [24] (excluding species only seen very sporadically). Asterisks (*) indicate unknown.

Group	Scientific name	Common name	Sound Production			
			Type	Bandwidth	Frequency peak	Comment
Invertebrates [19, 27]	Cirripeda	Barnacle	Peak pulses	**	**	1–3 ms
	Homarus gammarus	European lobster	Buz	**	**	**
			Rattle			
	Pecten maximus	Great scallop	Cough	20-27 kHz	**	Clusters or isolated
	Maja brachydactyla	**	Type 1	2-50 kHz	7 kHz	Short pulse, 2 peak freqs.
			Type 2	2-18 kHz	4 kHz	Pulse series
			Type 3	0-5 kHz	5 kHz	Feeding noises
	Athanas nitescens	Hooded shrimp	Snapping	5->50 kHz	9 kHz	Two peak freqs.
	Crepidula fornicata	Common slipper shell	Passive	8->50 kHz	45 kHz	Moving sounds, friction between shells
	Mimachlamys varia	Variegated scallop	Passive	4->50 kHz	37 kHz	Transient sound when jumping
	Patella vulgata	Common limpet	Rasping	**	**	**
	Psammechinus miliaris	Green sea urchin	Passive	11 -> 50 kHz	47 kHz	Moving
				47 - >50 kHz	49 kHz	Jumping

3.3.2 Anthropophony

The most important source of anthropogenic sound in the BPNS is shipping. Farcas et al. (2020) [28] reported that the shipping sound exceeded the modeled ambient noise by 20 dB more than 50% of the time in most of the Belgian Exclusive Economic Zone (EEZ) and more than 10% in the rest of the area. They also reported that the annual median ship noise excess in the BPNS is more than 20 dB from 63 Hz to 4 kHz for the shallower waters of Sea. For the deeper areas, the sound in these bands attenuated clearly in frequencies above 250 Hz. This means that there are very few places free from shipping noise in the BPNS.

Although knowing the total sound level generated by shipping is crucial for policymakers, shipping sounds can be very different from each other. Shipping noise has most of its energy at low frequencies (50–150 Hz), but sound can be up to 10 kHz [29]. Furthermore, several characteristics of the ship (type, size, speed, maintenance state, operational settings, etc.) influence its source levels and frequency distribution [30]. Small recreational vessels have a different sound signature than large cargo vessels, and anchored ships contribute differently from moving ships.

Most studies concerning shipping sound use data from the Automatic Identification System (AIS) combined with models of ship sound production [30]. In the North Sea, several propagation models have been developed, in the context of the JOMOPANS project, and AIS-based sound level models have been developed for the Dutch part of the North Sea [31], and for the entire North Sea [32, 33]. AIS data are commonly used to predict ship intensities on large spatial and temporal scales. They have been proven to be a valid source of information on global sound levels produced by shipping noise, which is useful for policy and management decisions [28]. Nonetheless, small vessels without AIS transponders and vessels switching their transponders off should not be underestimated in terms of sound pollution. This is particularly important in shallow waters, where the low frequencies are not propagated as well and therefore shipping contribution could be concentrated in higher frequencies generated by smaller recreational boats [34]. In addition to the unintentionally generated noise, some shipping activity also produces some intentional sounds using sonar. Sonar is used to navigate, measure distances, detect objects, and communicate. They use frequencies higher than those of sound unintentionally generated by ships, typically between 2 and 400 kHz, depending on the application.

In addition to all the shipping lanes, Belgium is also the country with the highest percentage of renewable offshore designated areas in the world [1]. Wind turbines do not produce loud sounds during operations. However, because of the cumulative effect, sound production of large offshore wind farms should not be underestimated [35]. During the construction phase, pile driving can reach very loud levels [36]. These impulsive sounds appeared to be correlated with a large-scale avoidance of the construction zone by porpoises, which led the Belgian Federal Government to implement a maximum threshold of 185 dB re 1 μPa (sound pressure level, zero to peak) at 750 m from the source for pile driving events. Therefore, since 2017, offshore wind farm constructors have developed and applied sound mitigation measures which have made incremental progress in complying with the national threshold, and that are currently used during their construction phase to reduce the sound produced [37, 38]. These mitigation strategies include bubble curtains (single or double), Hydro Sound Damper, noise mitigation screen or cofferdam [39, 40].

Other sources of anthropogenic sound present in the BPNS are seismic sources such as sparkers and air-guns (frequency range 200 Hz to 2 kHz), sub-bottom profilers (frequency range 2 to 20 kHz) and underwater explosions, which are characterized by very loud impulsive sound sources which are only active for a short period of time. The underwater explosions can be spontaneous, from old war munition present at the sea floor, or controlled, during munition removal exercises. Furthermore, in Belgium there are also dredging and sand extraction activities, which are also audible in the soundscape and have a clear and recognizable signature. All these differences should be reflected in soundscape analysis for ecological purposes.

3.3.3 Geophony

Some of the geophony sound sources in the BPNS are related to current: strong currents around hard structures at closer distances or very loud sources further away (several kilometers) such as coastal surf and tidal river currents. In [41], they relate these distant sound sources with the tide. Together with particle motion measurements, they propose that these low-frequency sound sources of 10-20 Hz and 20-40 Hz are possibly river mouths from some of the main Belgian, Dutch, and English rivers (Scheldt Estuary, Rhine Delta, North Sea Canal, and Thames River).

Current is also the main agent for sediment transport, where seabed material is brought into suspension by currents [42]. The collision of particles generates high frequency sound, which is largely dependent on the grain size and speed of the particles during impact [43]. The theoretical peak frequency of the generated sound can be approximated according to the formula in equation 3.1 [43]:

$$f_{\text{peak}} = 0.15 \left\{ \frac{E}{\rho_s(1-\sigma^2)} \right\}^{0.4} \frac{U^{0.2}}{a}, \tag{3.1}$$

where, for glass spheres (approximation of sand grains),
$E = 7.10^{10}\text{Nm}^{-2}$ is the Young's modulus,
$\sigma = 0.2$ is the Poisson's ratio,
$\rho_s = 2500\text{kg m}^{-3}$ is the density,
U is the collision speed and a is the radius.

Therefore, considering a collision speed between 0.2 and 0.6 ms^{-1} [44] and a radius between 50 μm and 1 mm [45], the frequency of sound produced by sediment transport in the BPNS should be between 12 and 300 kHz.

Besides, current generates flow-noise, which is not actual sound which propagates but just pressure turbulence generated by the shape of the hydrophone and/or the anchoring system. In any case, even if it is not a sound wave, this turbulence is picked up by the hydrophone and has to be removed before analyzing the data. Flow-noise and the other sounds generated by the current dominate the lower frequencies of the BPNS soundscapes, below 100 Hz. An example of this effect can be seen in Figure 3.3a.

Another source of abiotic natural underwater sound is wind, which appears between 100 Hz and 25 kHz. The generated sound level of wind depends on its speed [46]. The higher frequency components come from oscillating and bursting bubbles and splashes created by breaking wave crests [42]. Bubbles generated by waves in shallow waters also contribute to the decrease in the propagation of other sound sources [47].

Finally, rain can also contribute significantly to underwater sound between 1 and 50 kHz. The sound is generated from the splashes of raindrops falling in the water. The generated sound depends on the drops size distribution, the wind speed and the rainfall rate. In the lower frequencies, lightning is a less common event, but it is very loud compared to other sources of natural noise such as wind or rain. The electric discharges in the atmosphere over water produces strong underwater sound pulses [25]. Earthquakes are also an important source of sound in other places of the world, but not so present in Belgian waters.

3.4 Particularities and challenges of shallow water

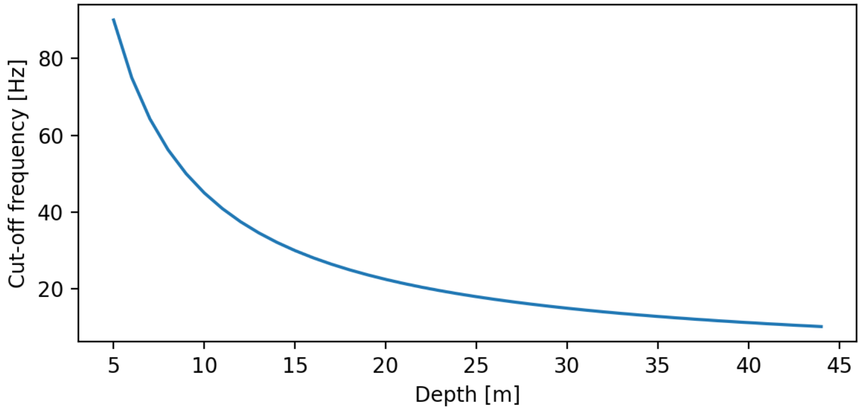
3.4.1 Propagation

As summarized in [48], in shallow water propagation is influenced by multiple reflections from the seabed and the sea surface, so geometrical spreading laws are not accurate enough. The resulting effect depends on the kind of source and its directivity. Several studies have focused on trying to understand the propagation patterns in shallow marine areas and simplify their computation [49, 31, 50].

In the BPNS the water mass is considered to be mixed, and therefore there are no big changes in salinity and temperature in the water column [51]. This makes the sound speed constant throughout the water column, and therefore isovelocity approximation could be sufficient for the shallow water propagation modeling [52].

An important aspect of shallow-water propagation is the presence of a waveguide cut-off frequency. Below this frequency, ducted propagation does not occur [53]. For shallow water, the cut-off frequency can be approximated by equation 3.2

Figure 3.2 Cut-off frequency values for an homogeneous sandy bottom for the equation provided in Ainslie (2010) [53] for all the depths of the BPNS



.

[53]:

$$f_{\text{cutoff}} = \frac{\pi - \rho_{\text{sed}}/\rho_{\text{w}}}{2\pi \sin \psi_{\text{c}}} \frac{c_{\text{w}}}{D}, \tag{3.2}$$

where,
f_{cutoff} is the cut-off frequency,
c_w is the velocity of the sound in water,
ψ_{c} is the critical angle,
ρ_{sed} is the sediment density,
ρ_{w} is the water density,
D depth in meters.

The cut-off frequency in the BPNS can be somewhere between 20 and 80 Hz (see Figure 3.2). Furthermore, because of the strong tides in the BPNS, the depth of shallower locations can practically double from low-tide to high-tide. Consequently, the local water depth has a clear influence in the propagation of low frequency farther sound sources [25].

On the other hand, mud is nearly acoustically transparent, and therefore an (theoretical) infinite layer of mud in the seafloor would not present such cut-off frequency. Therefore, to accurately predict the cut-off frequency in muddy bottoms it would be necessary to know the thickness of the mud layer. The importance of the seabed for shallow water propagation is illustrated in Ainslie (2010) [53], which shows how the propagation and the reflection loss between mud and sand sediments can differ more than 20 dB at a range of 10 km (and the difference increases with range).

Understanding the propagation in shallow waters not only allows us to understand better how can certain sounds be distorted but it can also provide an idea of how far can certain sounds reach. However, variability in acoustic propagation and differences in the source level and directivity of biological signals result in varying ranges over which sounds can be measured. This variability makes it challenging to estimate detection ranges [54], but knowing these detection ranges is crucial to understand which area is being monitored, especially when using sound event detection for species density estimation [55, 56]. For this reason, it is necessary to have sound propagation models to quantify variation in specific sound source detection ranges [8].

Even with accurate propagation models, it would be necessary to know the source level (SL) of a certain sound source (see ISO 18405:2017 for SL definition [57]) to estimate the area at which this source could potentially be recorded. The detection probability would then change depending on the ambient sound and the other sounds occurring which could potentially mask the detection. The "listening range" of a sensor will thus depend on the environmental parameters and the sound source. More detailed sound maps (including temporal variations) and propagation models featuring multiple sources would be necessary to get this information.

As an example, an exercise can be done for one source, at one specific frequency band, and considering one specific bathymetry profile. For example, a ship source level can be computed according to the model updated in the JOMOPANS project [30, 58] as a function of frequency (f), ship speed over ground (V), ship length (l) and ship type (C) (equations 3.3 and 3.4).

$$L_s(f, V, L, C) = L_{s0}(f) + 60\log_{10}(V/V_c) + 20\log_{10}(l/l_0) \tag{3.3}$$

$$L_{s0}(f, C) = 191 - 20\log_{10}(f_1) - 10\log_{10}(\left(1 - \frac{f}{f_1}\right)^2 + D^2) \tag{3.4}$$

Where,
L_s is the source level,
L_{s0} is the baseline spectrum for all the vessels,
V_c is the reference speed per vessel class,
l_0 is the reference lenght, 100 m,
D is 3 for all vessels except for cruise vessels, where it is 4, and
f_1 is $\frac{480}{V_c}$.

We will take as an example the 250 Hz in-band source level of a 32 m long fishing vessel sailing at 10 kn, and a container ship of 300 m long sailing at 18 kn. According to the model (see [58]), a container ship has a V_c of 18.0 kn, and a fishing vessel of 6.4 kn. Therefore, the fishing vessel would produce an in-band source level of 161 dB re 1 μPa^{2}m^2, and a container ship of 170 dB re 1 μPa^{2}m^2.

According to Küsel et al. (2019) [59], on a perfectly flat bathymetry of 50 m deep with an homogeneous sand bottom, the propagation loss (PL) for 250 Hz (for a broadband source at 15 m deep) at 30 km would be around 75 dB re 1 m^2. This would be a more or less realistic scenario at the offshore part of the BPNS. Considering an a sea state of 3-4 Beaufort, with no other ships (see Figure 4.1, in Chapter 4 for more information), the ambient sound would be around 85 dB re 1 μPa^2 (converted to power level at decidecade band 250 Hz). Therefore, the sound of the container ship would be received above the ambient sound (SNR $>$ 3 dB) everywhere in the sandy area ($>$ 30 km), while the fishing vessel only until 20 or 30 km.

As another example, Atlantic Cod is known to produce sound between 20 Hz and 600 Hz, with a frequency peak at $\sim$60 Hz [60], and at a (broadband) source level of 163.5 $\pm$ 7.9 dB re 1 μPa^{2}m^2, which can be approximated to a spectral level[1] of 135 dB dB re 1 μPa^{2}m$^2 Hz^{-1}$. Converting it back to the power level at the decidecade band of 250 Hz, this would lead to an in-band source level of 152 dB re 1 μPa^{2}m^2. This would only be heard at around 5 km away in ideal ship-free situation, as in the propagation curve from Küsel et al. (2012) [59].

However, the ship-free scenario is not realistic. When looking at actual sound levels (see Chapter 4, Figure 4.2 for details), the power level at the 250 Hz band is approximately of 97 dB re 1 μPa^2 when considering the median 50 % exceedance level (L50) for all the stations combined. Considering these data, the fishing vessel would only be received above ambient sound levels at around 5 km (SNR $>$ 3 dB), the container ship at around 7 to 10 km, and the cod at less than 2.5 km.

This exercise gets more complicated for animals vocalizing close to the sediment, as more research would be needed to be able to understand the interaction of sound propagation between the sediment and the water.

Finally, Sertlek et al. (2019) [61] showcases the influence of bathymetry changes in propagation in shallow waters. They analyze two cases, one flat bathymetry of 100 m and one bathymetry which starts at 100 m and has a steep slope to 30 m between 5 and 7 km. This could be a realistic scenario where a source would be situated outside the BPNS in a deeper area. Their results show that the propagation loss at 250 Hz is 10 dB higher for the second case, therefore the sound propagates less far.

3.4.2 Flow noise and bio-fouling

Another challenge of measuring underwater sound in the BPNS is flow-noise. The turbulence generated by the current around the hydrophone or the mooring system produces pressure changes that are not actual propagating sound. This pseudo-noise,

[1]This conversion was done considering a flat spectrum, which is unrealistic. Spectrum source levels were not available, so this was considered as a rough approximation.

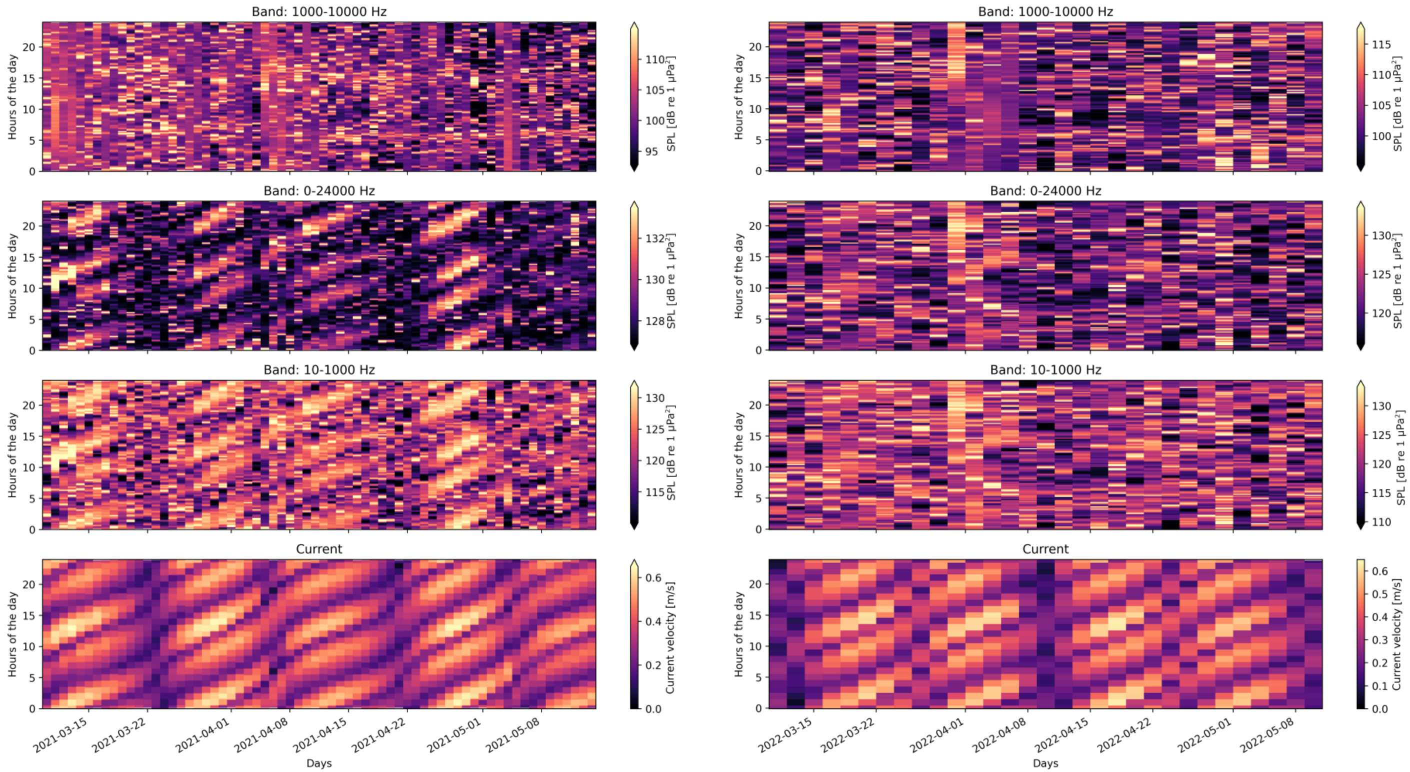
however, is picked up by the hydrophone anyway and can mask other sound in the lower frequencies below 1000 Hz, limiting the capability to reliably measure ambient noise [62, 63]. This turbulence depends on the current speed. In the North Sea, the tidal current is estimated to range from a few cms^{-1} to 67.7 cms^{-1} [44]. Furthermore, the flow noise recorded by the hydrophone depends on the shape and size of the hydrophone and/or the anchoring system.

When deploying bottom long-term observatories in the North Sea, the instruments will be exposed to bio-fouling. The growth of living organisms on top of the sensors can lead to a change in sensitivity [64]. The anchoring system is an artificial hard substrate that is sometimes rapidly colonized by many organisms, depending on the location and season. The fauna residing on top of the mooring therefore can affect the soundscape. Furthermore, the presence of bio-fouling on top of the hydrophone can lead to changes in sensitivity, particularly to the sensitivity of the hydrophone to generate and record flow noise. Moreover, the living organisms on top of the hydrophone can produce ”tapping” noises which lead to clipping of the hydrophone recordings.

We exemplify these particularities with data analysis from the stationary dataset presented in Section 2.2. For this example, 3 frequency bands were selected for analysis: broadband (up to 24,000Hz), low frequency (10-1,000Hz), and higher frequency (1,000-10,000Hz). For each of the band, the signal was filtered using a band-pass Butterworth filter of order 4, and data were decimated to twice the higher frequency limit. Next, the Root Mean Square Sound pressure level (SPL, entry 3.2.1.1 in ISO 18405:2017 [57]) value of the calibrated data were computed for each file, of equal duration (10 minutes and 15 seconds). For each processed point, the bottom current was retrieved from the COHERENS model for the North Sea [65] and assigned using the Python package bpnsdata [66].

This process was done for all the recording stations (see Figure 2.2). In some deployments, it could be observed that the low-frequency band was dominated by a source that is correlated with the current speed, which we assume is flow-noise (e.g. Figure 3.3a). However, this pattern did not repeat in other deployments (e.g. Figure 3.3b). This could be due to a different position of the anchoring system with respect to sand banks.

Figure 3.3 Root Mean Squared Sound pressure levels (SPL) in station Grafton in different bands (three top), and total current bottom speed (bottom). Station location can be seen in Figure 2.2 (Chapter 2). Deployments performed in **(a)** 2021 **(b)** 2022



(a) **(b)**

In the deployments where flow-noise was present, most of them seemed to show a decrease in correlation with the current after a certain time. The hypothesis was that the biofouling growing on top of the hydrophone would decrease the flownoise picked up by the hydrophone or its sensitivity. A linear mixed effects regression as described in Equation 3.5 was fitted to check whether the effect of the number of days deployed was significant. Linear mixed effects regression is a statistical method used to model relationships between a dependent variable and one or more independent variables while accounting for both fixed effects (variables of interest) and random effects (random variations from subjects or experimental units). For this particular case, the independent variables were the current and the number of days since deployment, and the random effect the deployment. Interaction between current and day since deployment was added. The dependent variable was the SPL from the band 10 to 1000 Hz. For this analysis, only deployments with a total correlation higher than 0.4 between the sound pressure in the low-frequency band and the current were included. The deployments with a low correlation with flow-noise were assumed to be in a spot where currents were less strong because of the sandbanks or the protection of a nearby shipwreck. The effect of the biofouling was not expected to change proportionally to the days but rather at a decreasing rate, as biofouling does not keep increasing infinitely but reaches a colonization plateau. For this reason, the variable expressing the days since deployment was converted to a logarithmic scale.

$$\texttt{SPL}_{\texttt{low\_freq}} \sim \texttt{current} + \texttt{d} + \texttt{current} : \texttt{d} + (1|\texttt{deployment}) \quad (3.5)$$

where, $\texttt{d}$ is the days since the deployment started (converted to log).

The results (see Table 3.2) showed that the days since deployment, the current and the interaction between the two were significant. The interaction and the days since deployment have a negative coefficient, which could be because the flow-noise is diminished due to bio-fouling (in time), and it has a bigger effect when the current is high than when it is low. However, this effect could also be due to sand burying increasing with time, change of depth because of the movement of the sandbanks, loss of sensitivity from the hydrophone for technical reasons, or other causes.

In conclusion, the recorded long-term sounds allow us to see clear current-related patterns in some of the deployments but not in others. These differences can be found also between deployments at the same location. This might be due to the ever-changing position of the sandbanks with respect to the hydrophone, both as a result of natural phenomena, such as sand waves and storms, or sediment extraction [67, 68]. It could also be due to differences in current strength or seasonality, or due to the orientation of the metallic frame. None of these hypotheses have been checked, and further analysis is necessary to understand the cause of these differences.

Table 3.2 Summary of the results of the fitted linear model. p-value of deployment: 0.13, not significant

Mixed Linear Model Regression Results				
	coef	std err	z	p-value
Intercept	116.60	3.72	31.37	0.000
days_since_deployment	-0.27	0.08	-3.44	0.0006
current	27.56	0.78	35.20	0.000
days_since_deployment:current	-2.72	0.21	-13.20	0.000

3.4.3 Unknown sound sources and low visibility

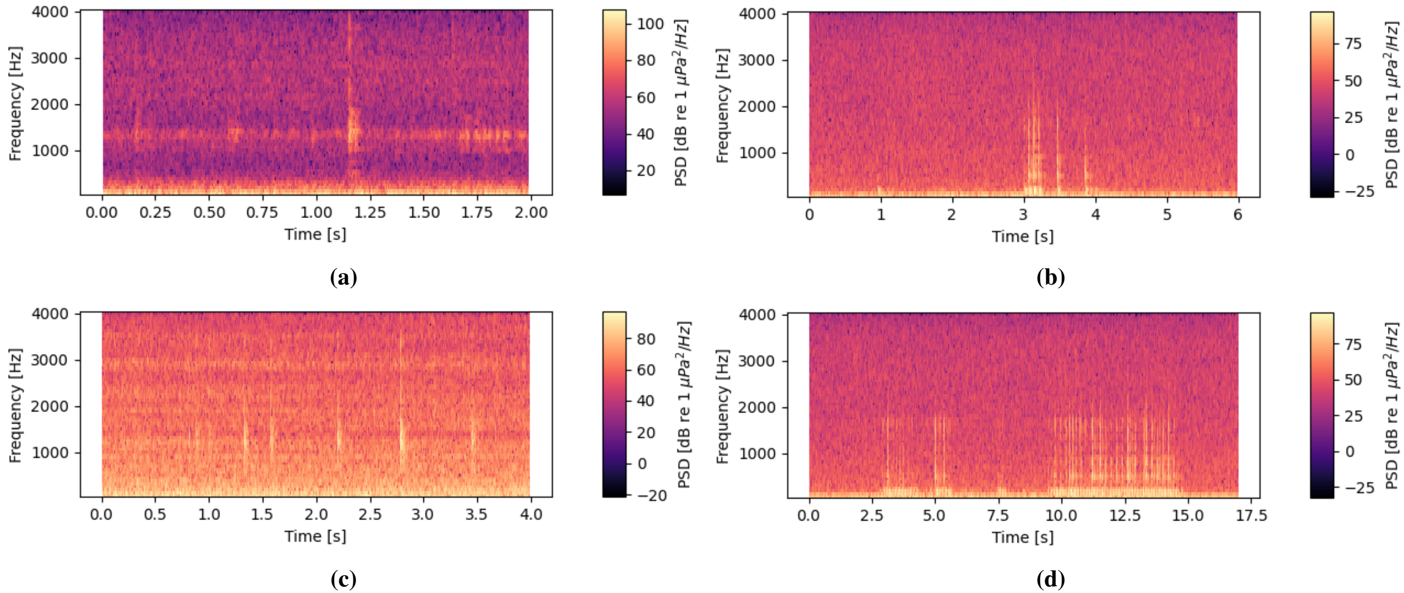
A common technique to link sound to its producer is visual correlation, both in the lab, in the wild with cameras deployed together with hydrophones or in the wild with visual surveys. Thanks to these efforts, all the marine mammals species are nowadays known to produce sound, and many of their signatures can now be recognized without the need for a visual confirmation.

However, many other biological sound sources are still unknown, particularly those of fish and invertebrates. The examination of fish and invertebrates for sound production can be complex, difficult, indecisive, and inefficient. In captivity, some animals present a different soniferous behavior than in the wild [15]. In the field, correlating sounds with their producers in areas with high turbidity, such as the BPNS [69], is often not possible due to low visibility.

Furthermore, in shallow waters the propagation patterns can be complex, so sound signatures can be distorted when received [48]. This complicates the recognition of certain sounds, as the received sound might not be consistent to those reported in literature.

As an illustration for these unidentified sound sources, some examples were manually selected. In Figure 3.4 the short spectrograms of these sounds can be found, illustrating sounds that could be from a biological source, but have not been identified at the moment. This topic is further developed in Part III. For more examples of sounds, please see the Supplementary Material, Section S3.1.

Figure 3.4 Several examples of unknown sound sources in the BPNS. For the spectrogram computation, the length of the Fast Fourier Transform (FFT) was chosen to bee 128, Hann window, with a 70% overlap, leading to a temporal resolution of 4.8 ms. The descriptions of the detected unknown sounds are **(a)** stridulation, possible source: crustacean **(b)** single “drum”, possible source: fish **(c)** consecutive “pocs”, possible source: unknown **(d)** continuous “drumroll”, possible source: fish



3.4.4 Masking

Auditory masking is defined in ISO 18405:2017 [57] as "auditory process by which the behavioural hearing threshold [3.7.2.1) or electrophysiological hearing threshold (3.7.2.2) for a sound [the signal) is raised by the presence of another sound [the noise)".

When interested in monitoring biodiversity using PAM, it is necessary to exclude almost constant anthropogenic noise to analyze biophony. This can be done by analyzing only quiet periods or with automatic denoising techniques. Shipping sounds have most of their energy from 50Hz to 10 kHz [29], which overlaps with most of the biological sounds produced by fish, invertebrates, seals and dolphins. This can lead to potential masking effect, where a louder sound source disrupts the reception of a weaker sound.

The sound produced by harbor porpoises is found at higher frequencies and is therefore not prone to be masked by human-made sounds. However, some sonar activities are in that frequency range, which are also detected by some click-detector algorithms such as the KERNO algorithm from the C-PODs (Chelonia Ltd, Penzance, UK), and could be a source of masking for these species.

Besides the impact that masking can have on marine fauna communication, it also complicates the soundscape analysis. Anthropogenic noise can decrease the ability of humans to detect certain sounds in short-term spectrograms. Furthermore, biological sounds can disappear in the general overview plots like Long Term Spectral Average (LTSA) or Spectral Probability Density (SPD) [70] because of their minor contribution to the soundscape compared to anthropogenic and geophonic sounds.

3.5 Discussion: collection and storage of underwater acoustic data in the BPNS

The effort necessary to record long-term underwater acoustic data should not be underestimated. To deploy an instrument in a fixed location, it is necessary to learn how to operate these instruments, and then design a suitable mooring system where the recorder can be attached to. For deep waters, these moorings usually consist of a weight at the seafloor with an acoustic release attached to several meters or kilometers of rope, held up by buoys. These systems, however, are not suitable for high-current, heavily used, shallow environments for several reasons. First, the high current generates noise due to the moving parts of the mooring. Second, such buoys need to be avoided by all ships, making it very complicated to find suitable locations where they do not interrupt human activities. Therefore, we recommend the use of the mooring designed at VLIZ to monitor underwater sound in shallow waters. The mooring design is open on request, so it can be accessed and build in other places. This mooring also can be used for deploying multiple sensors at the same time, and it has a modular approach, so sensors of any shape can be added and deployed simultaneously [71]. Deploying multiple instruments simultaneously can provide an extra dimension to the data analysis, as different measurements can be combined to be able to interpret the data from different perspectives [72, 73]. Another advantage is that, contrary to deep waters moorings, nothing is left on the sea floor after recovering it [74].

Deployments using a single hydrophone can be effective for projects such as recording acoustic habitats, measuring noise levels, and monitoring earthquake and geological activity, as demonstrated in this thesis. Single hydrophone recordings are suitable when the sound data does not need to be directional, meaning the direction and location of the sound source are not crucial for the project. However, even for projects that do not require source localization, we recommend deploying recorders with multiple hydrophones in shallow waters, if the budget allows. This approach has three advantages: first, it provides a backup in case one of the hydrophones fails to record; second, it allows for cross-correlation of the received signals to reduce flow noise [75, 63], third, it can be used to compute the direction of the sound (particle motion)[2]. For PAM projects which do require source localization, it is recommended to use an array of minimum 3 hydrophones, but preferably 4, in a tetrahedral position. Source localization was not in the scope of this dissertation, and therefore this has not been tested.

When starting a new underwater PAM network (or single deployment), deciding which underwater recorder to acquire is not a straightforward task, as the answer will depend on the research question at hand and the available budget. Some things to consider are memory capacity, hydrophone sensitivity, battery life, available sampling rates, and if multiple channels are necessary. In regions with high biofouling rates, such as the Belgian Part of the North Sea (BPNS), it is crucial to consider whether the hydrophones are constructed from materials that resist biofouling at their seams. From experience, biofouling (barnacles) can compromise molded silicone seals. Consequently, we do not recommend using hydrophones with such seals for monitoring in these environments. Even for hydrophones designed with biofouling-resistant materials, it is advisable to apply electrical tape at the bottom of the hydrophone to protect any seams, leaving only the tip of the hydrophone exposed. This method also provides effective protection for preamplifiers and cables.

When deploying acoustic recorders, other parameters need to be decided, such as the sampling rate and duty cycle. These will depend on the research question and the endurance of the recorder (battery/memory duration) at the desired sampling frequency. Higher sampling rates will fill the memory faster and use more power, potentially limiting the duration. If the calculated endurance using the desired

[2] Only possible if more than three hydrophones are recording simultaneously

sampling frequency is too short, recording duration can always be extended by using a duty cycle. Selecting an appropriate duty cycle also depends on the research question, but for analyzing data of areas which have not been studied acoustically before, we would recommend using a 24h-on 24h-off cycle. This allows detecting rare sounds which happen daily and/or seasonally, which could be otherwise missed when recording, for example, 30 minutes every 60 minutes. This is of particular importance in areas with sparse-occurring sounds, especially when twilight is very short (less than 20 minutes [76]). On the other hand, 24h-on 24h-off duty cycles can also miss very rare events.

Long-term recordings are best done at fixed locations, and they can be used to assess long-term changes. Autonomous recorders present several disadvantages with respect to fixed cabled stations. First, that data are accessible from the moment the recording is done. Secondly, malfunctions can be monitored and potentially corrected during recording, although immediate resolution may not always be feasible. Recording at a fixed location is however not the only option when recording a certain soundscape. Drifting recordings can be used to capture ”snapshots” of soundscapes. These recordings are a lot cheaper to deploy, as they can be done from small boats. Furthermore, usually the amount of data to manage and analyze afterwards is less, so if data storage is a constraint this can be a good option. These soundscape snapshots cannot be used to asses long-term changes and patterns. However, they can be used to assess the spatial variability of the soundscape in a certain habitat, which is of particular interest in shallow soundscapes. For example, in Lillis et al. (2018) [77], they use drifting recordings as a meaningful approach to underwater soundscape measurement in coastal benthic habitats. These recordings allow them to see how the soundscape changes when they approach an oyster reef. Therefore, drifting provides acoustic sampling at spatial scales that might typically be overlooked when using stationary hydrophone methods. Mobile and fixed hydrophone recording methods are complementary approaches, and can be used together to unravel different aspects of the soundscape and its change [77].

Moving recordings can also be done from robots with a silent propeller such as gliders or some Uncrewed Surface Vehicles (USV). These approaches present the same advantages than drift recordings regarding the information they can provide, but they have the extra advantage that they can be deployed for months. In this case, onboard processing might be of interest to obtain near-real-time information. In the particular case of gliders, the sound is also recorded at different depths, which can provide extra insight on how the sound changes in the water column [78, 79, 80, 81].

Once the instruments are deployed and retrieved, data and metadata should be stored following FAIR principles. There are several international initiatives currently working on the standardization of deployments metadata and data structure to make long-term underwater acoustic data FAIR. Some of them are the IQOE

Standardization working group or the IQOE Data Management and Access. Some projects are presenting their own recommendations, such as Tethys, SoundCoop, JOMOPANS, or OPUS. Even though the standardization is advancing and several guidelines are available, it is still hard for researchers to decide which guidelines to follow when starting a new project.

Regarding data cleaning, it is common to have artifacts when recording such as flow-noise, electronic noise from the instrument, or sounds produced by the anchoring system. Detecting and removing these artifacts is a time-consuming task when analyzing long-term data. In areas with high currents, it is very likely that flow noise will disturb the recordings. There is currently no solution which deals with this problem entirely, but some measures can be taken. First, depending on the research question, a valid point can be to only consider slack tide periods. This is valid to asses background levels, or if there is interest in finding a specific sound source but its temporal patterns do not matter. Another option is to only consider frequencies higher than the ones expected to be influenced by flow noise. This approach will be valid if the sounds of interest are not at lower frequencies. Theoretically, flow noise could also be reduced directly while recording. The Cetacean Research Technology company (CRT, USA) offers similar cages for flow noise reduction. In Weil et al. (2021) [82] they tested different hydrophone cages with simulations and in an anechoic tank, and presented one design with serrated trailing edges which was more successful at reducing flow noise. However, to the author's knowledge, these are not commercialized yet. For the data collected for this dissertation, we used similar cages to the ones from Cetacean Research Technology developed at VLIZ. However, testing the improvement when using a serrated cage remains as future work. Flow noise can also be reduced during post-processing if the recording is made with several hydrophones, as already mentioned, by cross-correlating the received signals [75, 63]. However, this approach can be costly due to the increased number of hydrophones needed and the increase in memory use. Furthermore, it also risks removing parts of the signal of interest [83, 63]. In this dissertation, due to the multiple problems encountered with the equipment, most of the obtained data ultimately contained only one (valid) channel. Finally, flow noise is limited if the recording system is drifting [84], because the hydrophone moves with the water column. Therefore, drifting can be chosen as an acoustic recording strategy, which allows for a higher spatial coverage but a lower temporal one, but it is not applicable for long-term monitoring stations.

When controlling the collected data for quality, we found that high levels of bio-fouling could corrupt the data. Bio-fouling produces short loud signals which alter the captured data by the hydrophone. These are not sound but pseudo-noise, coming from the direct contact between animals or moving particles and the hydrophone (bio-abrasion) [85]. These contacts generate extreme broadband transient sounds. When these sounds are not occurring constantly, using a 1-minute median instead of averaging would reduce the impact they have on the sound pressure levels [85]. However, after several months in the water, these events can become nearly constant, leading to unusable data. An interesting fact after analyzing the data is that by checking how the correlation between flow-noise and recorded low-frequency sound levels evolved over time, we could conclude that flow-noise influence is reduced after several months of deployment, either because of the bio-fouling itself reducing flow-noise or because of a drop in sensitivity from the hydrophone due to the bio-fouling. Further work is necessary to understand the underlying cause of it.

Once data are collected, stored, and quality-controlled, manual scrolling of those data is necessary before proceeding to any automatic analysis. In some cases, this will include annotating sound events. It is however not possible to know which species produces which sound without access to a library of reference sounds. Recent international efforts are leading to a more efficient way to access this knowledge. Initiatives like the Global Library of Underwater Biological Sounds (GLUBS) [86] or FishSounds [18] aim to create collective and global open-access platforms that provide reference libraries of known and unknown biological sound sources. Furthermore, a recent publication from GLUBS has created a Global inventory of species categorized by known underwater sonifery [23], linked to the WoRMS database register [24]. This effort collects all the publications to date which have assigned sound to species. The centralization of knowledge about sound-producing species is key to the advancement of marine bioacoustics.

3.5.1 Conclusions

Recording in an environment such as the BPNS is challenging. We have demonstrated to be aware of this challenging environment and the difficulties it brings to collect data and interpret it. These include high currents, biofouling, complex propagation patterns, unknown sound sources and masking. We have made this knowledge explicit as a basis for future underwater acoustics work in the BPNS.

Acknowledgments

The project was realized as part of the Flemish contribution to LifeWatch. The authors thank the Simon Stevin RV crew for their help with the deployment and retrieval of the acoustic instruments. We want to thank Bram Conings for sharing his knowledge of the species present in the BPNS with us, and Özkan Sertlek for reviewing the propagation part.

References

[1] Belgian Federal Public Service (FPS) Health, Food. *Something Is Moving at Sea. The Marine Spatial Plan for 2020-2026*, February 2020.

[2] F. Kerckhof and J.-S. Houziaux. *Biodiversity of the Belgian Marine Areas*. Technical report, Koninklijk Belgisch Instiuut voor Natuurwetenschappen, Brussels, April 2023.

[3] M. Vincx, W. Bonne, A. Cattrijsse, S. Degraer, A. Dewicke, M. Steyaert, J. Vanaverbeke, G. Van Hoey, E. Stienen, J. Waeyenberge, E. Kuijken, P. Meire, H. Offringa, J. Seys, F. Volckaert, E. Gysels, T. Huyse, M. Zietara, A. Geets, and B. Hellemans. *Structural and Functional Biodiversity of North Sea Ecosystems: Species and Their Habitats as Indicators for a Sustainable Development of the Belgian Continental Shelf*. Technical report, Belgian Federal Office for Scientific, Technical and Cultural Affairs (OSTC), 2004.

[4] Belgium Health, Food Chain Safety and Environment. *Biodiversity of the North Sea*. https://www.health.belgium.be/en/biodiversity-north-sea, January 2016.

[5] V. Zintzen and C. Massin. *Artificial Hard Substrata from the Belgian Part of the North Sea and Their Influence on the Distributional Range of Species*. Belgian Journal of Zoology, 140(1):20–29, 2010.

[6] E. Staaterman, C. B. Paris, H. A. DeFerrari, D. A. Mann, A. N. Rice, and E. K. D'Alessandro. *Celestial Patterns in Marine Soundscapes*. Marine Ecology Progress Series, 508:17–32, August 2014.

[7] C. A. Radford, A. G. Jeffs, C. T. Tindle, and J. C. Montgomery. *Temporal Patterns in Ambient Noise of Biological Origin from a Shallow Water Temperate Reef*. Oecologia, 156(4):921–929, July 2008.

[8] M. F. McKenna, S. Baumann-Pickering, A. C. M. Kok, W. K. Oestreich, J. D. Adams, J. Barkowski, K. M. Fristrup, J. A. Goldbogen, J. Joseph, E. B. Kim, A. Kügler, M. O. Lammers, T. Margolina, L. E. Peavey Reeves, T. J. Rowell, J. A. Stanley, A. K. Stimpert, E. J. Zang, B. L. Southall, C. C. Wall, S. V. Parijs, and L. T. Hatch. *Advancing the Interpretation of Shallow Water Marine Soundscapes*. Frontiers in Marine Science, 8, 2021.

[9] S. Derous, M. Vincx, S. Degraer, K. Deneudt, P. Deckers, D. Cuvelier, J. Mees, W. Courtens, E. W. M. Stienen, H. Hillewaert, K. Hostens, I. Moulaert, V. V. Lancker, and E. Verfaillie. *A Biological Valuation Map for the Belgian Part of the North Sea*. Global Change, page 162, 2007.

[10] C. M. Duarte, L. Chapuis, S. P. Collin, D. P. Costa, R. P. Devassy, V. M. Eguiluz, C. Erbe, T. A. C. Gordon, B. S. Halpern, H. R. Harding, M. N. Havlik, M. Meekan, N. D. Merchant, J. L. Miksis-Olds, M. Parsons, M. Predragovic, A. N. Radford, C. A. Radford, S. D. Simpson, H. Slabbekoorn, E. Staaterman, I. C. Van Opzeeland, J. Winderen, X. Zhang, and F. Juanes. *The Soundscape of the Anthropocene Ocean*. Science, 371(6529):eaba4658, February 2021.

[11] A. Cattrijsse and M. Vincx. *Biodiversity of the Benthos and the Avifauna of the Belgian Coastal Waters : Summary of Data Collected between 1970 and 1998*. Federale Office for Scientific, Technical and Cultural Affairs,, Brussels : Federale Office for Scientific, 1999.

[12] A. Aubert, O. Beauchard, R. de Blok, L. F. Artigas, K. Sabbe, W. Vyverman, L. A. Martínez, K. Deneudt, A. Louchart, J. Mortelmans, M. Rijkeboer, and E. Debusschere. *From Bacteria to Zooplankton: An Integrative Approach Revealing Regional Spatial Patterns During the Spring Phytoplankton Bloom in the Southern Bight of the North Sea*. Frontiers in Marine Science, 9, 2022.

[13] J.-S. Houziaux, M. Fettweis, F. Francken, and V. Van Lancker. *Historic (1900) Seafloor Composition in the Belgian–Dutch Part of the North Sea: A Reconstruction Based on Calibrated Visual Sediment Descriptions*. Continental Shelf Research, 31(10):1043–1056, July 2011.

[14] A. N. Popper, A. D. Hawkins, O. Sand, and J. A. Sisneros. *Examining the Hearing Abilities of Fishes*. The Journal of the Acoustical Society of America, 146(2):948–955, August 2019.

[15] A. Looby, K. Cox, S. Bravo, R. Rountree, F. Juanes, L. K. Reynolds, and C. W. Martin. *A Quantitative Inventory of Global Soniferous Fish Diversity*. Reviews in Fish Biology and Fisheries, 32(2):581–595, June 2022.

[16] M. C. Amorim. *Diversity of Sound Production in Fish*. Communication in Fishes, 1:71–105, January 2006.

[17] A. N. Popper and A. D. Hawkins. *An Overview of Fish Bioacoustics and the Impacts of Anthropogenic Sounds on Fishes*. Journal of Fish Biology, 94(5):692–713, 2019.

[18] A. Looby, S. Vela, K. Cox, A. Riera, S. Bravo, H. L. Davies, R. Rountree, L. K. Reynolds, C. W. Martin, S. Matwin, and F. Juanes. *FishSounds Version 1.0: A Website for the Compilation of Fish Sound Production Information and Recordings*. Ecological Informatics, 74:101953, May 2023.

[19] M. Solé, K. Kaifu, T. A. Mooney, S. L. Nedelec, F. Olivier, A. N. Radford, M. Vazzana, M. A. Wale, J. M. Semmens, S. D. Simpson, G. Buscaino, A. Hawkins, N. Aguilar de Soto, T. Akamatsu, L. Chauvaud, R. D. Day, Q. Fitzgibbon, R. D. McCauley, and M. André. *Marine Invertebrates and Noise*. Frontiers in Marine Science, 10, 2023.

[20] D. Gillespie, P. Berggren, S. Brown, I. Kuklik, C. Lacey, T. Lewis, J. Matthews, R. McLanaghan, A. Moscrop, and N. Tregenza. *Relative Abundance of Harbour Porpoises (Phocoena Phocoena) from Acoustic and Visual Surveys of the Baltic Sea and Adjacent Waters during 2001 and 2002*. Journal of Cetacean Research and Management, 7:51–57, January 2005.

[21] V. L. G. Todd, W. D. Pearse, N. C. Tregenza, P. A. Lepper, and I. B. Todd. *Diel Echolocation Activity of Harbour Porpoises (Phocoena Phocoena) around North Sea Offshore Gas Installations*. ICES Journal of Marine Science, 66(4):734–745, May 2009.

[22] U. K. Verfuß, L. A. Miller, and H.-U. Schnitzler. *Spatial Orientation in Echolocating Harbour Porpoises (Phocoena Phocoena)*. Journal of Experimental Biology, 208(17):3385–3394, September 2005.

[23] A. Looby, C. Erbe, S. Bravo, K. Cox, H. L. Davies, L. Di Iorio, Y. Jézéquel, F. Juanes, C. W. Martin, T. A. Mooney, C. Radford, L. K. Reynolds, A. N. Rice, A. Riera, R. Rountree, B. Spriel, J. Stanley, S. Vela, and M. J. G. Parsons. *Global Inventory of Species Categorized by Known Underwater Sonifery*. Scientific Data, 10(1):892, December 2023.

[24] WoRMS Editorial Board. *World Register of Marine Species*, 2024.

[25] M. A. Ainslie, C. a. F. de Jong, H. S. Dol, G. Blacquière, and C. Marasini. *Assessment of Natural and Anthropogenic Sound Sources and Acoustic Propagation in the North Sea*. UNCLASSIFIED TNO-DV 2009 C085, The Netherlands Ministry of Transport, Public Works and Water Affairs; Directorate-General for Water Affairs, February 2009.

[26] W. J. Richardson, C. R. Greene, C. I. Malme, and D. H. Thomson. *CHAPTER 7 - MARINE MAMMAL SOUNDS11By Denis H. Thomson and W. John Richardson, LGL Ltd.* In W. J. Richardson, C. R. Greene, C. I. Malme, and D. H. Thomson, editors, Marine Mammals and Noise, pages 159–204. Academic Press, San Diego, January 1995.

[27] L. Coquereau, J. Grall, L. Chauvaud, C. Gervaise, J. Clavier, A. Jolivet, and L. Di Iorio. *Sound Production and Associated Behaviours of Benthic Invertebrates from a Coastal Habitat in the North-East Atlantic*. Marine Biology, 163(5):127, May 2016.

[28] A. Farcas, C. F. Powell, K. L. Brookes, and N. D. Merchant. *Validated Shipping Noise Maps of the Northeast Atlantic*. Science of The Total Environment, 735:139509, September 2020.

[29] G. Zhang, T. N. Forland, E. Johnsen, G. Pedersen, and H. Dong. *Measurements of Underwater Noise Radiated by Commercial Ships at a Cabled Ocean Observatory*. Marine Pollution Bulletin, 153:110948, April 2020.

[30] A. MacGillivray and C. de Jong. *A Reference Spectrum Model for Estimating Source Levels of Marine Shipping Based on Automated Identification System Data*. Journal of Marine Science and Engineering, 9(4):369, April 2021.

[31] H. Ö. Sertlek, G. Aarts, S. Brasseur, H. Slabbekoorn, C. ten Cate, A. M. von Benda-Beckmann, and M. A. Ainslie. *Mapping Underwater Sound in the Dutch Part of the North Sea*. In A. N. Popper and A. Hawkins, editors, The Effects of Noise on Aquatic Life II, volume 875, pages 1001–1006. Springer New York, New York, NY, 2016.

[32] C. de Jong, B. Binnerts, P. de Krom, and T. Gaida. *North Sea Sound Maps 2019-2020*. Technical report, EU INTERREG Joint Monitoring Programme for Ambient Noise North Sea (Jomopans), 2020.

[33] R. L. Putland, C. A. F. de Jong, B. Binnerts, A. Farcas, and N. D. Merchant. *Multi-Site Validation of Shipping Noise Maps Using Field Measurements*. Marine Pollution Bulletin, 179:113733, June 2022.

[34] L. Hermannsen, L. Mikkelsen, J. Tougaard, K. Beedholm, M. Johnson, and P. T. Madsen. *Recreational Vessels without Automatic Identification System (AIS) Dominate Anthropogenic Noise Contributions to a Shallow Water Soundscape*. Scientific Reports, 9(1):15477, October 2019.

[35] J. Tougaard, L. Hermannsen, and P. T. Madsen. *How Loud Is the Underwater Noise from Operating Offshore Wind Turbines?* The Journal of the Acoustical Society of America, 148(5):2885–2893, November 2020.

[36] P. T. Madsen, M. Wahlberg, J. Tougaard, K. Lucke, and P. Tyack. *Wind Turbine Underwater Noise and Marine Mammals: Implications of Current Knowledge and Data Needs*. Marine Ecology Progress Series, 309:279–295, March 2006.

[37] S. Degraer, R. Brabant, B. Rumes, and L. Vigin. *Environmental Impacts of Offshore Wind Farms in the Belgian Part of the North Sea: Attraction, Avoidance and Habitat Use at Various Spatial Scales*. Technical report, Royal Belgian Institute of Natural Sciences, OD Natural Environment, Marine Ecology and Management, Brussels, December 2021.

[38] A. Norro. *On the Effectiveness of a Single Big Bubble Curtain as Mitigation Measure for Offshore Wind Farm Piling Sound in Belgian Waters*. Technical report, Royal Belgian Institute of Natural Sciences (RBINS), 2018.

[39] M. A. Bellmann. *Overview of Existing Noise Mitigation Systems for Reducing Pile-Driving Noise*. In Inter.Noise 2014, Melbourne, 2014.

[40] M. A. Bellmann, J. Schuckenbrock, S. Gündert, M. Müller, H. Holst, and P. Remmers. *Is There a State-of-the-Art to Reduce Pile-Driving Noise?* In J. Köppel, editor, Wind Energy and Wildlife Interactions: Presentations from the CWW2015 Conference, pages 161–172. Springer International Publishing, Cham, 2017.

[41] P. Rogers, E. Debusschere, D. de Haan, B. Martin, and H. Slabbekoorn. *North Sea Soundscapes from a Fish Perspective: Directional Patterns in Particle Motion and Masking Potential from Anthropogenic Noise*. The Journal of the Acoustical Society of America, 150(3):2174–2188, September 2021.

[42] A. J. R. Cotter. *The "Soundscape" of the Sea, Underwater Navigation, and Why We Should Be Listening More*. In Advances in Fisheries Science, chapter 19, pages 451–471. John Wiley & Sons, Ltd, 2008.

[43] P. D. Thorne. *An Overview of Underwater Sound Generated by Interparticle Collisions and Its Application to the Measurements of Coarse Sediment Bedload Transport*. Earth Surface Dynamics, 2(2):531–543, December 2014.

[44] H. Vindenes, K. A. Orvik, H. Søiland, and H. Wehde. *Analysis of Tidal Currents in the North Sea from Shipboard Acoustic Doppler Current Profiler Data*. Continental Shelf Research, 162:1–12, June 2018.

[45] B. J. Lee, M. Fettweis, E. Toorman, and F. J. Molz. *Multimodality of a Particle Size Distribution of Cohesive Suspended Particulate Matters in a Coastal Zone*. Journal of Geophysical Research: Oceans, 117(C3), 2012.

[46] G. M. Wenz. *Acoustic Ambient Noise in the Ocean: Spectra and Sources*. The Journal of the Acoustical Society of America, 34(12):1936–1956, December 1962.

[47] M. A. Ainslie. *Effect of Wind-Generated Bubbles on Fixed Range Acoustic Attenuation in Shallow Water at 1 – 4 kHz*. The Journal of the Acoustical Society of America, 3513(2005), 2011.

[48] M. A. Ainslie and P. H. Dahl. *Practical Spreading Laws: The Snakes and Ladders of Shallow Water Acoustics*. In Proceedings of Second International Conference and Exhibition on Underwater Acoustics, page 8, 2014.

[49] O. H. Sertlek. *Aria of the Dutch North Sea*. PhD thesis, Leiden University, June 2016.

[50] H. O. Sertlek, H. Slabbekoorn, C. J. Ten Cate, and M. A. Ainslie. *Insights into the Calculation of Metrics for Transient Sounds in Shallow Water*. In ECUA 2012 11th European Conference on Underwater Acoustics, page 070076, Edinburgh, Scotland, 2013.

[51] E. Ivanov, A. Capet, A. Barth, E. J. M. Delhez, K. Soetaert, and M. Grégoire. *Hydrodynamic Variability in the Southern Bight of the North Sea in Response to Typical Atmospheric and Tidal Regimes. Benefit of Using a High Resolution Model*. Ocean Modelling, 154:101682, October 2020.

[52] H. Ö. Sertlek, B. Binnerts, and M. A. Ainslie. *The Effect of Sound Speed Profile on Shallow Water Shipping Sound Maps*. The Journal of the Acoustical Society of America, 140(1):EL84–EL88, 2016.

[53] M. Ainslie. *Principles of Sonar Performance Modelling*. Springer Berlin Heidelberg, Berlin, Heidelberg, 2010.

[54] T. A. Mooney, L. Di Iorio, M. Lammers, T.-H. Lin, S. L. Nedelec, M. Parsons, C. Radford, E. Urban, and J. Stanley. *Listening Forward: Approaching Marine Biodiversity Assessments Using Acoustic Methods*. Royal Society Open Science, 7(8):201287, 2020.

[55] T. A. Marques, L. Thomas, S. W. Martin, D. K. Mellinger, J. A. Ward, D. J. Moretti, D. Harris, and P. L. Tyack. *Estimating Animal Population Density Using Passive Acoustics*. Biological Reviews, 88(2):287–309, 2013.

[56] E. Fleishman, D. Cholewiak, D. Gillespie, T. Helble, H. Klinck, E.-M. Nosal, and M. A. Roch. *Ecological Inferences about Marine Mammals from Passive Acoustic Data*. Biological Reviews, 98(5):1633–1647, 2023.

[57] International Organization for Standardization. *ISO 18405:2017*. Technical report, International Organization for Standardization, 2017.

[58] C. de Jong, B. Binnerts, S. Robinson, and L. Wang. *Guidelines for Modelling Ocean Ambient Noise*. Technical report, INTERREG North Sea Region, 2021.

[59] E. T. Kusel and M. Siderius. *Comparison of Propagation Models for the Characterization of Sound Pressure Fields*. IEEE Journal of Oceanic Engineering, 44(3):598–610, July 2019.

[60] S. G. Seri, M. E. Schinault, S. M. Penna, C. Zhu, L. D. Sivle, K. de Jong, N. O. Handegard, and P. Ratilal. *Characterizing Coastal Cod Vocalization Using a Towed Hydrophone Array*. ICES Journal of Marine Science, 80(6):1727–1745, August 2023.

[61] H. O. Sertlek, M. A. Ainslie, and K. D. Heaney. *Analytical and Numerical Propagation Loss Predictions for Gradually Range-Dependent Isospeed Waveguides*. IEEE Journal of Oceanic Engineering, 44(4):1240–1252, 2019.

[62] C. Bassett, J. Thomson, P. H. Dahl, and B. Polagye. *Flow-Noise and Turbulence in Two Tidal Channels*. The Journal of the Acoustical Society of America, 135(4):1764–1774, April 2014.

[63] A. M. von Benda-Beckmann, P. J. Wensveen, F. I. P. Samarra, S. P. Beerens, and P. J. O. Miller. *Separating Underwater Ambient Noise from Flow Noise Recorded on Stereo Acoustic Tags Attached to Marine Mammals*. Journal of Experimental Biology, 219(15):2271–2275, August 2016.

[64] M. R. Heupel, K. L. Reiss, B. G. Yeiser, and C. A. Simpfendorfer. *Effects of Biofouling on Performance of Moored Data Logging Acoustic Receivers*. Limnology and Oceanography: Methods, 6(7):327–335, 2008.

[65] MFC ODNature RBINS. *Physical State of the Sea - Belgian Coastal Zone - COHERENS UKMO*, 2023.

[66] C. Parcerisas. *Lifewatch/Bpnsdata: First Release of Bpnsdata*. Zenodo, November 2021.

[67] L. Kint, F. Barette, K. Degrendele, M. Roche, and V. Van Lancker. *Sediment Variability in Intermittently Extracted Sandbanks in the Belgian Part of the North Sea*. Frontiers in Earth Science, 11, August 2023.

[68] L. Kint, N. Terseleer, and V. Van Lancker. *Multi-Scale (or Multiresolution) Analysis of Sandbank Features Optimising Geomorphological Mapping of Sandy Shelf Environments: Belgian Part of the North Sea*. In Books of Abstracts of the International Conference of Marine and River Dune Dynamics, pages 127–133, Bremen, Germany, April 2019.

[69] M. Fettweis and D. Van den Eynde. *The Mud Deposits and the High Turbidity in the Belgian–Dutch Coastal Zone, Southern Bight of the North Sea*. Continental Shelf Research, 23(7):669–691, May 2003.

[70] N. D. Merchant, T. R. Barton, P. M. Thompson, E. Pirotta, D. T. Dakin, and J. Dorocicz. *Spectral Probability Density as a Tool for Ambient Noise Analysis*. The Journal of the Acoustical Society of America, 133(4):EL262–EL267, April 2013.

[71] F.-W. Fourie, K. Langedock, R. Develter, H. Loop, C. J. Peck, L. Ponsoni, H. Pirlet, and W. Boone. *Sonarlogger: Enabling Long-Term Underwater Sonar Observations*. HardwareX, 18, June 2024.

[72] J. Hubert, J. M. Demuynck, M. R. Remmelzwaal, C. Muñiz, E. Debusschere, B. Berges, and H. Slabbekoorn. *An Experimental Sound Exposure Study at Sea: No Spatial Deterrence of Free-Ranging Pelagic Fisha)*. The Journal of the Acoustical Society of America, 155(2):1151–1161, February 2024.

[73] A. Calonge, J. Goossens, C. Muñiz, J. Reubens, and E. Debusschere. *Importance of Multi-Sensor Observations to Advance Species Co-Occurrence Knowledge: A Demonstration of Two Acoustic Technologies*. Marine Ecology Progress Series, 727:49–65, January 2024.

[74] J. Goossens, M. T'Jampens, K. Deneudt, and J. Reubens. *Mooring Scientific Instruments on the Seabed—Design, Deployment Protocol and Performance of a Recoverable Frame for Acoustic Receivers*. Methods in Ecology and Evolution, 11(8):974–979, 2020.

[75] J. Y. Chung. *Rejection of Flow Noise Using a Coherence Function Method*. Journal of the Acoustical Society of America, 62(2):388–395, 1977.

[76] H. H. Kimball. *The Duration and Intensity of Twilight*. Monthly Weather Review, 44(11):614–620, 1916.

[77] A. Lillis, F. Caruso, T. A. Mooney, J. Llopiz, D. Bohnenstiehl, and D. B. Eggleston. *Drifting Hydrophones as an Ecologically Meaningful Approach to Underwater Soundscape Measurement in Coastal Benthic Habitats*. Journal of Ecoacoustics, 2(February):STBDH1, 2018.

[78] J. J. Luczkovich and M. W. Sprague. *Soundscape Maps of Soniferous Fishes Observed From a Mobile Glider*. Frontiers in Marine Science, 9, 2022.

[79] M. Bolgan, C. Gervaise, L. Di Iorio, J. Lossent, P. Lejeune, X. Raick, and E. Parmentier. *Fish Biophony in a Mediterranean Submarine Canyon*. The Journal of the Acoustical Society of America, 147(4):2466–2477, April 2020.

[80] K. A. Kowarski, B. J. Gaudet, A. J. Cole, E. E. Maxner, S. P. Turner, S. B. Martin, H. D. Johnson, and J. E. Moloney. *Near Real-Time Marine Mammal Monitoring from Gliders: Practical Challenges, System Development, and Management Implications*. The Journal of the Acoustical Society of America, 148(3):1215–1230, September 2020.

[81] C. Wall, C. Lembke, and D. Mann. *Shelf-Scale Mapping of Sound Production by Fishes in the Eastern Gulf of Mexico, Using Autonomous Glider Technology*. Marine Ecology Progress Series, 449:55–64, March 2012.

[82] Y. Wei, M. Li, J. Liu, Y. Chen, F. Qu, and M. Zhang. *Flow Noise and Its Suppression in Exposed Hydrophone*. Applied Acoustics, 182:108190, November 2021.

[83] S. P. Beerens, S. P. van IJsselmuide, C. Volwerk, E. Trouve, Y. Doisy, and T. F. e. E. Laboratorium. *Flow Noise Analysis of Towed Sonar Arrays*. UDT Europe 1999 - Conference Proceedings Undersea Defence Technology, 29 June - 1 July 1999, Nice, France, page 392, 1999.

[84] N. D. Merchant, K. M. Fristrup, M. P. Johnson, P. L. Tyack, M. J. Witt, P. Blondel, and S. E. Parks. *Measuring Acoustic Habitats*. Methods in Ecology and Evolution, 6(3):257–265, 2015.

[85] J. P. Ryan, J. E. Joseph, T. Margolina, L. T. Hatch, A. Azzara, A. Reyes, B. L. Southall, A. DeVogelaere, L. E. Peavey Reeves, Y. Zhang, D. E. Cline, B. Jones, P. McGill, S. Baumann-Pickering, and A. K. Stimpert. *Reduction of Low-Frequency Vessel Noise in Monterey Bay National Marine Sanctuary During the COVID-19 Pandemic*. Frontiers in Marine Science, 8, 2021.

[86] M. J. G. Parsons, T.-H. Lin, T. A. Mooney, C. Erbe, F. Juanes, M. Lammers, S. Li, S. Linke, A. Looby, S. L. Nedelec, I. Van Opzeeland, C. Radford, A. N. Rice, L. Sayigh, J. Stanley, E. Urban, and L. Di Iorio. *Sounding the Call for a Global Library of Underwater Biological Sounds*. Frontiers in Ecology and Evolution, 10, 2022.

Part II

Holistic Characterization of Soundscapes

4

Understanding BPNS soundscapes using traditional sound metrics

The Introduction of this chapter is adapted from the publication:

- *Parcerisas, C., Botteldooren, D., Devos, P., Hamard, Q., Debusschere, E. (2023). Studying the Soundscape of Shallow and Heavy Used Marine Areas: Belgian Part of the North Sea. In: Popper, A.N., Sisneros, J., Hawkins, A.D., Thomsen, F. (eds) The Effects of Noise on Aquatic Life. Springer, Cham.* https://doi.org/10.1007/978-3-031-10417-6_122-1

Overview

Natural marine soundscapes are being threatened by increasing anthropogenic noise, particularly in shallow coastal waters. To preserve and monitor these soundscapes, understanding them is essential. There are several ways to do so, each with its limitations. In this chapter, we summarize the most commonly used approaches, and then we use some of these techniques to describe the BPNS soundscape. With this, we introduce the standard baseline approach to answer the research question:

Can we holistically describe and characterize the soundscape of the BPNS, and identify the external environmental factors (such as wind, rain, and currents) that influence it?

These methods include visualization and interpretation of Long Term Spectral Averages (LTSA) and Spectral Probability Density (SPD) plots, comparisons of spectrum and broadband sound pressure levels between locations, and linear correlations with environmental factors. This chapter has not been published and has been written specifically for this thesis.

4.1 Introduction

As identified in the report of the International Quiet Ocean Experiment (IQOE) on marine bioacoustics standards [1], there are currently no standards to measure, characterize, and analyze underwater soundscapes. In 2018 the Marine Strategy Framework Directive (MSFD) proposed the use of sound levels at the 1/3-octave bands centered at 63 and 125 Hz as shipping noise indicators [2]. In 2023 the guide to set EU threshold values related to anthropogenic continuous noise in water was released. There it is proposed to select the octave band more adapted to the target species when assessing the impact of anthropogenic noise [3].

Regarding data processing, there are no standards for analyzing and reporting these soundscapes, which makes it difficult to compare them. Some studies focus on temporal patterns, while others report broadband frequency spectrum per site. Nonetheless, some practices have become widely used to analyze underwater soundscapes, as summarized in [4]. Some of these analyses focus on obtaining a general overview. They can be used to describe the soundscape of a particular habitat or to compare multiple habitats. Data are generally explored using spectrograms for shorter periods (minutes or hours) and the long-term spectral average (LTSA) for longer periods (days, months, or years) [4]. Spectral Probability Density (SPD) plot is a common way to plot a soundscape summary, it gives an idea of how often certain frequencies appear at a certain sound pressure level [5]. Another common way of analyzing soundscapes is to compute the Sound Pressure Level (SPL) broadband or at specific frequency bands of interest. These frequency bands are sometimes selected according to the expected sound sources of interest or known sensitivity of certain species (e.g. [6, 7]), but third-octave bands or decidecade bands are also commonly used. A common approach is to measure the variation in pressure within a specified frequency band and time interval; however, only using sound pressure levels can limit the scope of interpretations [8]. Characterizing marine soundscapes using sound levels of frequency bands can be very useful in some cases, for example when monitoring a particular biological chorus. However, in other cases it offers limited information as it lacks comprehensive details about the frequency and temporal distribution of the soundscape and its unique acoustic characteristics. For example, when studying biophony in areas dominated by anthropogenic sound, time-averaging can lead to the fading of some shorter and less common biological sounds.

Making (raw) acoustic data available can be challenging because of its size (order of several TB or even PB). Currently it seems like decentralized storage is preferred, as storing all the PAM data which has been collected on one single place would require access to huge storing spaces. One option to make large raw datasets available is to use cloud storage, but obtaining several TB of storing space in the cloud might not be an option for all the institutions collecting data. In this context, hybrid millidecade band (sound pressure) levels (HMBL) were proposed by Miksis-Olds et al. (2021)[9] as an alternative to decidecade bands. This approach allows for higher resolution spectral comparisons between stations while maintaining data dimensions that are manageable for distribution and use. Hybrid millidecade bands (HMB) are at a higher spectral resolution than decidecade bands but share the frequency band limits. Therefore it is not necessary to re-compute the band sound pressure levels to convert HMBL to decidecade band levels (not possible the other way around without recomputing the spectrum). For this reason, HMBL were chosen for the LifeWatch Broadband Acoustic Network [10] as a reasonably small data product which could be used to compare sound levels across the globe, both for the data products from the U.S. (SoundCoop, SanctSound projects) and the European ones (JOMOPANS) in a fast and accessible way. To this aim, the long-term collected data described in Section 2.2 were processed to 1-minute HMBL [9] using the package pypam [11] which was developed as part of this PhD dissertation.

To gain insight into the soundscapes of the BPNS, the obtained long-term data were processed and analyzed to understand dynamics and main sound contributing agents mentioned in Section 3.3 in a holistic way instead of focusing on sound sources. We assess also want to asses which information can we extract from HMBL to describe the BPNS soundscape, and if these are suitable metrics to report it. This analysis aims to answer the following questions:

1. What are the acoustic similarities and differences between measurement locations? What drives these differences?
2. Do short and loud events contribute to HMBL? Can they be detected with simple threshold levels?
3. Which frequency bands are correlated to environmental parameters? How do these parameters influence the soundscape?
4. How is shipping correlated with HMBL?

To answer question (1), we looked at the acoustic differences between locations and discussed the possible drivers. We compared the SPDs of all the locations, and compared broadband levels of several locations over the same period. Then, for question (2) we tested the possible contribution of sparse biophony to the long-term averaged sound levels, and analyzed if the difference between the median and mean hybrid millidecade band (sound pressure) levels can be used to detect such events. For question (3), we investigated the correlation between the sources explained in Section 3.3.3 - wind, rain, current - with hybrid millidecade band (sound pressure) levels. Finally, to answer question (4) we computed the correlation between AIS data and sound levels of the hybrid millidecade bands in the shipping frequency bands. As complementary information to answer (3) and (4), we created daily LTSAs plotted together with shipping and current and analyzed them manually.

During all these analyses, we compared the obtained spectrum levels to levels measured by Wenz (1962) for shallow waters [12] (see Figure 4.1). This comparison allowed us to understand what sources were driving which characteristics of the spectrum from the different locations. From now on, we will refer to the ensemble of these methods as the traditional statistical methods.

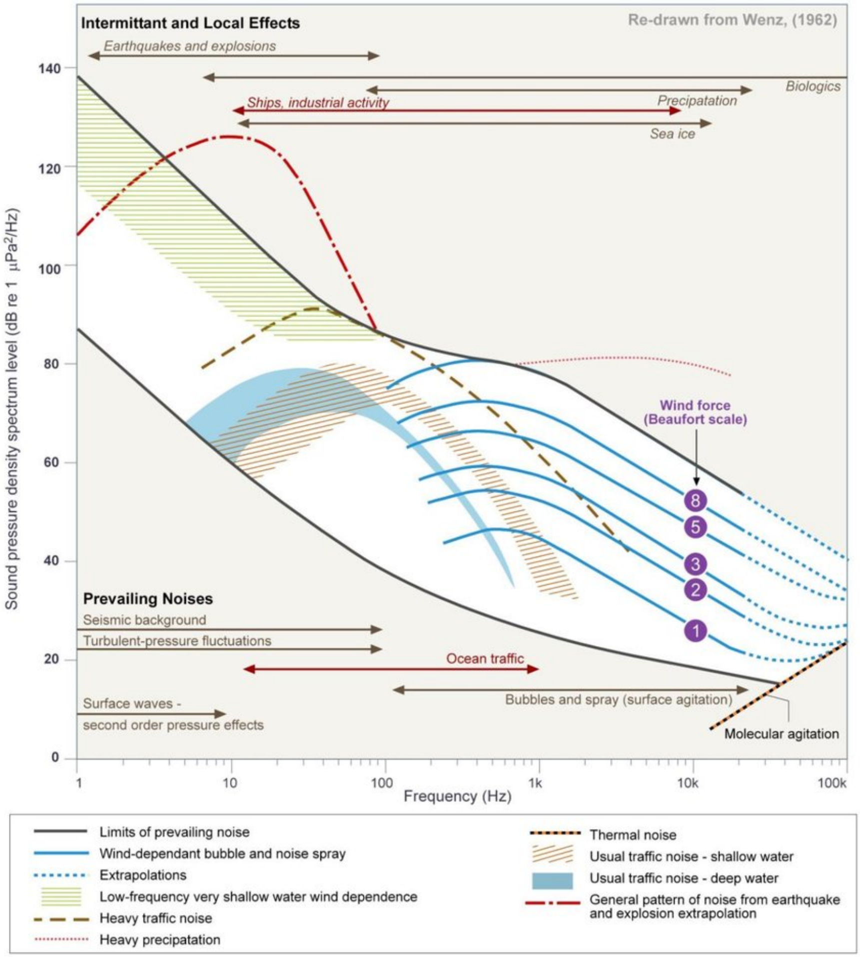
4.2 Methods

4.2.1 Data

The deployments from Table 2.1 (Section 2.2 from Chapter 2) marked with a 1 on the 'HMBL' column were processed to hybrid millidecade band (sound pressure) levels (HMBL) [9] using pypam [11]. The hybrid millidecade bands data product explained in Miksis-Olds et al. (2021) [9] proposes to compute the spectrum at 1-second temporal resolution and 1 Hz spectral resolution (FFT bins equal to sampling rate). This is obtained using a Hann window, with 50 % overlap. This process is repeated for each non-overlapping 1-minute window. The 1 Hz resolution spectrum is then grouped per hybrid millidecade band following the process described in Martin et al. (2021) [13], only at frequencies above 455 Hz. The spectrum below 455 Hz is left at 1 Hz resolution. Millidecade bands share band frequency limits with decidecade bands, but at higher resolution, comprising 1000 bands each octave instead of 10. The data products obtained with pypam are daily netcdf files at 1-minute resolution. Before computing the HMBL, data were convereted from WAV to μPa as explained in Section 2.4, using a flat calibration (not frequency dependent).

Only deployments from March 2021 to November 2022 were considered. The resulting dataset included 7 locations, with 15 different deployments whose durations are reported in Figure 2.4. See Figure 2.2 for more detailed information about the locations. Note that Belwind station is situated in an offshore windmill park. In such a shallow area as the BPNS, the depth of the water column can have a strong impact on the soundscape, as it can cause low-frequency sound to be strongly attenuated, as explained in Section 3.4.1. The depths of the locations are Buitenratel (7 m), Fairplay (14.5 m), Grafton (19 m), Faulbaums (19 m), Belwind (26 m), GardenCity (28 m) and Birkenfels (37.5 m). The increase in depth is also

Figure 4.1 The typical sound levels of ocean background noises at different frequencies, as measured by Wenz (1962) [12]. This graph is therefore also referred to as the Wenz curves. The sound levels are given in underwater dB re 1 $\mu Pa^2/Hz$. Adapted (after National Research Council, 2003) from Wenz, G. M. (1962) [12]. Acoustic ambient noise in the ocean: Spectra and sources. The Journal of the Acoustical Society of America, 34(12), 1936–1956. Copyright Acoustical Society of America, reprinted with permission.



linked to an increased distance to the coast. The differences in depth are expected to have an impact on the soundscape at these locations.

4.2.2 Acoustic differences between locations

To compare the acoustic characteristics between the locations we plotted the SPD [5] and the percentile exceedance values of the hourly median hybrid millidecade data over the whole measurement period at each location.

Because not all the locations were recorded at the same time, the SPD comparison comprises different periods for different locations. Therefore, differences that might look spatial could be temporal instead. For this reason, we analyzed if the locations presented different temporal patterns when comparing the hourly HMBL, aggregated by week, in the most dominant shipping frequency band during the period when most locations were recording simultaneously. This turned out to be from the 8th of March 2021 to the 31st of May 2021. The frequency band of comparison was chosen to be 50 to 200 Hz, as this includes the loudest frequencies for shipping in shallow waters (see Figure 4.1) but excludes lower frequencies known to be influenced by flow noise.

4.2.3 Contribution of short events to long-term averaging

In some areas, loud and continuous biophony can be the main contributors to the soundscape of an area [14]. Usually, this phenomenon is caused by chorusing fauna [15]. This is not the case for the Belgian Part of the North Sea, as no fish chorus has been reported to date in the BPNS to the author's knowledge. However, that does not mean short, loud events cannot be present in long-term data. These events affect the mean when averaging, but not the median. To further assess the main contributors to the soundscape a test was performed to determine whether any bioacoustic activities could be extracted from the 1-minute hybrid millidecade band (sound pressure) levels.

For this analysis the 7th of November of 2022 (from Grafton station, Stationary dataset, see Section 2.2) was selected because during manual screening fish sounds were found. The HMBL for this test were computed following two approaches. First the one explained in Section 4.2.1 (HMBL-mean), and second the same method but the aggregation from 1-second spectra to 1-minute was done using a median instead of a mean (HMBL-median). For this test, only frequencies lower than 4 kHz were considered. This frequency range was chosen as it is the frequency range known for fish vocalizations [16, 17]. Prior to computing the HMBL (both mean and median), data were filtered to the frequency range of interest and then decimated to 8000 Hz. The difference between the mean and the median was computed.

Manual annotations of short salient events of interest were conducted using Raven Pro 1.6.5 [18] for that particular day. Only annotations below 4000 Hz

Table 4.1 Confusion matrix between detections and threshold exceedance of the difference between mean and median. Diff stands for difference between mean and median. Th stands for threshold (6 dB).

		Manual annotations	
		Detection	No detection
Diff	Diff≥Th	a	b
	Diff<Th	c	d

were selected for this analysis. These events were manually classified into different categories. These categories included: 'Click', 'Grunt', 'Poc', 'Crustacean stridulation', 'Tick', 'Metallic sound', 'Metallic bell', 'Knock', 'Jackhammer', 'Croak', and 'Siren'. Most of the sources of these categories remain unknown, but the Jackhammer sound has been associated with fish sounds in other studies (see Chapters 7 and 8). These events ranged from 10 ms to 3.5 s.

The event detections were converted to binary data for each minute, so 1 would represent a minute where an event was present and 0 a minute with no event (variable d). The difference between the mean and the median was also converted to binary data per frequency band, where 1 was used to code that the mean exceeded the median by 6 dB in that frequency band (variable f). 6 db was chosen as a good trade-off between small and big differences. The Phi (ϕ) correlation coefficient was computed between d and f. This coefficient is commonly used in statistics as a measure of association between two binary variables (see equation 4.1). This correlation coefficient was then plotted for each frequency band and assigned label.

$$\phi = \frac{ab - bc}{\sqrt{(a+b)(c+d)(a+c)(b+d)}} \tag{4.1}$$

Where, X and Y are binary random variables and a, b, c, and d, are the number of observations (i.e., manual detections or difference exceeding the threshold) specified in Table 4.1.

4.2.4 Correlations with environmental data

To match the time resolution from the environmental data, the 1-minute HMBL were aggregated into median hourly HMBL. Wind, current, tide, and rain data were assigned to each HMBL spectra (each hour). Current and tide data were obtained from the COHERENS UMO model from RBINS [19] (same data as the one mentioned in Table 5.1), available at 1 hour resolution. The tide variable was represented by the height above the average sea level (in cm), while current was the total speed (m/s) at the bottom. Wind and rain were extracted from the closest monitoring buoy from the Meetnet Vlaamse Banken [20], also at 1 hour resolution.

We evaluated the correlation between the environmental factors tide, current, wind, and rain, and all the hourly median hybrid millidecade band (sound pressure) levels, following a similar analysis to the one done in Basan et al. (2024) [7]. A Spearman correlation coefficient [21] was computed for each location, frequency band, and each environmental parameter. Spearman correlation coefficient was chosen to allow for the non-parametric distributions of the compared parameters (see equation 4.2). The correlation obtained at each frequency band was then plotted per location. Then the hourly median hybrid millidecade band (sound pressure) levels were plotted for different environmental conditions of wind and current speed to assess the acoustic influence of the wind and the tide for Belwind and Buitenratel, respectively. This visualization helps understand exactly which environmental conditions raise the sound levels and the dominant frequencies of these raises, and it can provide further insight into understanding the obtained correlation values.

$$r_s = 1 - \frac{6\sum d_i^2}{n(n^2-1)} \tag{4.2}$$

Where[1],
d_i is the difference between the two ranks of each observation,
n is the number of observations,
r_s is the is the Spearman's rank correlation coefficient.

4.2.5 LTSAs manual exploration

Visually examining LTSAs can offer insights into the primary contributors to a soundscape. We generated daily LTSAs using the computed 1-minute HMBL, allowing us to visually identify the dominant frequencies during specific hours and days. We manually scrolled through the generated daily LTSAs from full deployments, to investigate the spectral and temporal variations of the deployments, as previously done in multiple soundscape analyses [4]. LTSAs can be used to highlight repetitive elements of the acoustic environment, such as the constant presence of sounds like shipping activity or storms, recurrent contributors such as biological choruses, and significant singular events like earthquakes. They enable the selection of particular days or hours containing sounds of interest (if known), the differentiation between quiet and loud periods, and the exploration of connections between acoustic patterns and environmental variables.

The AIS data was obtained via de AISHub present at VLIZ. These data were collected from October 2021 onward, so for this exploration only data from after this date were considered. As a preliminary examination, all the messages closer than 0.2 degrees in latitude and longitude were considered. The messages were then joined by Maritime Mobile Service Identity (MMSI). AIS data were aggregated

[1]This equation is only valid if all n ranks are distinct integers

to form segments. Each segment represented the trajectory of that vessel when crossing the area. For each of these segments, an average distance to the sensor was computed using the Cartesian equation. AIS data were plotted and superimposed with the LTSAs. The minimum distance from any vessel to the recorder was computed per minute as a function of distance from the recorder and time. These plots allowed us to asses up to what distance the sound of vessels was just part of the background noise, and at what distance they were individually distinguishable. The current data from Section 4.2.4 was also plotted simultaneously with the LTSAs.

4.2.6 Correlation with shipping

To investigate underwater sound from vessel traffic in the LTSA, the correlation between hybrid millidecade band (sound pressure) levels and AIS data was explored. Several extra filters were applied to the AIS data from 4.2.5: 1) a minimum speed of 0.1 knots for ships to be considered, 2) the segment distance to the measurement location has to be closer than 10 km. This decision was taken after the preliminary analysis, where only closer vessels would influence the received sound levels.

Two different metrics were computed to assess the correlation between AIS data and median hourly HMB. These metrics were computed for each frequency band at every location. For each hour, we collected (1) the number of (unique) ships present in the area, as done in Ryan et al. (2021) [15], and (2) a metric extracted from Basan et al. (2024) [7] (see equation 4.3). As described in Basan et al. (2024), this metric (CS) is a simplified index representing the contribution of shipping to overall received sound levels. It is distance-weighted and accounts for the presence of multiple ships at varying distances. This approach is suitable for identifying frequency bands correlated with shipping but not for estimating the energy received from ships. Only considering the number of ships might be too simplistic, as it might not capture the complex relationship betwween ships presence and received SPL at a certain frequency band, which is influenced by numerous factors [7]. Spearman's correlation coefficient was computed for each frequency band and each of the two metrics.

$$CS = 10\log_{10}\left(\sum_{i=1}^{N}\left(\frac{r_i}{r_0}\right)^{-0.1B}\right)\mathrm{dB},\ \text{with } r_0 = 1 \text{ m and } B = 20 \tag{4.3}$$

where r_i is the distance to the vessel and N the number of vessels.

4.3 Results

4.3.1 Acoustic differences between locations

All the locations except Fairplay, and Buitenratel, both locations in very shallow waters (14.5 and 7 m deep, respectively), and far away from any major shipping lane, present a spectral distribution with a maximum around 60 Hz (see Figure 4.2), typical of vessel sounds [22]. This curve matches the theoretical one from Wenz (1962) [12] for heavy traffic in shallow waters (see Figure 4.1), which is to be expected when looking at the human activities in the BPNS.

Faulbaums, Grafton, and Belwind (mid-depth locations) present a dip in frequencies between 10 and 100 Hz, while Buitenratel and Fairplay (shallow locations) present a rise in level at the lower frequencies. This is the frequency range of the flow noise disturbance (below 100 Hz) [7]. The location at Belwind, which is at an offshore wind farm, presents an interesting pattern of apparent harmonics between 1 and 10 kHz. It is the only location presenting this pattern, and also the only location at a wind farm, so these sounds could be attributed to some operational sounds only present at wind farms. They are not generated by the turbines themselves as they do not produce tonal sounds [23].

Figure 4.2 Comparison of the median hourly spectrum of each location. Lx lines represent the xth percentile of exceedance level (i.e. L50 is the 50th percentile exceedance level). Each SPD per location is computed with all the available data for that location. Therefore, different locations comprised data from different periods. The dashed line is at 60 Hz (typical for shipping noise, [24]). The colors of the boxplot are aligned with the colors of the titles of each location's SPD, and also with the L50 comparison. D stands for deep, M+ for medium-high, M- for medium-low and S for shallow.

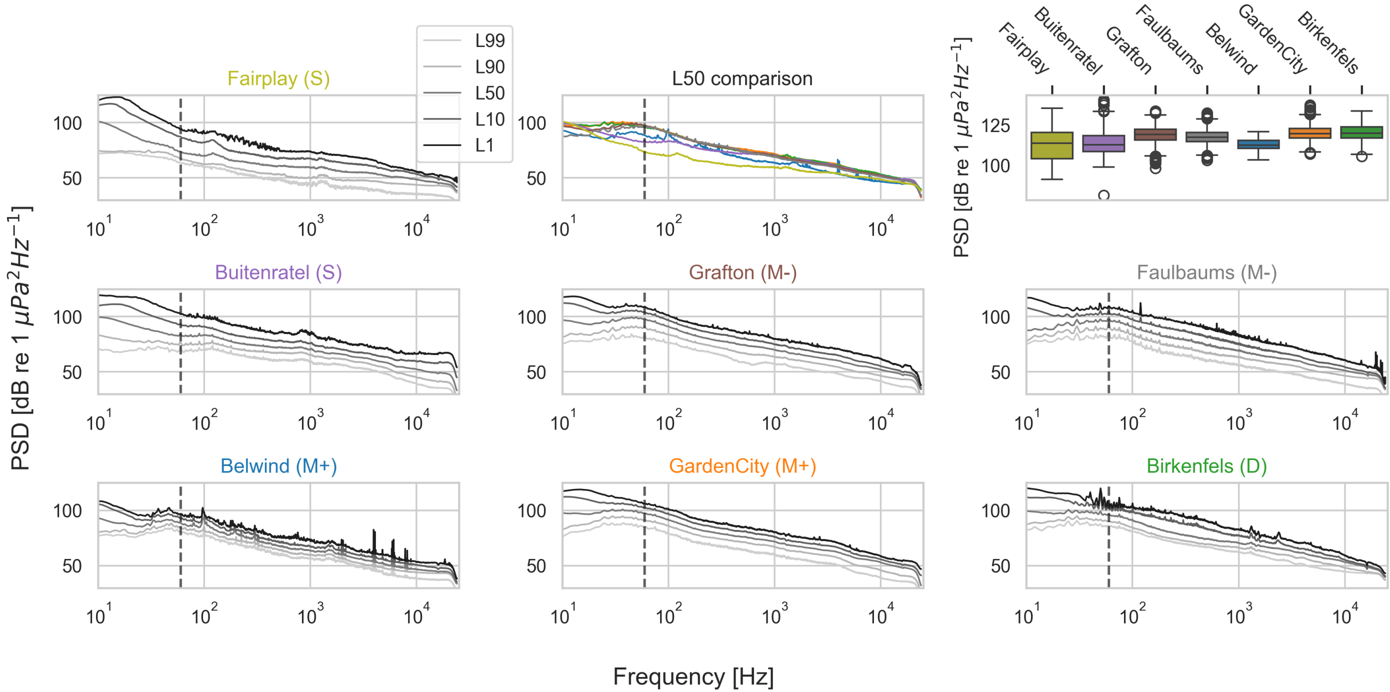


Figure 4.3 Comparison of median hourly HMBL between 50 Hz and 200 Hz for several locations from March to May 2021, aggregated by week (boxplot).

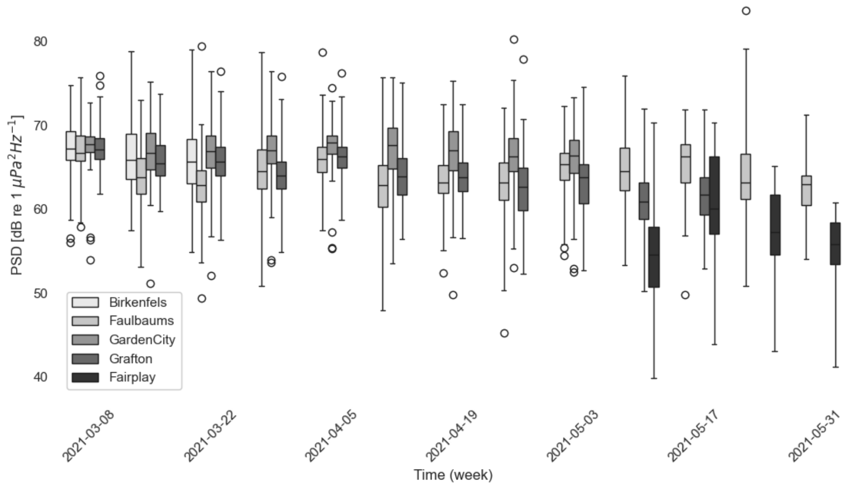


Figure 4.3 shows the result of the median weekly comparison from the 8th of March 2021 to the 31st of May 2021. For this period, the locations Birkenfels and GardenCity are the loudest in this frequency range. These locations are also the closest to the shipping lanes (see Figure 2.2). On the other hand, Faulbaums and Grafton present lower SPLs. Fairplay is significantly quieter than the rest.

4.3.2 Contribution of short events to long-term averaging

The minimum of 6 dB difference between the HMBL-mean and HMBL-median showed that short and loud sound events (10 ms to 3.5 s) can affect the LTSA when using the mean (see Figure 4.4), as they did not raise the levels when using the median, but they did when using the mean. After the correlation analysis between the difference mean-median and the presence of sound events it could be concluded that this different at 1-minute resolution was greater when short loud events are present (see Figure 4.5). These differences occur only at the frequency range of these short transient events.

The jackhammer presented the strongest correlation with the difference between mean and median. ‘Fish grunt’, ‘knock’, ‘poc’, and ‘tick’ presented weak correlations. The other annotated events did not present a correlation. The annotated events do not present any clear diel patterns, so they could be from abiotic, biotic, or man-made sound sources.

Figure 4.4 Comparison of the LTSA from 1-minute HMBL of a full day (7th November 2022) at Grafton location using (top) mean or (middle) median when aggregating the obtained 1-second HMBL into the 1-minute HMBL data product. (bottom) Difference between the mean and the median.

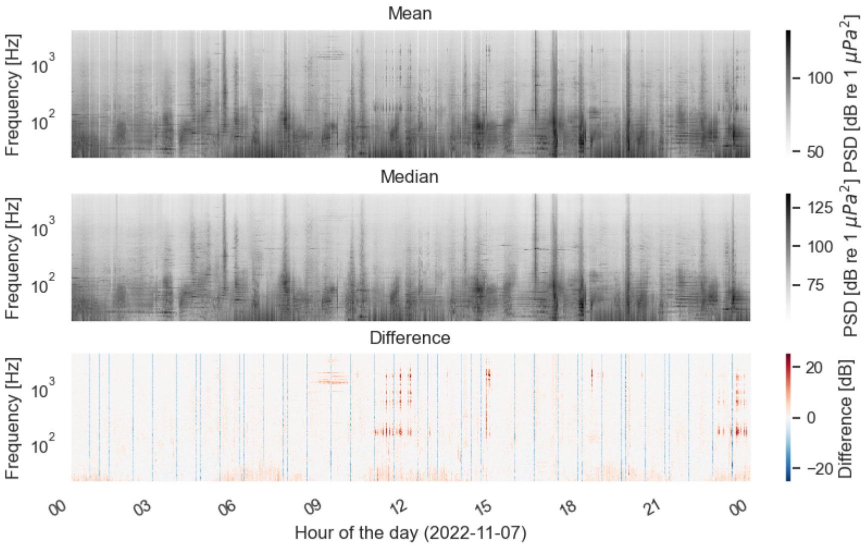
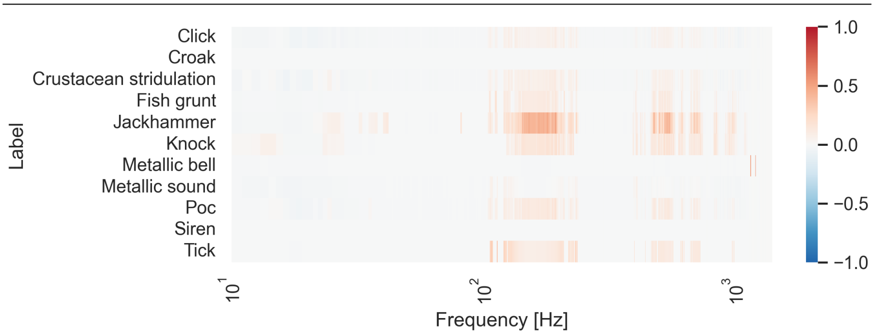


Figure 4.5 Phi (ϕ) coefficient computed between the presence of each of the labels (1 if any event was detected in a minute, 0 if not) and the difference between mean and median of each hybrid millidecade band (1 if the difference in that frequency band exceeds 6 dB). Only data from 7-11-2022 from Grafton was considered.



4.3.3 Correlation with environmental data

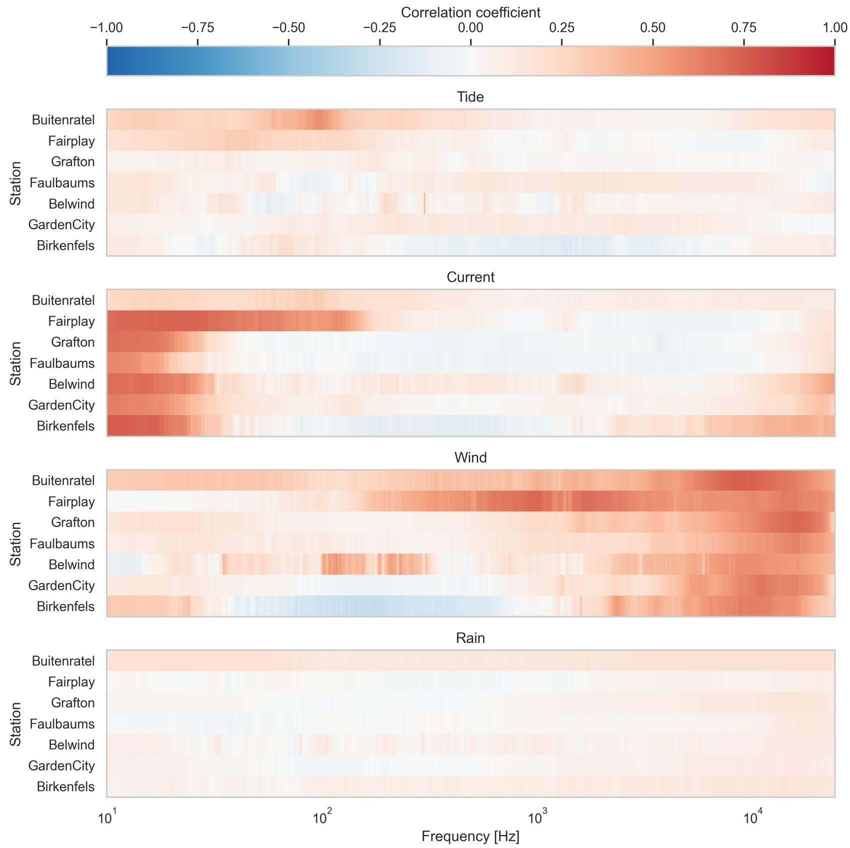
The considered environmental parameters (tide, current, wind, and rain) appeared to be strongly correlated with some frequency bands except rain. Results can be seen in Figure 4.6. The current is correlated with frequencies below 60 Hz for all the locations except Buitenratel. For Fairplay, the correlation is high up to 100 Hz. Buitenratel presents a very weak correlation with current at low frequencies, but a higher correlation at 100 Hz with tide, which the other locations do not present. This could be because Buitenratel is the shallowest location, featuring only 7 meters deep (average). In Figure 4.2 Buitenratel presents a high peak at the lower frequency range. Therefore, the lack of correlation could be caused not by an absence of flow noise but by a constant presence of it. Because of the shallow water column, there might be constant movement of water due to waves breaking and turbulence. The correlation with the tide around 100 Hz could be due to the change in the propagation of sound as the water column depth changes with the tide. The highest observed height above mean sea level was 3.3 m, and the minimum was -2.6 m . This means that the water depth changes from 4.4 to 10.6 m, more than double. Higher tides would allow for low frequencies coming from far away sound sources such as shipping to be recorded.

All the locations have strong correlations with wind at higher frequencies (>2 kHz). The shallower locations again present a bit of a different pattern. At Buitenratel correlation is present throughout the frequency spectrum, even though weaker at lower frequencies. For Fairplay, strong correlations can be observed on all the frequencies higher than 200 Hz. An interesting observation is the correlation between wind and frequency bands between 100 and 200 Hz at Belwind. This is not the typical range where wind is known to produce sound (see Figure 4.1), and it is very different at the other locations. This could therefore be related to the operational wind turbines present at the Belwind location (next to a offshore wind farm). This frequency range is in agreement with the one reported in Tougaard et al. (2020) [23] and Madsen et al. (2006) [25] when studying operational wind turbine sound. Even though the sound produced by one operational turbine is rapidly attenuated and falls below the ambient noise, several turbines operating in a wind farm might have a cumulative effect reflecting in an increment of sound levels at frequencies between 100 and 200 Hz when close to the wind park.

Furthermore, the influence of current and wind on the spectrum can be seen when plotting the spectra for different environmental conditions. The influence of the current-generated sounds and the sound associated with wind for specific locations can be seen in Figure 4.7. To gain a deeper understanding, we plotted the spectrum for different tide values at Buitenratel and different wind intensities for Belwind.

For Buitenratel, this plot allows us to disentangle the effect of tide and current on the spectrum. This is not possible just by looking at correlations, as these two

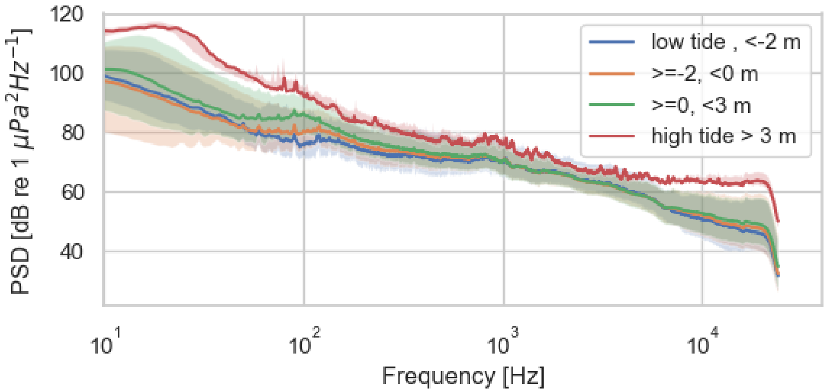
Figure 4.6 Spearman Correlation coefficients for the tide (height above mean sea level), current, wind, and rain with all the hybrid millidecade band (sound pressure) levels. locations are ordered from shallower (Buitenratel, 7 m) to deeper (Birkenfels, 37.5 m) locations.



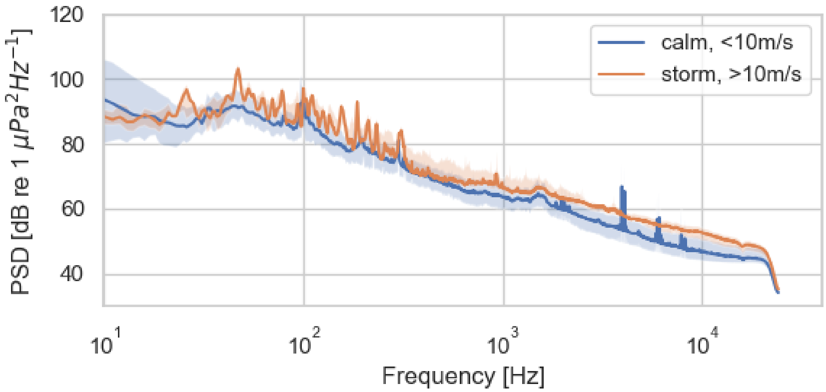
parameters are tightly correlated (see Figure 4.7a). In this case, we see that the depth of the water column seems to have a great influence on the received sound pressure levels rather than the current. Sound pressure levels received during high tide are a lot higher than in other conditions. This suggests that the sound is not propagating when the water column is not deep enough. The level increase for the high tide is however larger for lower frequencies ($<$ 50 Hz), typical of flow noise, and higher frequencies ($>$ 5 kHz). On the other hand, when the tide is lower than average ($<$ 0 m), the peak at 100 Hz often related to shipping noise is barely present. For medium tide (0 to 3 m above average) the shipping bump at 100 Hz increases, and so do the lower frequencies.

Following the same method for Belwind, we can also see an increase in sound levels when the wind is stronger than 5 Beaufort (see Figure 4.7b). In the low frequency, this happens in two peaks, one from 20 to 100 Hz and another one from 100 to 200 Hz. While the second one is known to be in the range of wind noise, the first one is not (see 4.1). Yet this first peak is in the frequency range of sound generated by operating wind turbines [23]. In the high frequency ($>$ 1 kHz), stronger winds lead to higher received sound levels in general, as expected. However, two peaks of tonal sounds (at 4 and 6 kHz) are not present wind is stronger, and therefore they are most likely not caused by the wind.

Figure 4.7 Spectrum of all the processed HMBL according to several environmental conditions.



(a) Tide influence on the spectrum for Buitenratel location (7 m deep)



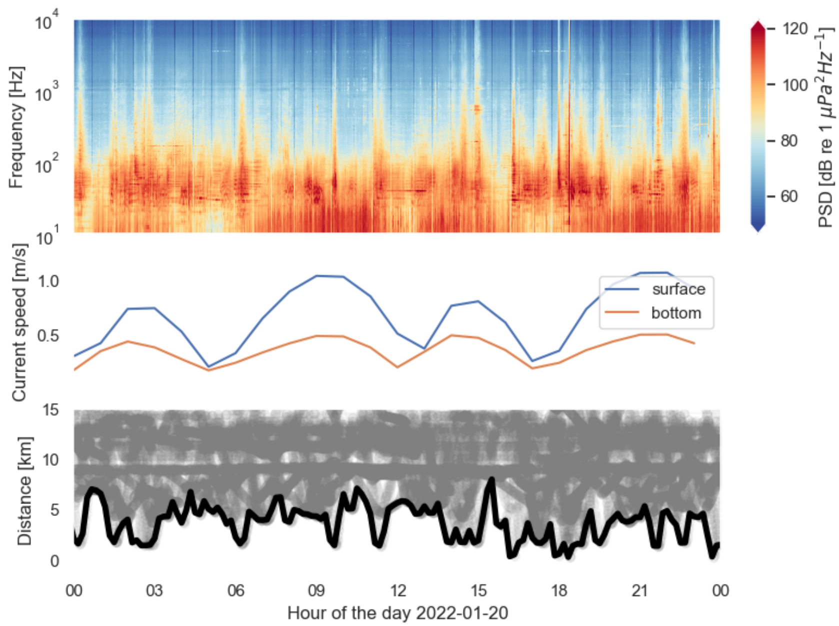
(b) Wind influence on the spectrum for Belwind (wind park) location (26 m deep)

4.3.4 LTSAs manual exploration

The manual analysis of the LTSAs shows that there are very few periods with no vessels within a radius of 10 km from the recording location. Farther vessels could not be identified as distinct sound events. The LTSAs from the same period inspected in Figure 4.3 can be seen in Figure 4.9. It is clear that Fairplay is quieter than the rest, and it is a clear showcase of how noisy the lower frequencies are in the BPNS. As a showcase of the information one can obtain when visually inspecting LTSAs, other examples of daily LTSAs can be found in the Supplementary Material (see Figure S4.1).

A LTSA example from the Grafton location (20th January 2022) was selected as representative of one day of the dataset (Figure 4.8). Grafton is approximately at 7 km from a major shipping lane. For this reason, there seems to be a constant presence of vessels at that distance. The current presented a high correlation with the lower frequencies (<60 Hz), while the vessels closer than 10 km seemed to increase the received levels at the frequency bands between 50 and 200 Hz.

Figure 4.8 (top) Example of the long-term spectrogram (LTSA) of one full day at Grafton location together with: (middle) the bottom and surface current speed extracted from [19] and (bottom) the distance to the vessels in the region obtained from AIS data. Increased levels between 60 and 150 Hz can be observed whenever a vessel is closer to the recording location. Flow noise can be seen in the lower frequency range ($\leq$ 60 Hz) throughout the day.



4.3.5 Correlation with shipping

During the LTSAs manual analysis, we saw that only ships closer than approximately 10 km generated a visible transient signature which could be individually identified in the LTSA. Further away ships blended in the background. Therefore, only ships closer than 10 km were included in the correlation analysis. The results showed that, for the BPNS, hourly correlations are higher for the total number of ships passing by. The distance-weighted metric from Basan et al. (2024) [7] (see equation 4.3) was less correlated with received levels (see Figure 4.10).

When looking at the correlation values with the number of ships (Figure 4.10), for Buitenratel there are only weak correlations from 50 Hz to 8 kHz. Grafton is also correlated in the same frequency range but the correlations are strong. GardenCity presents a similar pattern to Grafton, but lower frequencies are also correlated with shipping. Belwind does not present strong correlations over the entire frequency range, only weak correlations from 200 Hz to 5 kHz. The distance-weighted metric does not seem to be strongly correlated with the shipping activity in Grafton, but it is weakly correlated with most of the frequency bands for Buitenratel and GardenCity.

4.4 Discussion

As a result of a first qualitative analysis, it can be reported that the BPNS is dominated by anthropogenic sound sources. Furthermore, the analysis given in this chapter proves that even in a small area such as the BPNS the soundscape presents differences depending on the location. As mentioned in the Introduction, this is typical of shallow waters, where acoustic changes in the soundscape can occur at very small spatial scales [8]. The greatest differences observed among locations seem to be driven by water depth and distance to shipping lanes. Furthermore, differences in the soundscape could be seen in the Belwind location, a recording location at a wind farm, and Fairplay, a shallow coastal location far from the shipping lanes.

Areas where only recreational vessels go such as Fairplay, are much quieter than other locations such as Grafton and Faulbaums where fishing might be heavier than in the very coastal region, and cargo ships pass close by. Overall, GardenCity and Grafton, which can be seen as representative of the BPNS, are affected over the entire frequency range by shipping noise. This analysis suggests that locations close to the shipping lane will receive higher sound levels in the shipping frequency band (50 to 200 Hz).

A recent study by Basan et al. (2024) [7] characterized the underwater soundscape of the entire North Sea, including one location in the Belgian Part. Of the 19 locations of the study across the North Sea, the Belgian location was found to

Figure 4.9 Median hourly HMBL plotted as LTSA for all the data from March to May 2021. D stands for deep, M+ for medium-high, M- for medium-low and S for shallow. White spaces represent no data. Faulbaums, Gardencity and Fairplay were recording with a duty cycle of 24h-on 24h-off.

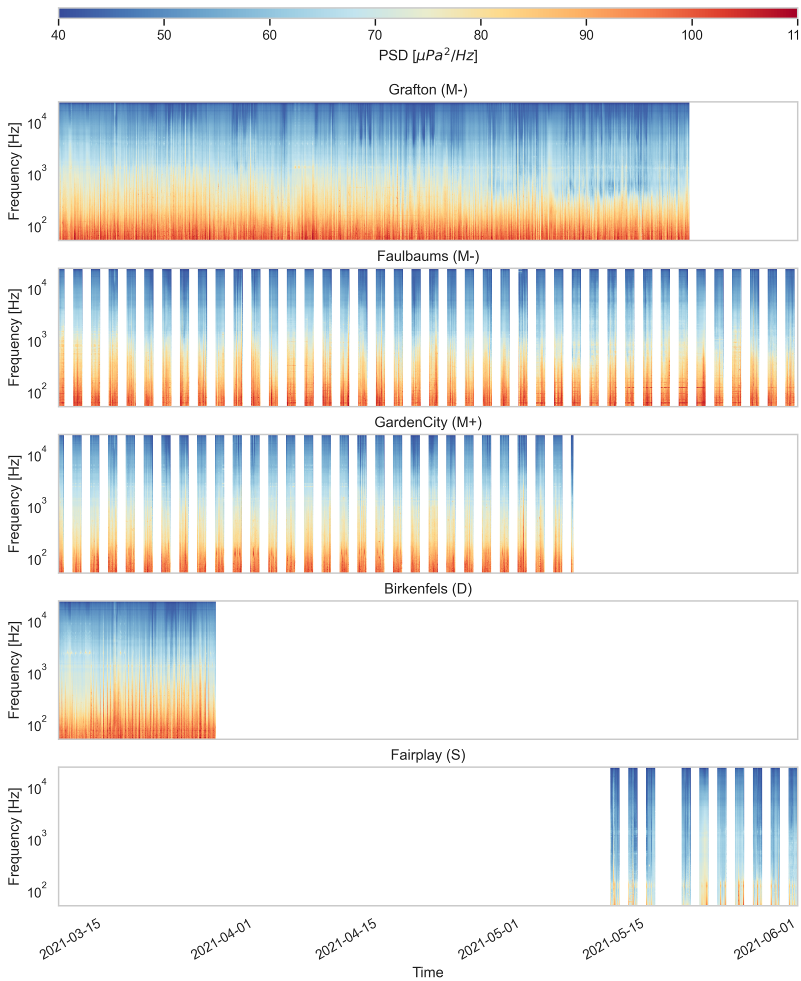
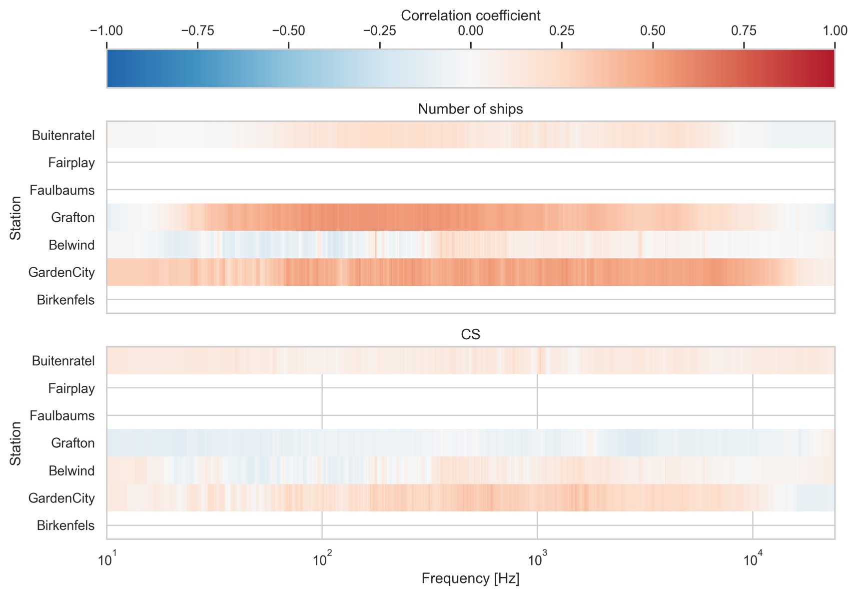


Figure 4.10 Spearman Correlation coefficients for the shipping activity. locations are ordered from shallower (Buitenratel, 7 m) to deeper (Birkenfels, 37.5 m) locations. locations in white represent missing data, as AIS data was only collected since October 2021. CS represents the distance-weighted metric from Basan et al. (2024) [7]



be the loudest. Basan et al. (2024) reported the location of Westhinder to be the location from their study where the most vessel passages were registered within a 35 km radius, and also the one with the maximum bottom current (0.99 ms^{-1}). The challenges, and the environmental and acoustic characteristics reported for the Belgian location in Basan et al. (2024) are consistent with the information reported in Chapters 2 and 3. We can conclude that there are big ($>$ 10 dB) temporal and spatial variations in the BPNS. These differences illustrate the high acoustic complexity of this area over time and space. Consequently, it is not ideal to generalize the measurements of one location for the entire BPNS in such a complex shallow sea.

In areas such as the BPNS where there are no loud biological choruses, quiet biological sounds get masked by vessels and flow-noise. For this reason, these biological sounds are difficult to detect on long-term averages. The analysis of long-term hybrid millidecade bands' mean levels does not provide straightforward insight into biological temporal acoustic patterns. When analyzing the main contributors we analyzed the difference between using mean and median metrics and the effects of short loud sounds on long-term averaging metrics. The results showed that these short loud sounds such as fish sounds do have an influence on the mean but not the median, as already proposed in other studies [26, 27]. Therefore, when computing the hybrid millidecade band (sound pressure) levels and taking the median instead of the mean, the difference between these two measures can be used to detect the presence of short, loud, and transient sounds such as fish sounds. Our results suggest that 1-minute mean HMBL can provide information about frequent sounds such as anthropophony and some physical events, but by themselves provide little information about biophony because no animal choruses have been found in the BPNS so far. The biophony in the BPNS is very sparse-occurring and the produced sounds are short, so no daily patterns were observed in the lower frequencies at 1-minute average resolution except the ones correlated to flow-noise. However, the difference between the mean and the median could be reliably used to detect the presence of certain short and loud events such as those from fish sounds, as already proposed by previous studies [26, 27]. This would not be the case for biological choruses, as they are usually constant spanning a longer time frame, and therefore would affect the median and the mean both equally. This difference threshold can be used to characterize general patterns of the soundscape and quantify the presence of salient sound events. Further work is necessary to assess if it can be used to detect biological sounds in long-term recordings.

All the studied environmental parameters except rain could be correlated with multiple frequency bands. The lack of correlation between rain and HMBL could be due to the lack of reliable environmental data, as the meteorological buoys are several km away from the recording locations. It could also be that rain is more correlated with higher frequencies than the ones studied here, or that the correlation does not exist because other sound sources are louder. Further manual analysis of the data to determine rain periods would be necessary to understand this contribution.

Although our findings are consistent with the big picture shown in Basan et al. (2024) [7], the frequency ranges differ, and they are not consistent among locations. The obtained correlations were not the same at all the locations. The differences seemed to be driven by location-specific characteristics, such as water depth or present human activities.

Low-frequency current-related sound seems to be present in all the locations. Flow noise has a strong influence mostly below 60 Hz (except for Fairplay, up to 500 Hz). For this reason, on all the pre-processing done in this dissertation we apply a high-pass filter to all our data at 50 Hz, in part because of flow noise and in part because of the floor noise of the hydrophone. A clear increase in sound levels of low frequencies can also be seen when the current speed is higher. This could be because the higher current causes higher flow-noise and masks all the sounds below 100 Hz. However, as proposed in Rogers et al. (2021) [28], instead of flow-noise, this could be sound coming from far away sound sources such as big river mouths, which are also synchronized with the tide. However, further research involving the measurement of particle motion would be necessary to determine if the current-related sound levels in the low frequencies are created by pseudo-noise generated by current (flow-noise) or actual sound sources coming from far such as tidal flows at river mouths [28].

Compared to the data from Basan et al. (2024) for the Belgian location, where they found weak correlations with the current velocity up to 500 Hz, the correlation of the lower frequencies of our data with current velocity was mostly below 100 Hz. However, for location Fairplay, the correlation was high up to 200 Hz. These results suggest that either the data acquired within the LifeWatch Acoustic Broadband Network [10] are less influenced by flow-noise than the ones collected in Westhinder during the JOMOPANS project, or other sources are masking the flow-noise at those frequency ranges. Furthermore, for the deeper locations (GardenCity and Birkenfels, 28 and 37.5 m, respectively) current also showed a medium correlation with higher frequencies (above 10 kHz), possibly due to the high-frequency sounds generated by the current moving the sediment. This was also found in the Belgian location of Basan et al. (2024).

The fact that current was only weakly correlated with lower frequencies in the Buitenratel location could be because flow-noise is constantly present there due to the shallowness of the location, or because the water depth at that location is too shallow to allow the propagation of these low-frequency far away sounds. For the other locations, because there seems to be only a correlation with current but not with the tide, the low frequency received levels are likely to be flow-noise.

Another finding is that in very shallow locations (average 7 m deep), the tide seems to have an influence on received sound pressure at around 100 Hz (from 50 to 200 Hz approximately, see Figure 4.7). This is probably because of the difference in depth created by the tide, which allows for the propagation of these frequencies that would otherwise be attenuated or cut off. When studying the frequency distribution at different tide conditions in Buitenratel we could further elaborate on the influence of these parameters in the received levels. The received levels at high tide were a lot higher, both in the lower and the higher frequencies. The increase at low and mid frequencies can be explained by the change in depth due to the tide. High tide would allow the propagation of frequencies that otherwise have a too large wavelength to be propagated in such a shallow environment (< 10 m). The increase at higher frequencies could be attributed to sediment transport noise.

The noise caused by the wind could also be identified by looking at the increase in spectrum levels, and it matched the expected sound levels from the Wenz curve (1962) for shallow waters with heavy traffic. The correlation with wind appears to be strong for all the locations above 4 kHz, and weak from 1 to 4 kHz. At frequencies above 20 kHz, for deeper locations, the correlation seems to weaken. Correlation with wind was also high at lower frequencies around 100 Hz at the location Belwind, which is at the offshore wind farm. When looking at the frequency distribution of Belwind at different wind speeds, we get extra insight into the effects of the wind on the received levels. Whenever the wind speed increases there is a clear increase in recorded sound levels at frequencies below 60 Hz and from 200 Hz to 10 kHz. The increase in received sound pressure levels at lower frequency bands could be attributed to the operating wind turbines, which are known to produce sound in that frequency range. Furthermore, very narrow band frequency peaks at 4 and 6 kHz seem to diminish when the wind is stronger. These tonal sounds are not typically produced by wind, however strong, so they are most likely human-made. Because they are only present in calm sea conditions, they could potentially be associated with shipping sounds coming from maintenance work on the turbines. These activities happen less often in rough sea state conditions. This could therefore be an explanation of why these narrow frequency bands are quieter when the wind is stronger.

The correlation of median hourly hybrid millidecade band (sound pressure) levels with shipping is present from 60 Hz to 8 kHz. The highest correlation was between 100 and 1000 Hz for Grafton and GardenCity. These results are in line with the ones presented in Basan et al. (2024) [7], where they found 630 Hz to be the frequency band more correlated with shipping activity. However, in our analysis, the correlation was greater when counting the number of vessels than when using the distance-weighted metrics proposed by Basan et al. (2024). Furthermore, we saw that further away vessels ($>$ 10 km) were not correlated with the received HMBL. Several reasons might explain why the lower frequencies, typically associated with peak frequencies produced by vessels, do not correlate so strongly with shipping presence. One reason can be because the current (mainly flow-noise) is strongly influencing the lower frequencies ($<$ 100 Hz), as already mentioned above. Another reason would be that because of the business of the area and the complex propagation reflections, those frequencies are always loud regardless of how many vessels are close by. It could also be that lower frequencies are not propagating due to the shallowness of the area and the sandbanks present. The fact that the analyzed data only contained vessels with active AIS can also be a factor for the lack of correlation. In shallow areas, vessels that do not require AIS by law are more abundant (such as recreational vessels), potentially contributing to the sound field without being accounted for on the shipping metrics [29]. Finally, in shallow waters, there are shipping activities that regularly produce noise in higher frequency ranges ($>$ 4 kHz) when not underway. This can be for example when doing other activities requiring Dynamic Positioning. These activities would not contribute to the lower frequency range when close to the recording position.

In conclusion, in shallow waters such as the BPNS, 63 and 125 Hz might not be the best 1/3 octave bands to monitor shipping noise, as proposed by the MSFD [2], unless specific measures are taken. An approach to remove flow-noise would be to record always with multiple synchronized hydrophones simultaneously and remove the non-coherent part of the signal [30]. However, this is of little practical value when flow-noise exceeds ambient noise levels by more than 50 dB [31]. Another solution would be to quantify shipping noise only during slack tide, but that would lead to incomplete data. Adding an extra measuring frequency of around 600 Hz would be an additional value to the reporting of shipping noise on shallow waters, as so would be reporting current velocity together with sound levels. Finally, we adhere to the proposal made by Hermannsen et al. (2019) [29] where they propose that to improve vessel noise models and impact assessments in shallow waters it is necessary that faster and more powerful recreational vessels also carry AIS transmitters.

4.5 Conclusions

This chapter describes the main contributors to the BPNS soundscape based on traditional soundscape analysis metrics such as LTSA visualization, broadband sound pressure comparison, spectrum comparison, or correlation with environmental parameters. With this analysis, we reveal the distribution of sound levels at different stations of the Belgian Part of the North Sea. All the obtained measures can be used as a baseline to compare for future developments or conservation measures. The provided soundscape analysis can also be used to compare the BPNS to other regions of the world. Using hybrid standardized 1-minute hybrid millidecade band (sound pressure) levels [9] is key to making ambient noise trends comparable between sites.

In an area so intensively used such as the BPNS, it becomes essential to monitor the human contribution to underwater sound to asses long-term trends and predict possible negative effects on the ecosystem. After seeing how the spectrum of all the locations is dominated by shipping sound, it is necessary to evaluate different measuring and reporting strategies to measure the shipping contribution to the underwater soundscape.

References

[1] IQOE – Marine Bioacoustical Standardization Working Group. *IQOE – Inventory of Existing Standards and Guidelines Relevant to Marine Bioacoustics*. Technical report, IQOE, 2021.

[2] M. Tasker, M. Amundin, M. Andre, A. Hawkins, W. Lang, T. Merck, A. Scholik-Schlomer, J.Teilmann, F. Thomsen, S. Werner, and M. Zakharia. *Marine Strategy Framework Directive : Task Group 11 Report : Underwater Noise and Other Forms of Energy, April 2010*. Publications Office, LU, 2010.

[3] Joint Research Centre (European Commission), J. F. Borsani, M. Andersson, M. André, A. Azzellino, M. Bou, M. Castellote, L. Ceyrac, D. Dellong, T. Folegot, D. Hedgeland, C. Juretzek, A. Klauson, R. Leaper, F. Le Courtois, A. Liebschner, A. Maglio, M. A, A. Norro, A. Novellino, O. Outinen, A. Popit, A. Prospathopoulos, F. Thomsen, T. J, P. Vukadin, L. Weilgart, P. Sigray, J. N. Druon, M. Casier, and G. Hanke. *Setting EU Threshold Values for Continuous Underwater Sound*. Publications Office of the European Union, 2023.

[4] R. P. Schoeman, C. Erbe, G. Pavan, R. Righini, and J. A. Thomas. *Analysis of Soundscapes as an Ecological Tool*. In C. Erbe and J. A. Thomas, editors, Exploring Animal Behavior Through Sound: Volume 1: Methods, pages 217–267. Springer International Publishing, Cham, 2022.

[5] N. D. Merchant, T. R. Barton, P. M. Thompson, E. Pirotta, D. T. Dakin, and J. Dorocicz. *Spectral Probability Density as a Tool for Ambient Noise Analysis*. The Journal of the Acoustical Society of America, 133(4):EL262–EL267, April 2013.

[6] M. Picciulin, M. Bolgan, N. Rako-Gospić, A. Petrizzo, M. Radulović, and R. Falkner. *A Fish and Dolphin Biophony in the Boat Noise-Dominated Soundscape of the Cres-Lošinj Archipelago (Croatia)*. Journal of Marine Science and Engineering, 10(2):300, February 2022.

[7] F. Basan, J. G. Fischer, R. Putland, J. Brinkkemper, C. A. F. de Jong, B. Binnerts, A. Norro, D. Kühnel, L. A. Ødegaard, M. Andersson, E. Lalander, J. Tougaard, E. T. Griffiths, M. Kosecka, E. Edwards, N. D. Merchant, K. de Jong, S. Robinson, L. Wang, and N. Kinneging. *The Underwater Soundscape of the North Sea*. Marine Pollution Bulletin, 198:115891, January 2024.

[8] M. F. McKenna, S. Baumann-Pickering, A. C. M. Kok, W. K. Oestreich, J. D. Adams, J. Barkowski, K. M. Fristrup, J. A. Goldbogen, J. Joseph, E. B. Kim, A. Kügler, M. O. Lammers, T. Margolina, L. E. Peavey Reeves, T. J. Rowell, J. A. Stanley, A. K. Stimpert, E. J. Zang, B. L. Southall, C. C. Wall, S. V.

Parijs, and L. T. Hatch. *Advancing the Interpretation of Shallow Water Marine Soundscapes*. Frontiers in Marine Science, 8, 2021.

[9] J. L. Miksis-Olds, P. J. Dugan, S. B. Martin, H. Klinck, D. K. Mellinger, D. A. Mann, D. W. Ponirakis, and O. Boebel. *Ocean Sound Analysis Software for Making Ambient Noise Trends Accessible (MANTA)*. Frontiers in Marine Science, 8:703650, August 2021.

[10] C. Parcerisas, D. Botteldooren, P. Devos, and E. Debusschere. *Broadband Acoustic Network Dataset*, 2021.

[11] C. Parcerisas. *Lifewatch/Pypam: Pypam, a Package to Process Bioacoustic Data*. Zenodo, February 2022.

[12] G. M. Wenz. *Acoustic Ambient Noise in the Ocean: Spectra and Sources*. The Journal of the Acoustical Society of America, 34(12):1936–1956, December 1962.

[13] S. B. Martin, B. J. Gaudet, H. Klinck, P. J. Dugan, J. L. Miksis-Olds, D. K. Mellinger, D. A. Mann, O. Boebel, C. C. Wilson, D. W. Ponirakis, and H. Moors-Murphy. *Hybrid Millidecade Spectra: A Practical Format for Exchange of Long-Term Ambient Sound Data*. JASA Express Letters, 1(1):011203, January 2021.

[14] M. P. Rey Baquero, C. Parcerisas, K. Seger, C. Perazio, N. Botero Acosta, F. Mesa, A. L. Acosta, D. Botteldooren, and E. Debusschere. *Comparison of Two Soundscapes: An Opportunity to Assess the Dominance of Biophony Versus Anthropophony*. Oceanography, pages 62–65, December 2021.

[15] J. P. Ryan, J. E. Joseph, T. Margolina, L. T. Hatch, A. Azzara, A. Reyes, B. L. Southall, A. DeVogelaere, L. E. Peavey Reeves, Y. Zhang, D. E. Cline, B. Jones, P. McGill, S. Baumann-Pickering, and A. K. Stimpert. *Reduction of Low-Frequency Vessel Noise in Monterey Bay National Marine Sanctuary During the COVID-19 Pandemic*. Frontiers in Marine Science, 8, 2021.

[16] M. C. Amorim. *Diversity of Sound Production in Fish*. Communication in Fishes, 1:71–105, January 2006.

[17] K. S. Cole, editor. *Reproduction and Sexuality in Marine Fishes: Patterns and Processes*. Univ. of California Press, Berkeley, 2010.

[18] K. Lisa Yang Center for Conservation Bioacoustics at the Cornell Lab of Ornithology. *Raven Pro: Interactive Sound Analysis Software (Version 1.6.4)*. The Cornell Lab of Ornithology, 2023.

[19] MFC ODNature RBINS. *Physical State of the Sea - Belgian Coastal Zone - COHERENS UKMO*, 2023.

[20] Agentschap Maritieme Dienstverlening en Kust. *Meetnet Vlaamse Banken*.

[21] C. Spearman. *The Proof and Measurement of Association between Two Things*. The American Journal of Psychology, 15(1):72–101, 1904.

[22] A. MacGillivray and C. de Jong. *A Reference Spectrum Model for Estimating Source Levels of Marine Shipping Based on Automated Identification System Data*. Journal of Marine Science and Engineering, 9(4):369, April 2021.

[23] J. Tougaard, L. Hermannsen, and P. T. Madsen. *How Loud Is the Underwater Noise from Operating Offshore Wind Turbines?* The Journal of the Acoustical Society of America, 148(5):2885–2893, November 2020.

[24] Commission Directive (EU). *Commission Notice on the Threshold Values Set under the Marine Strategy Framework Directive 2008/56/EC and Commission Decision (EU) 2017/848*, March 2024.

[25] P. T. Madsen, M. Wahlberg, J. Tougaard, K. Lucke, and P. Tyack. *Wind Turbine Underwater Noise and Marine Mammals: Implications of Current Knowledge and Data Needs*. Marine Ecology Progress Series, 309:279–295, March 2006.

[26] T.-H. Lin, Y. Tsao, Y.-H. Wang, H.-W. Yen, and S.-S. Lu. *Computing Biodiversity Change via a Soundscape Monitoring Network*. In 2017 Pacific Neighborhood Consortium Annual Conference and Joint Meetings (PNC), pages 128–133, November 2017.

[27] G. Bazile Kinda, Y. Simard, C. Gervaise, J. I. Mars, and L. Fortier. *Under-Ice Ambient Noise in Eastern Beaufort Sea, Canadian Arctic, and Its Relation to Environmental Forcing*. The Journal of the Acoustical Society of America, 134(1):77–87, July 2013.

[28] P. Rogers, E. Debusschere, D. de Haan, B. Martin, and H. Slabbekoorn. *North Sea Soundscapes from a Fish Perspective: Directional Patterns in Particle Motion and Masking Potential from Anthropogenic Noise*. The Journal of the Acoustical Society of America, 150(3):2174–2188, September 2021.

[29] L. Hermannsen, L. Mikkelsen, J. Tougaard, K. Beedholm, M. Johnson, and P. T. Madsen. *Recreational Vessels without Automatic Identification System (AIS) Dominate Anthropogenic Noise Contributions to a Shallow Water Soundscape*. Scientific Reports, 9(1):15477, October 2019.

[30] J. Y. Chung. *Rejection of Flow Noise Using a Coherence Function Method*. Journal of the Acoustical Society of America, 62(2):388–395, 1977.

[31] C. Bassett, J. Thomson, P. H. Dahl, and B. Polagye. *Flow-Noise and Turbulence in Two Tidal Channels*. The Journal of the Acoustical Society of America, 135(4):1764–1774, April 2014.

5

Understanding soundscapes by categorization

This chapter is adapted from two merged publications:

- *Parcerisas, C.; Roca, I.T.; Botteldooren, D.; Devos, P.; Debusschere, E. Categorizing Shallow Marine Soundscapes Using Explained Clusters. J. Mar. Sci. Eng. 2023, 11, 550.* https://doi.org/10.3390/jmse11030550
- *Parcerisas, C., Roca, I.T., Botteldooren, D., Devos, P., & Debusschere, E. Clustering, categorization, and mapping of shallow coastal water soundscapes. Forum Acusticum, Sept 2023, Turin, Italy.* https://www.doi.org/10.61782/fa.2023.1070

Summary

Natural marine soundscapes are being threatened by increasing anthropic noise, particularly in shallow coastal waters. To preserve and monitor these soundscapes, understanding them is essential. There are several ways to do so, each with their limitations. To complement the description provided in Chapter 4, in this Chapter we propose a new method for semi-supervised categorization of shallow marine soundscapes, with further interpretation of these categories according to concurrent environmental conditions using SHAP. This approach is focused on continuous sound or sounds that are repeated frequently in a (complex) sequence (e.g., a fish

chorus or breaking waves), including the combination of all sounds that occur under certain conditions at specific places. We tested this methodology in the two different datasets explained in Chapter 2. In this section, we approach the same research question than in Chapter 4, but using ML tools instead of traditional soundscape analysis:

Can we holistically describe and characterize the soundscape of the BPNS, and identify the external environmental factors (such as wind, rain, and currents) that influence it?

5.1 Introduction

Studying marine soundscapes holistically, instead of focusing on specific sound events, can provide us with information at habitat levels. In a human-centered way, ISO 12913-1 [1] includes perception and understanding of the definition of soundscape; therefore, a soundscape is defined by how it is perceived and understood. Because humans do not spend a significant amount of their lives underwater, the human perception of underwater soundscapes is not representative. However, one could extend the perception concept to the underwater world by including the perception of a variety of species or, in other words, considering ecological relevance.

Understanding the acoustic environment in an machine learning context is also referred to as acoustic scene recognition, which refers to one of the important functions of the acoustic scene: supporting context and situation awareness. Acoustic scene identification does not need to be based on specific event recognition, but can be based on an overall sound impression [2]. In waters where visibility is low, asserting situation and context may be relevant for many species. The holistic soundscape contains many different sounds, and some species may use auditory stream segregation to disentangle these sounds depending on their relevance. However, in a multi-species and ecological approach, a more holistic technique may be more appropriate than methods that separate and classify individual sound sources [3].

We propose an unsupervised method of assigning a label to different underwater acoustic scenes, with the aim of categorizing them. Previous studies have shown how different environments have distinct acoustic signatures that cluster together [4]. Other studies have successfully used unsupervised clustering algorithms to discriminate between terrestrial ecological habitats [5], and to test if the combination of several acoustic indices could capture the difference in the spectral and temporal patterns between under-shelf and pelagic marine soundscapes [6]. Sethi et al. (2020) [4] proposed unsupervised clustering as a method of detecting acoustic anomalies in natural terrestrial soundscapes. Michaud et al. (2023) [7] and Ulloa et al. (2018) [8] used unsupervised clustering on terrestrial soundscapes, to group pre-selected regions-of-interest from terrestrial soundscapes. Clustering has also been used as a tool to speed up labeling efforts [9], and has been proposed as a method of monitoring changes in the acoustic scene in terrestrial habitats [10]. However, in these cases the obtained categories had to be manually analyzed to have an ecological meaning.

Knowledge about the marine acoustic scene is still limited, and there are no defined marine soundscape categories because, during human history we did not need to name them. Currently marine soundscapes are named depending on the habitat they represent. To further understand the categories we obtain in an unsupervised way, we propose to explain them according to the spatiotemporal context in which they occur. This is the first study to describe soundscape categories in an automatic way, and also the first to consider the time component in soundscape categorization. The proposed solution is particularly useful in areas where the underwater acoustic scenes (soundscapes) have not yet been described. This is often the case in areas where the water is often too turbid to employ camera or video sampling techniques, and where the sound signatures of most of its sound sources (especially biological ones) are not known. To understand when and where these categories occur, we linked them to environmental data, using a supervised machine learning model. Interrelationships were checked, using SHAP [11]. These tools allowed for assessing which of the environmental features were important in predicting each class, and had already been successfully implemented in some ecology fields [12]. The classes that could not be explained from the environmental parameters were not considered soundscape classes, but just a certain sound class. Afterwards, by interpreting the SHAP values outcome, we could infer when and where these categories were found. The tools for the machine learning models' interpretation led to understanding which environmental parameters were representative for different soundscapes, and what differentiates soundscapes from one another, ecologically. Furthermore, it allowed us to predict which soundscape category would be predominant and map it across the BPNS, and dominant frequency maps could be predicted.

The relevant environmental conditions of each cluster could then be used to describe and understand each category, without the need for a large annotation effort: this helped explain acoustic dissimilarity between habitats, and also provided a baseline for the soundscapes in their current state. In conclusion, we propose an automatic solution to extract relevant ecological information from underwater soundscapes, by assessing which environmental factors are contributing to the soundscape, and quantifying their significance. This should allow for monitoring of relevant ecological processes and major changes in the underwater ecosystems. Furthermore, we propose a semi-supervised method by which to remove artifacts from a dataset before processing. The analysis was performed on two datasets recorded in the BPNS, one based on short moving recordings (see section 2.3) and another one based on long-term recordings (see section 2.2). Different temporal resolutions were used for the two datasets to showcase two different approaches to soundscape analysis: a more detailed one (Drifts), where the focus lays on small changes in time and space, and a more global one (Stationary) where the focus lays on global patterns and differences.

5.2 Methods

In this section we present how the acoustic data was processed for both datasets, and how the environmental data was obtained. Then we explain how did we obtained the acoustic categories in an unsupervised way, and we proceed to show the proposed methodology to link the obtained categories to the environmental data using SHAP values. Finally, we describe a method to use the link between the environmental parameters and the categories to predict the soundscape category distribution, and how this distribution can be used to estimate sound levels at a certain frequency band.

5.2.1 Acoustic Processing - drifting data

The acoustic processing was applied to the data from the drifting dataset described in Section 2.3. Acoustic changes in the soundscape when drifting were expected to be found at a small spatial (meters) and temporal (seconds) resolution: for this reason, the data were processed in time windows of 1 s (hereafter, s_i), with no overlap. Then, these 1-second segments, s_i, were arranged in groups of 5. These groups of 5 had an overlap of 60%, (hereafter, S_j) (Equation 5.1). The time window and the overlap were chosen so that the spatial resolution would be of minimum 2 samples every 5 m, considering an average drifting speed of $1\text{m}s^{-1}$.

$$S = \begin{bmatrix} S_0 \\ S_1 \\ \dots \\ S_j \\ \dots \\ S_n \end{bmatrix} = \begin{bmatrix} s_0 & s_1 & s_2 & s_3 & s_4 \\ s_2 & s_3 & s_4 & s_5 & s_6 \\ \dots & & \dots & & \dots \\ s_i & & \dots & & s_{i+4} \end{bmatrix} \tag{5.1}$$

Where,
S_j is one aggregated segment of 5 x 1-second segments.

All the acoustic processing was done using pypam [13], an open-source python tool for processing acoustic files in chunks. Each recorded file was converted to sound pressure, using the calibration factor for the equipment used. For the SoundTrap, the calibration was done according to the value given by the manufacturer. The B and K equipment was calibrated by generating a calibration tone at the beginning of each file, and was afterwards removed from the file. The rest of the recording was processed according to the obtained calibration value. The sound pressure values obtained by the two instruments recording simultaneously were compared, to make sure the calibrations were accurate. Per time window, a Butterworth band pass filter of order 4 was applied between 10 and 20,000Hz, and the signal was down-sampled to twice the high limit of the filtered band. The frequency band was chosen to cover all the biophony and geophony of interest that we could record, given the sampling rate limitations of the equipment. After filtering, the root mean squared value of the sound pressure of each one-third octave band (base 2) was computed per each 1-second segment s_i. Then, every 5 s were concatenated, to create an acoustic sample, S_j, which resulted in 5×29 one-third octave bands sound pressure levels per time window S_j (5.1). Each sample S_j was flattened to an array of 145×1, for further analysis. One-third octave bands sound pressure levels are considered to be appropriate for most considerations of environmental impact [14], and they are commonly used for soundscape analysis [15]. Furthermore, they had previously been used to describe soundscapes [16, 17, 18].

5.2.2 Acoustic Processing - Stationary data

The acoustic processing explained in Section 5.2.1 was applied to the data from the stationary network defined in Section 2.2. 14 different deployments from 6 different stations (see Figure 2.2, in Chapter 2) were considered, ranging between March 2021 and September 2022. Because of computing CPU limitations, 10 minutes of every recorded hour were chosen randomly for the analysis.

All acoustic data were processed using the Python package pypam [13]. Data with a higher sampling rate than 48,000 kS/s were filtered with a low-pass Butterworth order 4 filter with a lower limit of 10 Hz and an upper limit of 24,000 Hz and then downsampled to 48,000 kS/s to match the rest of the data. Then the filtered audio data records were processed to 1-minute hybrid millidecade bands (sound pressure) levels (HMBL) [19] because their frequency resolution is well-suited for long-term spectral averages and soundscape comparisons [20]. The HMBL were computed as explained in Section 4.2.1. The HMBL were used as an input for the dimension reduction. In this first analysis, a 1-minute non-overlapping window temporal resolution was chosen. This implies however that short-term patterns smear out and cannot be distinguished. In this approach, a sample S_j did not have any temporal component, and was only one spectrum per minute.

5.2.3 Environmental Data

A summary of the considered environmental parameters is shown in Table 5.1. These included (1) season, which was computed as the week number, (2) bathymetry, (3) moon phase, considered as the angle giving the difference between the geocentric apparent ecliptic longitudes of the Moon and Sun, where 0 is New Moon and π is Full Moon, (4) day moment (Day, Night, Twilight), (5) type of substrate on the sea bottom, (6) seabed habitat output of the 2021 EUSeaMap broad-scale predictive model, produced by EMODnet Seabed Habitats[1] [21], (7) the benthic community type predicted in Derous et al. (2007)[2], (8) the modeled surface salinity, (9) the modeled surface temperature, (10) the total surface and the bottom current speed obtained from computing the norm between eastward and northward sea water velocity, (11) the surface elevation above mean sea level (m.a.s.l), (12) the distance to the closest known shipwreck in the BPNS, and (13) the monthly averaged shipping density from EMODnet Human Activites [22]. The environmental variables used for each dataset were adjusted to accommodate the shift from a higher spatial coverage and lower temporal coverage (drifts dataset) to a higher temporal coverage and a lower spatial coverage (stationary dataset). The variables used for each dataset are marked with a D (drifts) or an S (stationary) in Table 5.1. The list of all the considered parameters is referred to as $F = [F_0, ..., F_n]$ in Figure 5.1.

For the drifts dataset, each acoustic sample, S_j, was matched to the closest point in time stored in the GPS tracker. If S_j did not have a GPS point closer than 5 s in time, it was eliminated from the dataset. For the stationary dataset the location was assigned per station. The selected environmental data were paired to each sample, S_j, using the python package bpnsdata [23]. We will refer to the environmental data paired to sample S_j as E_j. Some data were not available in the public environmental datasets during some of the recording periods, so any S_j with a corresponding incomplete E_j was also removed from the dataset.

The cyclic variables, season and moon phase, were split into their cosine and their sine values, with the intention of representing the cyclic continuation of the values. Categorical variables were encoded using the one-hot-encoder[3] function of sklearnDF [24]. Distance to closest shipwreck was converted into logarithmic scale, as its effect was expected to be relevant only at a very close distance, because of the short propagation of the calls of the expected benthic fauna [25]. Furthermore, the decay of sound pressure with distance was expected to be logarithmic. Bathymetry was converted to a positive value. See Table 5.1 for more details.

[1] see Figure 2.3 for a map of the sediment type

[2] see Figure 2.3 for a map of these habitats

[3] One-hot encoding is a technique used in machine learning to represent categorical variables as binary vectors. Each category is converted into a unique vector where one element is "1" (indicating the presence of the category) and all other elements are "0".

Table 5.1: Summary of the all the available environmental variables. m.a.s.l. stands for meters above sea level. D stands for the drifting dataset and S for the stationary dataset.

Parameter	**Encoding**	**Resolution**	**Source**	**Dependency**	**Dataset**
Season	week_n_cos: winter (1) vs summer (-1), week_n_sin: spring (1) vs autumn (-1).	Weekly	NA	Time	D
Bathymetry [m]	Converted to positive.	1/16 arc min	[26]	Space	D,S
Moon phase [rad]	growing_moon: growing (1) vs decreasing (-1) new_moon: new moon (1) vs full moon (-1).	Continuous	[27]	Space, Time	D,S
Moment of the day	Categorical encoding				
Shipping Density [$1\text{km}^{-2}\text{month}^{-1}$]	None	1 km^2, 1 month	[22]	Space, Time	D,S
Substrate	Categorical encoding	0.00104	[21]	Space	D
Seabed habitat					
Benthic habitat	Categorical encoding	3×3 km	[28]	Space	D,S
Salinity [PSU]	None	lon: 0.042, lat: 0.083, time: 1h	[29]	Space, Time	D
Surface Temp. [K]					
Current [ms^{-1}]					
m.a.s.l. [m]					
Shipwreck dist. [log(m)]	$log_{10}(distance)$	NA	[30]	Space	D
Dist. to coast [m]	None	Continuous	[31]	Space	D
Wave Height [m]	None	lon: 0.042, lat: 0.083, time: 1 h	[32]	Space, Time	S
Wave period [s]					

5.2.4 Acoustic Categorization

5.2.4.1 Dimension Reduction and Acoustic Dissimilarity

A dimension reduction algorithm was applied to the processed acoustic data (5×29 spectral levels). The resulting clusters were visually presented in a 2D space for easy visual inspection. No linear combination of the features was expected to represent the data distribution properly, so Uniform Manifold Approximation and Projection (UMAP) was considered an appropriate dimension reduction algorithm to deal with the non-linearity of the data [33]. UMAP is a dimension reduction technique that can be used for visualization, similar to t-SNE [34]. It models the manifold with a fuzzy topological structure. The embedding is found by searching for a low dimensional projection of the data, which has the closest possible equivalent fuzzy topological structure. UMAP was preferred over t-SNE, because it better preserves the global structure of the data [35]. UMAP has successfully been used in several bioacoustics studies [36, 37, 38]. Different configurations of the UMAP parameters were tested, to try to achieve an optimal distribution of the data, where acoustic sample points were distributed in separated clusters. Using UMAP, the data were represented by more similar points closer to each other, and by more different points further apart: the idea was to exaggerate local similarities, while also keeping a structure for global dissimilarities.

Similar sounds were thus expected to create clusters, and clusters further apart to represent more different sounds than clusters closer together, in the 2D space [4]. The 2D dimension could then be used to compute distances between clusters, as a measure of acoustic dissimilarity between two groups of sounds. However, not all of these clusters represented a soundscape, because specific sounds which had occurred in the dataset multiple times, but were not related to their recording location or time, also formed clusters; therefore, some of the obtained clusters might not represent a soundscape, but only a certain foreground sound. We tackled this distinction using SHAP, and we considered soundscape-clusters to be the ones closely linked to the environmental parameters. Clusters not correlated to the environmental parameters were not considered to be soundscape classes in this chapter.

5.2.4.2 Data Cleaning: Artifacts Removal

During underwater acoustic deployments, a lot of artifacts can affect the recordings: these do not contain any ecologically relevant information, nor do they mask existing information, and thus they need to be removed. They can be caused by small particles hitting the hydrophone, or by sounds generated by the deployment structure or the instrument itself. In this chapter, a semi-supervised approach was applied, to remove the artifacts. All the recordings made with the B and K equipment were manually checked for artifacts segments, and were annotated in a non-detailed fashion. Visual inspections of 10-minute spectrograms were conducted in Audacity [39], and artifacts that were clearly visible were annotated. Each annotation comprised the start and end of the artifact event, and the designated label: *clipping; electronic noise; rope noise; calibration signal; and boat noise*. All the samples S_j which contained more than 1 s of artifact data were labeled with the corresponding tag.

These labels were then plotted in the UMAP space, to check visually if there was any visible clustering. Then, all the data embedded in the 2D space were clustered using Density-Based Spatial Clustering of Applications with Noise (DBSCAN), using the scikit-learn package [40]. This algorithm finds core samples of high density, and expands clusters from them; it is appropriate for data which contain clusters of similar density. Samples that are not close enough to any cluster were considered ”noise”, and therefore are not be classified as one particular class. After clustering, the percentage of artifact samples in each of the resulting clusters was assessed. If the ratio of artifact samples in that cluster exceeded 2 times the ratio of total samples labeled as artifacts in the dataset (Equation (5.2)), it was considered to be an artifact-cluster, and was therefore removed from the dataset. Consequently, all the samples that were clustered in an artifact-cluster, but which had not been manually labeled as artifacts, were also considered artifacts (Figure 5.2). The DBSCAN ”noise” class was also considered a cluster, and treated accordingly.

$$Ni_{\text{artifacts}}/Ni > 2N_{\text{artifacts}}/N_{\text{total}} \tag{5.2}$$

where Ni represents the samples classified as cluster i, and underscore artifacts means samples classified as any of the artifact labels.

5.2.4.3 Acoustic Categories

The UMAP dimension reduction was applied for a second time, to the clean dataset: this was done because the artifact clusters were expected to be acoustically very dissimilar to the environmental clusters; therefore, re-computing the UMAP without them would help obtain clearer distinctions between environmental clusters. The data obtained in the new, embedded 2D space were clustered again, using DBSCAN. Samples classified as noise were removed from the analysis, as they were considered outliers, and could not be classified as one particular acoustic category. The obtained classes after the noise reduction were considered as the different acoustic categories present in the dataset. A diagram of the workflow for artifact removal and final clustering can be seen in Figure 5.2.

5.2.5 Characterization of Acoustic Classes

To understand which environmental parameters were correlated to the categorized acoustic environments (clusters), and could be used to describe them, a Random

Forest (RF) Classifier was built, with the environmental variables as the independent variables, and the obtained acoustic clusters as the dependent variable. The input variables were encoded, as described in Section 5.2.3. The RF was subsequently analyzed, using the inspector system from the FACET gamma package, which is based on SHAP (SHapley Additive exPlanations) [41]. SHAP is a method of explaining individual predictions of instance X, by computing the contribution of each feature to the prediction. The SHAP explanation method can be used to understand how each feature affects the model decisions, by generalizing local relationships between input features and model output. The python SHAP library contains an algorithm called TreeExplainer, made specifically to interpret tree-based Machine Learning models [42]. For this analysis, the TreeExplainer approach was used, considering a marginal distribution (interventional feature perturbation). However, the output of the SHAP analysis can give misleading conclusions if redundant variables are present in the dataset, because the importance of variables that share information will be split between them, as their contribution to the model is also split: the real impact of that redundant pair—and the ecological phenomena they represent—will therefore not be assessed. To overcome this problem, the model was evaluated for redundancy between variables [43]. In facet-gamma, redundancy quantifies the degree to which the predictive contribution of variable u uses information that is also available through variable v. Redundancy is a naturally asymmetric property of the global information that feature pairs have for predicting an outcome, and it ranges from 0% (full uniqueness) to 100% (full redundancy).

The RF was trained multiple times in an iterative way, where one of the possible redundant variables was removed at each iteration, as displayed in Figure5.1. At the start of each iteration, the Boruta[4] algorithm [44] was used, to check if any of the variables were not relevant. Then, to find out the best hyper-parameters for the algorithm, a grid search with cross-validation with 5 folds was run on all the data, with accuracy as the scoring metric. The parameters of the grid can be seen in Table S5.2. The results given for all the models are considering the cross-validation approach.

[4] Boruta is a feature selection algorithm which trains certain ML models (RF by default) while permuting the values of the input features, and evaluates the change in feature importance. If this change is not significant, the feature is discarded as noise.

Figure 5.1 Diagram of the trained Random Forest and the environmental characteristics assessment: k is the number of categories; i is the iteration number; F are all the available variables; n is the number of available variables; p_i is the performance of iteration i; RM is the redundancy matrix nxn, with the "perspective" feature as rows; the dashed line illustrates the basis of the influence.

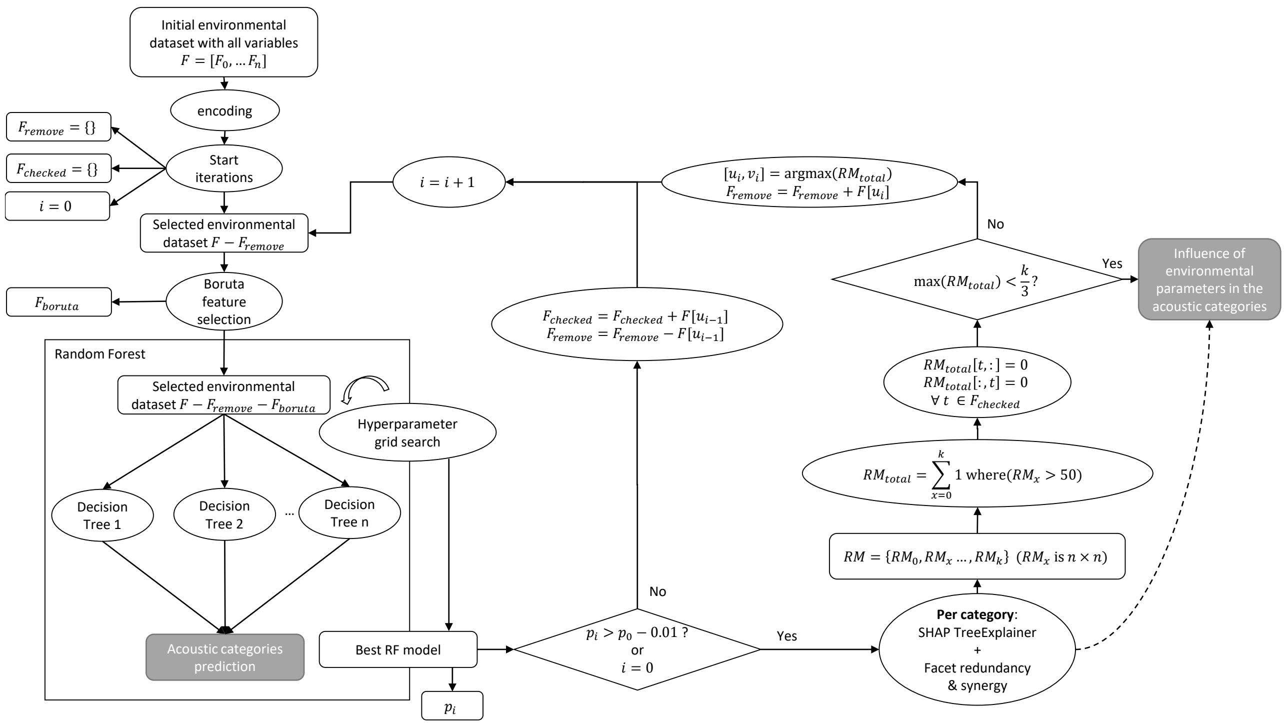
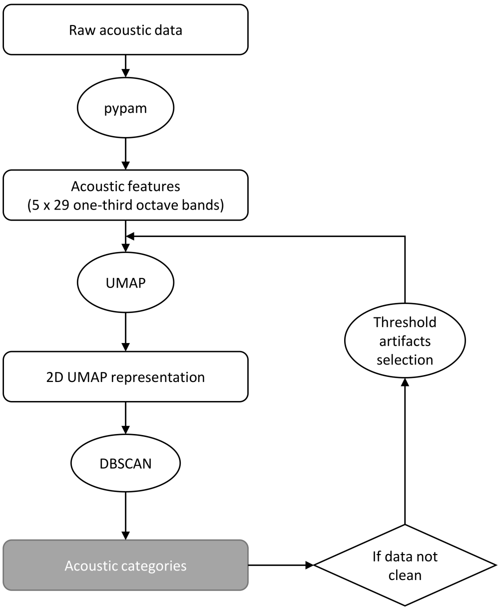


Figure 5.2 Flowchart to obtain the acoustic categories.



To decide if a variable should be removed from the analysis because it was too redundant, the obtained best RF was evaluated for each class, using the FACET [45] package. First, the redundancy matrix was computed per class [45]. Then, we counted in how many classes a variable pair was more than 50% redundant. The pair exceeding the threshold in the highest number of classes was selected for removal, and the ”perspective” variable was removed from the dataset in the following iteration. The whole training process was repeated without this variable, and the accuracy of the model was compared to the model from the first iteration. If the performance did not decrease by more than 1%, the variable was excluded from all further iterations. If the performance decreased more than 1%, then the variable was considered too important to be removed, was kept, and was not checked anymore in the following iterations. Iteratively, this process was repeated, until there were no pairs that had a redundancy higher than 50% in more than 1/3 of the classes. The last best RF model obtained was used for further analysis, with all the redundant features already eliminated.

Once all the redundant variables were removed, the last RF model was interpreted, using SHAP to further depict the relative importance of the remaining environmental features, to characterize the acoustic clusters. Using the computed SHAP values, a summary of the most important global features was plotted and analyzed. The SHAP python package incorporates the option to produce beeswarm plots, which allow for checking for feature importance per category, and understanding which values of each feature increase or decrease the probability of belonging to a certain class: this allowed for assessing why a certain spatiotemporal location was classified as a certain type of acoustic category.

SHAP offers an interactive plot, whereby we can see the probability contribution of each feature per class in time. In these plots, we can visualize the shapely values of each feature value as a force (arrow) that pushes to increase (positive value) or decrease (negative value) the probability of belonging to that specific class.

To test if the deployment settings had a significant influence on the recorded soundscapes, a categorical variable representing each deployment (deployment_id) was added, and the whole process was repeated. The selected variables were compared, and the influence of the deployment_id was compared to the other variables. The performance (total of data explained) of the RF trained with the deployment_id variable was compared to the model obtained, using only environmental variables. If the performance of the model did not improve by adding the deployment_id variable, it was assumed that it was redundant, with the data being already present in the dataset.

Furthermore, to understand whether there were specific clusters that were not understood by the model because they represented a certain sound not directly linked to the available environmental parameters, the data incorrectly classified by the obtained RF model was plotted in the 2D space. It was then visually checked if the samples wrongly classified grouped together. The percentage of each cluster correctly explained by the model was computed.

5.2.6 Predicting and mapping

With the obtained RF-model and the available environmental parameters, the presence of different soundscape categories can be predicted in the whole BPNS. This allows for visualizing the different soundscapes present in the BPNS. As a showcase on how this prediction and mapping can be used, we provide an example on the Stationary dataset.

Each obtained cluster was assigned a characteristic spectrum by averaging the HMBL in that cluster. Then categories were predicted for every hour during a month. The mean power density for each month for the bands centered at 63 Hz, 125 Hz, and 2000 Hz (center band 2002.16 Hz) was computed and mapped according to the weighted mean considering how often a certain category was predicted and averaging according to the mean value of each cluster at the specified frequency band. The 63 and 125 Hz bands were selected because they are the frequency bands selected by the EU to monitor the Good Environmental Status of marine waters [46]. 2000 Hz was selected because of its higher relevance for marine mammals, in line with other studies [47].

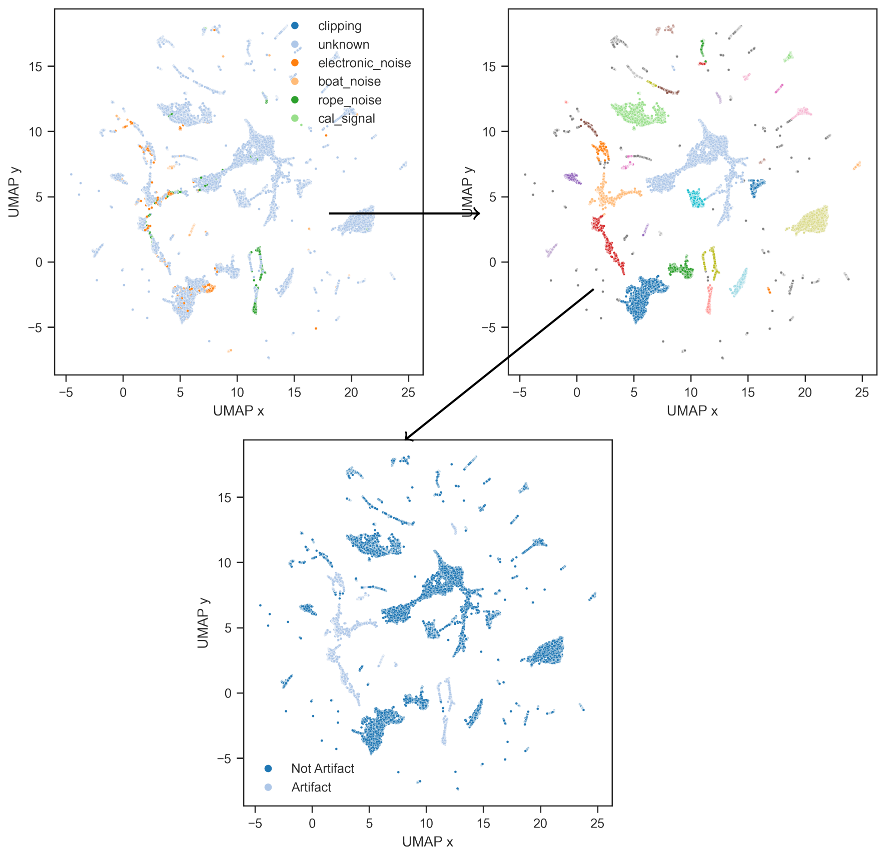
5.3 Results for drift data

5.3.1 Acoustic Categories

5.3.1.1 Data Cleaning

A total of 107 artifacts were manually labeled. The labeled data were plotted in a 2D UMAP representation, clearly grouped according to the label (Figure 5.3). This highlighted that points closer to each other in the UMAP space were similar acoustically. The DBSCAN algorithm generated 37 clusters: from these clusters, 9 were considered artifacts, and were removed from the dataset. The sum of all the samples of these clusters represented 15.30% of the dataset. The parameters used for UMAP and DBSCAN are in Table S5.1.

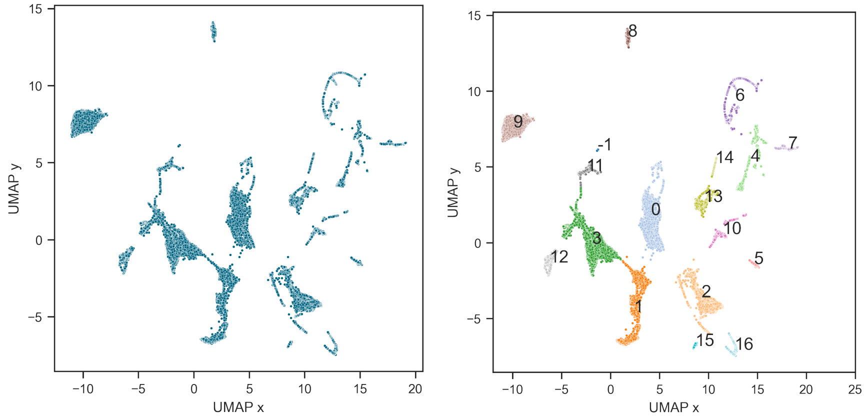
Figure 5.3 Drifts Dataset. **(a)** distribution of the labeled artifacts on the 2D UMAP dimension; **(b)** obtained 37 clusters after applying DBSCAN; **(c)** data considered to be artifact.



5.3.1.2 Acoustic Categories

UMAP was applied to the dataset without artifacts for a second time. The clean data plotted in this second UMAP space also presented clear clusters (Figure 5.4).

Figure 5.4 Drift dataset. (**a**) dimension reduction, using UMAP on the clean dataset; (**b**) obtained clusters using UMAP projection: -1 represents the noise class.



(a) **(b)**

A DBSCAN algorithm was run on the new UMAP space, and 17 clusters were obtained. Of the samples, 6.01% were classified as noise, and therefore they were removed from the dataset, for further analysis (Figure5.4). The mean and standard deviations of each cluster's one-third octave bands sound pressure levels evolution can be seen in Table S5.4. The parameters used for the UMAP and DBSCAN are listed in Table S5.1.

5.3.1.3 Characterization of Acoustic Classes

During the Boruta analysis, seabed_habitat_A5.23 or A5.24 was selected for removal. All the results of the iterative training of the RF and the redundant variables removal are summarized in Table 5.2. In the final model temperature, new_moon, substrate_Sandy mud to muddy sand and week_n_cos were removed.

The accuracy from the last best RF trained was 91.77 %: this percentage represented the proportion of the data that could be explained by the model. In the visual analysis of the data incorrectly classified, it was seen that the data incorrectly classified clearly clustered or were at the border of a cluster (Figure 5.5): this supported the hypothesis that these clusters represented similar sounds not correlated to the considered environmental parameters. After computing the percentage of each cluster correctly predicted by the model, cluster 14 was discarded as a soundscape,

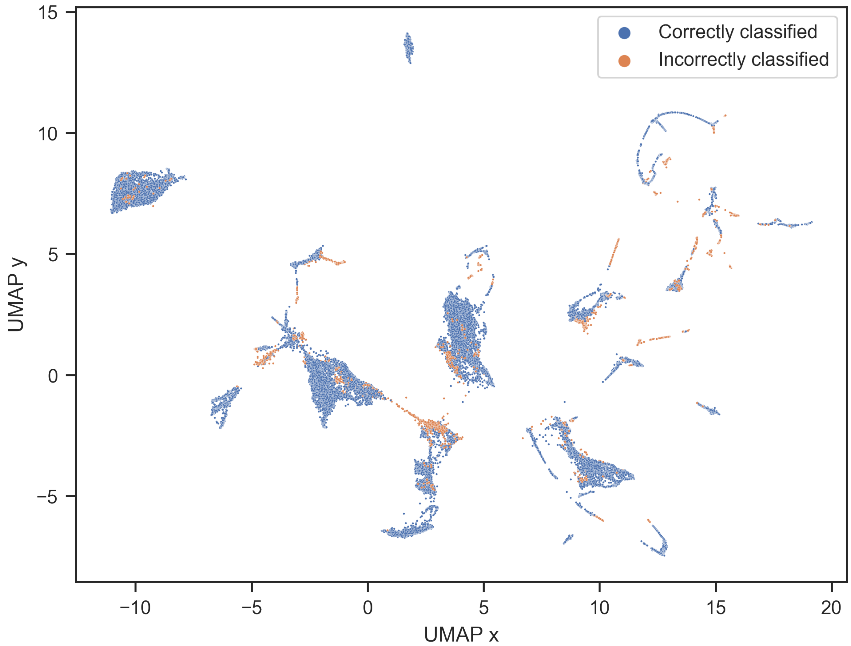
Table 5.2 Drift dataset. Summary of the selected model at each iteration, and their performance with the consequent redundant pairs and the features removed. Mean Accuracy represents the mean accuracy of the best model across the cross-validations. SH stands for seabed_habitat. In the column Maximum Redundant pair, the selected variable of the pair is in **bold**

i	**Mean Accuracy**	**Mean Std**	**Removed Features Boruta**	**Maximum Redundant Pair**
0	91.10	0.73	SH A5.23 or A5.24	**temperature** week_n_sin
1	90.66	0.43		**new_moon** week_n_cos
2	91.03	0.71		**substrate_Sandy mud to muddy sand** SH A5.35
3	91.08	0.59	SH A5.23 or A5.24	**week_n_cos** week_n_sin
4	91.77	0.60		**day_moment_Day** day_moment_Twilight
5	90.35	0.63	SH A5.23 or A5.24	day_moment_Day selected and discarded
6	91.77	0.60		**coast_dist** salinity
7	90.02	0.56		coast_dist selected and discarded
8	91.77	0.60		

because 0% of its samples were correctly explained by the environmental variables.

Using the last RF, the SHAP values were computed for the total dataset. The parameters most distinctive between classes globally were day_moment, instrument_depth, coast_dist, salinity and growing_moon (Figure 5.6). With the obtained plots of SHAP values per class, it could be understood which values were affecting each class. For example, from Figure 5.7, it can be interpreted that high values of instrument_depth increased the probability of a sample belonging to class 0. With coast_dist it was the other way around, and lower values decreased the probability of a sample belonging to class 0. For class 1, low values of shipping increased the probability of belonging to class 1, and high values of instrument_depth likewise. The SHAP plots per category can be see in Figure S5.1, and the results are manually summarized in Table 5.3. To improve visualization, variables corresponding to

Figure 5.5 Drift dataset. Samples correctly and incorrectly classified using the RF obtained in the last iteration.



an encoded value of a categorical variable were grouped, and plotted in gray in the SHAP plots: this was done because categorical variables do not have higher and lower values. In all the interpretations, medium values of week_n_cos were considered summer, because there was no sampling done in winter.

The total performance of the model, when adding the deployment_id variable, did not improve. It was concluded that deployment_id could safely be removed, and not considered a necessary factor to explain the differences between acoustic categories.

Figure 5.6 Drift dataset. Global summary of the influence of the environmental parameters on the acoustic categorization in the entire dataset. The SHAP values represent the probability that each feature can have on a sample to be classified as one particular class. The values shown here are the average of these probabilities in absolute values, both negative and positive. This plot shows the impact of each feature to each class (colors). The addition of all the impacts per class results in the total impact for the dataset for that feature. For example, moment of the day has the largest impact on the dataset, but for class 16 growing moon has a higher impact than moment of the day.

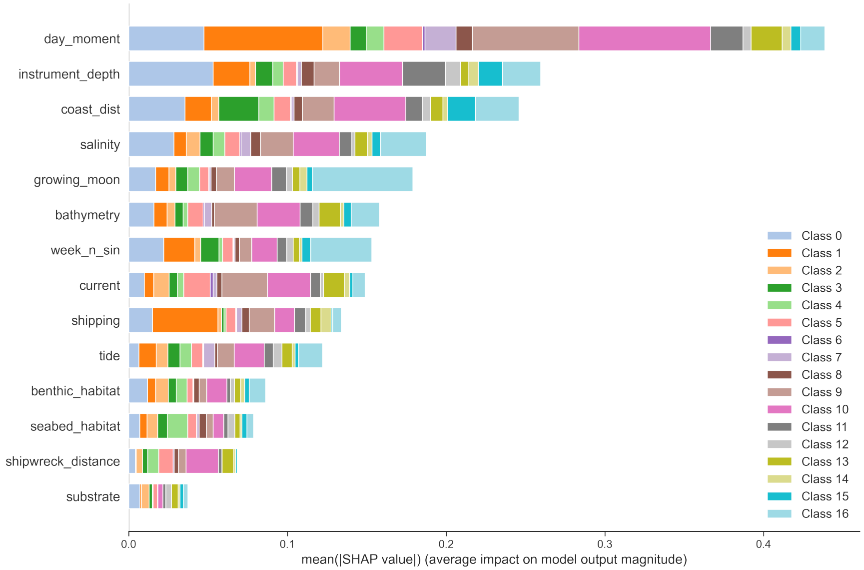
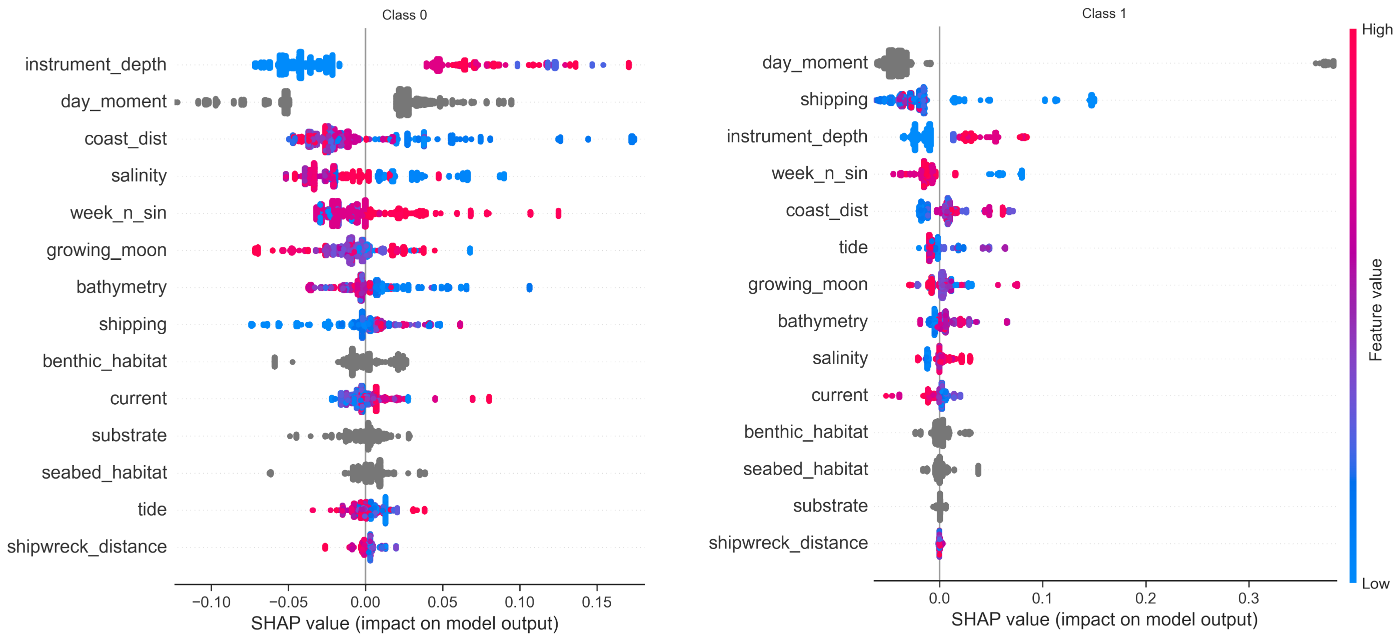


Figure 5.7 Drift dataset. Summary of the SHAP values for all the environmental parameters for (**a**) class 0 (**b**) class 1. SHAP values for other classes can be seen on Figure S5.1 (Supplementary Material). Variables in gray are categorical variables. The color of the dots represents the value of that variable, normalized (high, mid or low). The position of the dots with respect to the vertical line at 0 represents the impact that feature has on the probability of a sample to be classified as that class. Values to the right have a positive influence (i.e. increase the probability), while values to the left have a negative one (i.e. decrease the probability).



(**a**) (**b**)

Table 5.3: Drift dataset. Manual interpretation of the influence of the environmental variables in cluster differentiation. Because there was no sampling done in winter, medium values of week_n_cos were considered summer. % S stands for percentage of samples in a cluster, % E stands for percentage of the cluster correctly classified (explained). i stands for cluster number

i	% S	% E	Environmental Description
0	15.41	96.48	high instrument depth, day, close to the coast, low salinity, spring
1	12.4	90.38	night, low shipping, high instrument depth, autumn, mid-offshore, low–mid-tide
2	11.34	92.12	twilight, high current, deep waters, high salinity, mid-offshore, low instrument depth, summer
3	19.58	89.27	twilight, mid-offshore, low instrument depth, mid–high salinity, low–mid-current, shallow waters, summer
4	5.1	76.38	high instrument depth, day, mid-offshore, spring, deep waters, high salinity
5	1.53	99.63	high instrument depth, high tide, mid-offshore, day, seabed habitat A5.15, spring, deep waters
6	2.86	82.50	day, medium bathymetry, high current, medium salinity, mid-offshore, far from shipwreck, mid–high shipping
7	1.21	98.38	low shipping, high instrument depth, day, decreasing moon, high current, mid-offshore
8	2.58	99.78	close to the coast, very low instrument depth, day, summer, low salinity, shallow waters
9	12.17	99.15	growing moon, spring, low salinity, close to the coast, low instrument depth, shallow waters, day
10	2.83	81.07	day, low–mid-current, mid-salinity, not benthic habitat 3 (*Macoma balthica* community), high tide, seabed habitat A5.14, medium depth
11	3.26	79.93	close to the coast, summer, low instrument depth, day, low salinity, high tide, new or full moon, seabed habitat A5.35
12	2.56	99.67	seabed habitat A5.35, day, close to the coast, full or new mooon, low tide, far from shipwreck

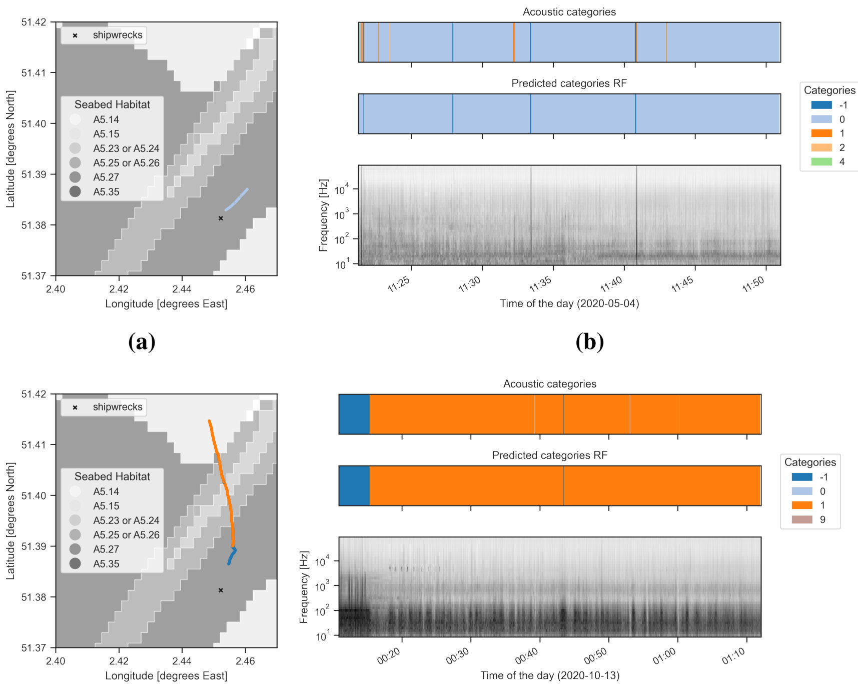
13	4.39	79.80	day, high current, mid-offshore, medium depth, medium salinity, far from shipwreck, low instrument depth, benthic habitat 3 (*Macoma balthica* community)
14	0.33	0.00	(not a soundscape) high current, day, mid-offshore, medium depth, medium salinity, far from shipwreck, low instrument depth, high tide, benthic habitat 3 (*Macoma balthica* community)
15	0.8	100.00	twilight, low–mid-tide, high salinity, deep waters, low shipping, low–mid-current, low instrument depth, far from the coast
16	1.65	92.66	day, low instrument depth, high salinity, mid-offshore, seabed habitat A5.27, low shipping, benthic habitat 3 (*Macoma balthica* community)

We manually analyzed the output categories in space and time. We observed a variation in the soundscapes recorded across the 40 sites, and also a shift within some individual deployments. The changes were driven by both spatial and temporal features: as an example, we examined one location recorded during different days and moments (Figure 5.8). The deployment plotted in (a)–(b) fell mostly under class 0, which was characterized as spring. The deployment in (c)–(d) fell under class 1, which was characterized as autumn.

Some deployments changed soundscape categories during the deployment, as in Figure 5.9, where the deployment started with class 3, then switched to class 13, and then to class 6. Referring to Table 5.3 and Figure S5.1, the biggest differences between class 3 and classes 13 and 6 were the current and distance to a shipwreck. This description correlates with the fact that the current was lower when class 3 appeared, and higher during classes 13 and 6. It also correlates with the fact that there was a shipwreck close to class 3, while there were no shipwrecks in the vicinity of class 6 and 13. Furthermore, the change of cluster from class 13 to 6 may have been due to the change in the sea bottom substrate, where it changed from coarse sediment to sand. These descriptions match the spatial changes (Figure 5.9a), where class 13 was found usually on coarse sediment bottoms, and class 6 was not defined by the substrate. In the spectrogram, we can also see a change in the sound. Some parts were classified as noise, because of the loud shipping noise present. In this example, class 6 and 14 are wrongly predicted by the model for a very short time. Class 14, which was not considered a soundscape category, so it was just a foreground event not considered by the model. Other examples of deployments can be found in Figure S5.2.

In Figure 5.10, we can see the interactive "force" plots for two of the classes from the same deployment as Figure 5.9.

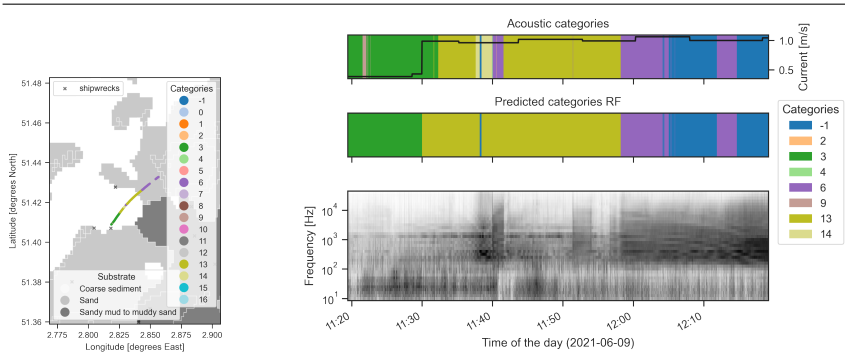
Figure 5.8 Deployment of the drift dataset in spring plotted **(a)** spatially and **(b)** temporally. Deployment in autumn, plotted spatially **(c)** and temporally **(d)**.



(a) **(b)**

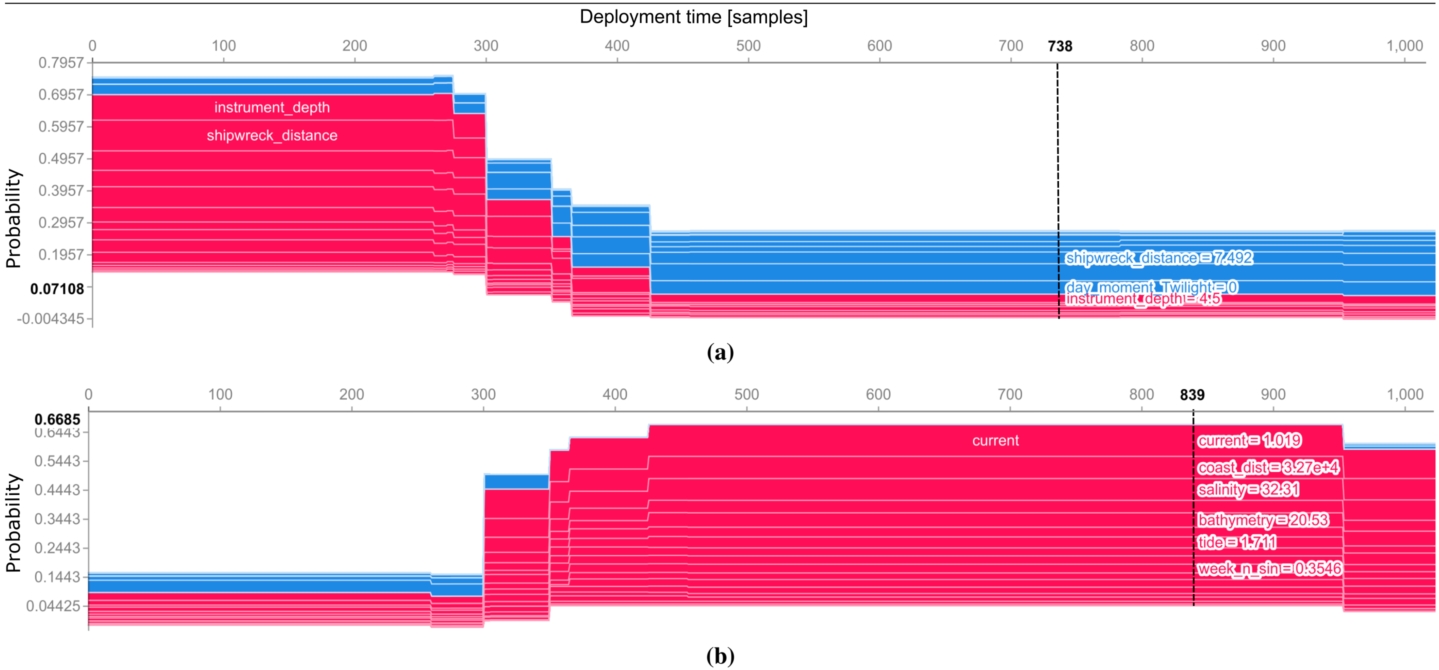
(c) **(d)**

Figure 5.9 Zoom of one deployment of the drift dataset recorded the 9 June of 2021: (**a**) plotted on top of a map of the substrate type; (**b**) plotted in time.



(a) **(b)**

Figure 5.10 Drift dataset. SHAP force plots showing samples' probability to be classified as class **(a)** 3 and **(b)** 13 during deployment from Figure 5.9. Blue represents negative influences of a variable, and red positive ones. With these plots the user can understand how the probability of belonging to one class or another changes in time. For example, at the beginning of the deployment the probability of the samples being class 3 is high because of the values of instrument_depth and shipwreck_distance, and it changes at around sample 350 because of the values of shipwreck_distance and day_moment. Dashed lines indicate the sample at which the values are shown. Deployment length has been cut to increase readability of the image.



5.4 Results for stationary data

5.4.1 Acoustic Categories

No data cleaning was performed in the long-term data because of lack of annotations. Four soundscape categories were clearly identified (see Figure 5.11). The parameters used for the UMAP and DSBSCAN are listed in the Table S5.1.

The obtained RF-model classified the soundscape categories with 85.75% accuracy. The description of each cluster is shown in Table 5.4. The principal variables selected to explain the clusters were bathymetry and seabed habitat, which in this case were linked to location (see Figure 5.11). This suggests that the acoustic characteristics of the habitats in the BPNS differ more between habitats than within. Category 1 was not correctly predicted by the model, which suggests that it represents a certain acoustic situation not linked to the selected environmental variables. This hypothesis was supported by the fact that category 1 comprised samples from all the different locations and deployments.

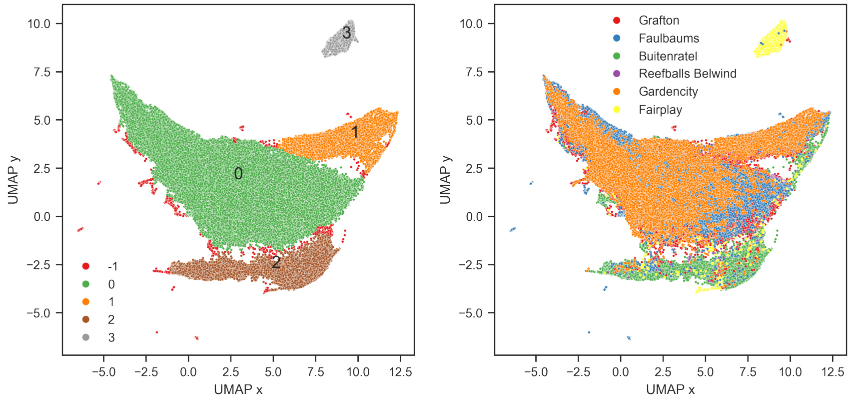
Table 5.4 Stationary dataset. Manual description of the SHAP values of each category.

Category	Description
0	low bathymetry, seabed habitat 5.27 or A5.14
1	(not a soundscape class) low bathymetry, low wave height, medium temperature
2	high bathymetry, seabed habitat A5.23 or A5.24, low temperature, high wave height
3	high bathymetry, seabed habitat A5.25 or A5.26, mid-high temperature, low wave height

Two examples of a map of the predicted categories in March and November can be seen in Figure 5.13. Category 0 seems to be the most present, and category 1 seems to be linked to shallower areas. This is also reflected in the frequency distribution. The power density at 63 Hz is low in areas classified as category 2 or 3 (Figure 5.14). In areas classified as category 0, the SPL at 63 and 125 Hz is higher than average while it is lower than average at 2000 Hz. By analyzing the obtained SHAP plots per class, it can be seen that category 0 is linked to low bathymetry (see Figure 5.12 and Table 5.4)). This is in line with shipping sound production, which is usually characterized by having most of the energy at lower frequencies.

The authors emphasize that these results report only a preliminary partial analysis. Several seasonal effects have yet to be explored in detail. Moreover, the limited temporal resolution of 1 minute does not allow for the identification of

Figure 5.11 Stationary dataset. **(a)** Obtained clusters using DBSCAN in the UMAP space. -1 represents samples classified as noise and not belonging to any cluster. **(b)** Distribution of the stations in the UMAP space.



(a) **(b)**

specific biological sounds that may contribute to the soundscape only at certain locations and certain times of the year. Further analysis will be reported elsewhere.

Figure 5.12 Stationary dataset. SHAP values of the obtained category 0 resulting from the trained RF-model.

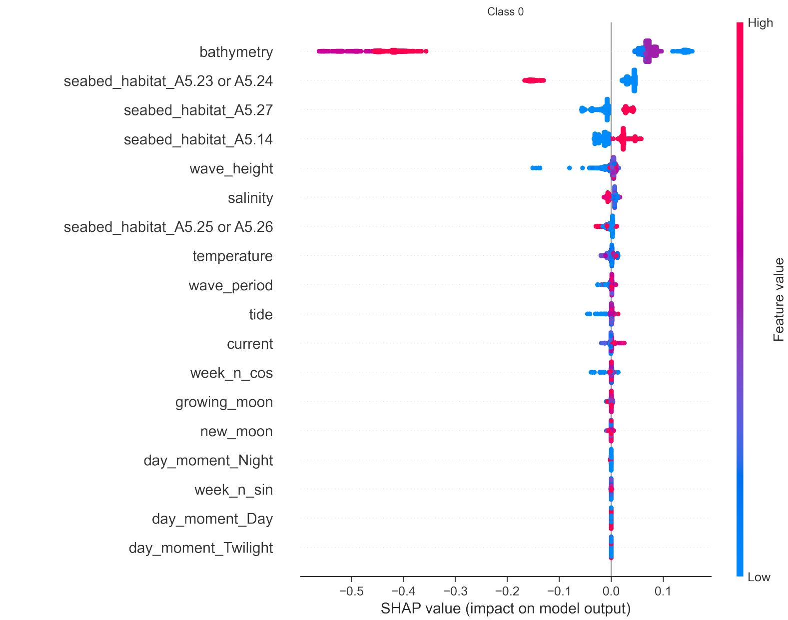
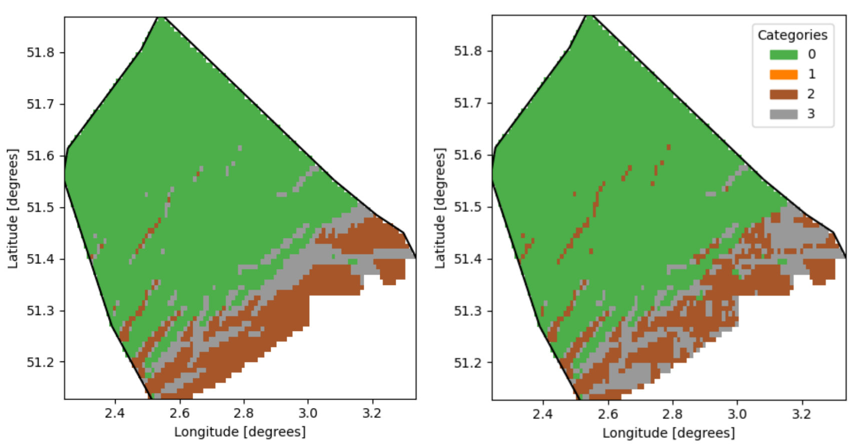
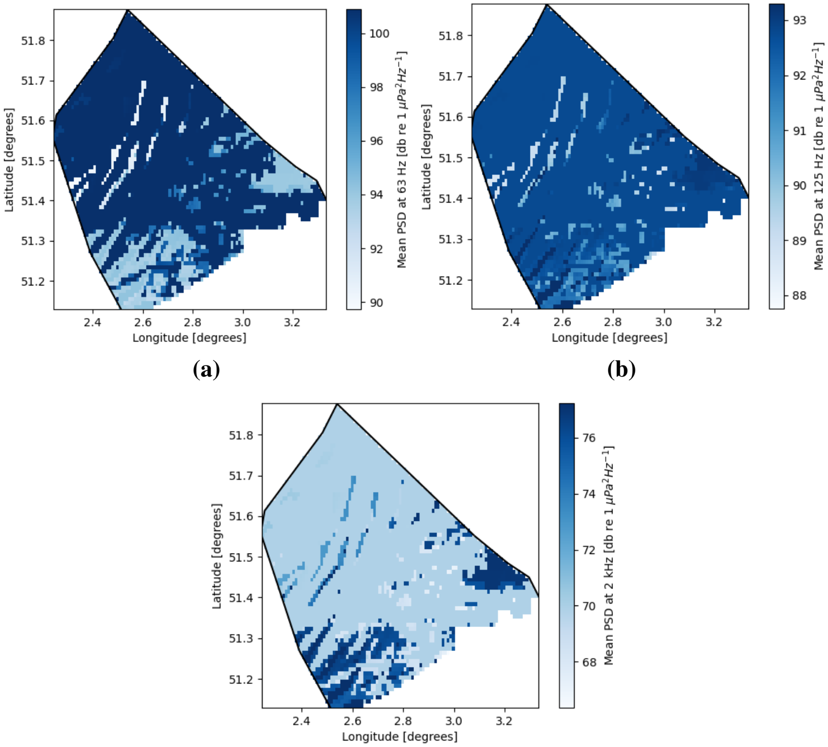


Figure 5.13 Stationary dataset. Predicted soundscape categories in the BPNS (black line is the delimitation of the Exclusive Economic Zone) according to the environmental parameters using the RF-model, from two randomly selected timestamps. **(a)** prediction of 8th of March of 2022 at 12:00 am. **(b)**, prediction of 25th of November at 00:00 am.



(a) **(b)**

Figure 5.14 Stationary dataset. Predicted mean power density for the three selected frequency bands during January 2022. Categories predicted hourly for the entire month. Mean power density computed considering the assigned power density at each frequency for each category and its hourly occurrence during the month. **(a)** 63 Hz. **(b)** 125 Hz. **(c)** 2000 Hz.



(a) **(b)**

(c)

5.5 Discussion

The proposed method was tested on a long-term dataset with high temporal coverage but low spatial coverage, and on a drifting dataset with high spatial coverage but low temporal coverage. The long-term dataset led to a smaller number of clusters, mostly based on bathymetry and the type of benthic habitat and substrate present (therefore, location). The drifting dataset led to a greater number of clusters, differentiated by more environmental variables such as time of the day, instrument depth, distance to the coast, salinity, and moon phase.

With this method we showed that by using an unsupervised approach we were able to categorize different marine shallow water soundscapes. In addition, we demonstrated that by using an automatic (supervised) approach, based on Explainable machine learning tools, we were able to characterize these categories ecologically. Our method was able to group soundscapes in different categories, which could be used to understand the spatiotemporal acoustic variations of a the marine environment. The obtained categories were afterwards proved to be connected to environmental parameters through an RF classifier. To understand the predictions made by the trained RF, SHAP was used: this allowed for the assessment of the main environmental parameters shaping the acoustic categories in general, but also per category, providing a practical description for each soundscape category.

One added value of the proposed methodology is that the obtained model can be used to forecast the effects of environmental change on the soundscape. This can eventually be used to create habitat suitability maps. Forecasting can be achieved by predicting the different categories for a given set of environmental parameters. The predictions can then help understand the consequences of certain changes or decisions in the environment regarding sound.

Our results show that the acoustic data analyzed by UMAP and DBSCAN clustered mostly in clear, independent groups. This indicates that there were major and quantifiable differences in BPNS' underwater soundscapes, and that one-third octave bands and hybrid millidecade bands (sound pressure) levels encoded enough information to capture these dissimilarities. This represents an advantage, due to the current availability of built-in implementations in some recorders, and the different available tools to compute them. We chose the frequency band of study to capture the sound sources of interest, and it captured enough information to obtain distinct clusters. Increasing the frequency range might have increased the number of clusters. Furthermore, similar artifact sounds clustered together. This is in line with findings from Sethi et al. (2020) [4], where artifacts could be detected by using an unsupervised clustering technique. Accordingly, the semi-supervised process used in this chapter could be applied to detecting artifacts in acoustic datasets. This would be especially useful for long-term deployments, where exhaustive manual analysis is too time consuming, and it would be a rapid solution for detecting instrument malfunction events.

The RF classifier was able to correctly classify more than 85% of the BPNS' acoustic data for both datasets: this high accuracy suggests that the environmental parameters included in the analysis were good indicators of the observed acoustic patterns. The SHAP values showed that in our case study, time of the day, instrument depth, distance to the coast, salinity and moon phase were the most important environmental parameters shaping and differentiating the soundscape categories. These results should be carefully interpreted. The importance of the environmental parameters was not necessarily correlated to their influence on the total sound: rather, it described their relevance to discriminating between categories. For example, if all the categories from the study area had had a considerable and equally distributed sound contribution from shipping, the feature would not have had a great effect in differentiating them: as a consequence, it would have had a low importance score in the model. However, if we had expanded the dataset with acoustic data from other areas in the world with less shipping influence, shipping would have become a very dominant feature in explaining the categories' differences. In addition, the redundant variables that were removed should not be ignored, but should be considered together with the redundant pair. For example, in the drift analysis, temperature was removed, because of its redundancy with the season (week_n_sin). Consequently, in all the clusters where season has an influence, we would not be able to distinguish if the real effect was the temperature or the season: we would know, however, that these two correlated parameters had an influence on the soundscape. If distinguishing between two redundant features would be relevant, more data should be collected to that effect, in a way that the two variables are not correlated.

Instrument depth showed a strong influence on discriminating between soundscape categories when drifting. This result could have been expected: as the recording depth was not kept at the same level and in such shallow environments, acoustic changes in the soundscape occurred at very small spatial scales, and vertically in the water column [14]. It is therefore important to always consider and report the hydrophone depth when comparing different soundscapes, to avoid any misleading conclusions. Furthermore, to exclude the position effect, and in order to better assess the biotic-driven soundscapes in shallow environments, recordings should be taken at a fixed depth.

Bathymetry and seabed habitat was the main factor differentiating the clusters when looking at long-term patterns. This suggests that habitats have different sound signatures which differ more in space than in time when not looking at short sounds.

The obtained categories showed the expected acoustic variation: this reflected the dynamic environment in the BPNS, and the need to study these soundscapes in more detail, in order to better understand this specific marine acoustic environment.

Spatial or temporal change in environmental variables could be noticed in the acoustic scenes, and the obtained categories reflected these changes.

To be able to generalize our conclusions and test the robustness of the method, it should be applied to new long- and short-term data from different underwater acoustic contexts. If these datasets came from ecosystems that were well-studied acoustically, the results would be contrasted. We would thereby be able to assess whether the obtained categories were representative and informative enough, and if they matched the currently existing knowledge or if they complemented it. Furthermore, the incorrectly classified data could also be analyzed manually, to detect specific events, and to have more insight into the missing explanatory parameters.

The method proposed here could be particularly useful in environments where the visual correlation between ecological factors and the underwater soundscape cannot be established: this includes low visibility and other challenging conditions, such as those occurring in remote areas with high latitudes, where the winter season prevents traditional ways of surveying, or in highly exploited areas. In these cases, a rapid and automated tool, capable of characterizing the soundscape, and of monitoring its potential changes in relation to relevant environmental drivers, would be very valuable.

Ecologically characterizing the soundscape categories is only possible if data from all the environmental parameters are available. If not, the method could still be applied to categorize the different recorded soundscapes into acoustically relevant categories that could help guide conservation decisions on, e.g., areas with diverse soundscape patterns. It would also be possible to use the categories to optimize the sampling effort, and to only sample for potential drivers where the soundscape categories are, e.g., most distinct. If no environmental data from a specific site were available, it would be possible to train the model on a similar dataset, but with available environmental data. The acoustic data could then be explored, according to the obtained classification, to assess whether there were similarities between the soundscape categories obtained in both datasets, thus establishing a potential relationship with analogous drivers. In addition, the acoustic categories obtained in such an unsupervised way could be manually analyzed and labeled, and subsequently used as a baseline for future monitoring, to assess the acoustic change in time or the spatial acoustic (dis)similarities in a certain environment.

5.6 Conclusions

This work constitutes a significant contribution to the development of a methodology for monitoring and characterizing underwater soundscapes in a fast and automatic way, thereby complementing previous works on underwater soundscape analysis [48, 49, 50]. Classifying soundscapes into categories assigns a label to each acoustic environment. This is an easy way to refer to, and identify them, which can be very useful for policy or conservation programs. One application of using categories is to analyze long-term data. The categories then can point out trends, status and seasonal patterns. Working with fully unsupervised acoustic categories has potential to assess the dynamic character of the soundscape. The fact that no labeling effort is needed is a step forward to solve the problem of coping with the analysis of the increasing amount of acoustic data that the technological advances allow to collect and store. In the global context of rapid environmental change and increasing anthropogenic pressure it is of critical importance to assess the acoustic research gap and study bias towards more accessible ecosystems [51]. The method we propose here constitutes an innovative and practical tool to categorize and characterize marine soundscapes, particularly in poorly known acoustic contexts. It has potential to be used for conservation purposes, disturbance detection and ecosystem integrity assessment.

Acknowledgments

The project was realized as part of the Flemish contribution to LifeWatch. We would like to thank the crew of the RV Simon Stevin. We would also like to thank Jan Vermaut for his essential work in driving the Zeekat drifting all over the BPNS, and the Sailawaywithme sailing team who made it possible to sail to further places and acquire night recordings. We would also like to acknowledge the contribution of Roeland Develter and Klaas Deneudt, for helping with the data collection. The author would also like to thank Peter Rubbens and Olaf Boebel for their advice during the study, and Marine Severin for her help with the manuscript.

Data availability statement

The dataset of the processed acoustic data for the Drifts dataset (one-third octave band sound pressure levels) can be found in the Marine Data Archive https://doi.org/10.14284/586 (accessed on 19 January 2023).

References

[1] International Organization for Standardization (ISO). *Acoustics - Soundscape - Part 1: Definition and Conceptual Framework*. Technical report, International Organization for Standardization, Geneva, 2014.

[2] J. Abeßer. *A Review of Deep Learning Based Methods for Acoustic Scene Classification*. Applied Sciences, 10(6):2020, January 2020.

[3] J. L. Miksis-Olds, B. Martin, and P. L. Tyack. *Exploring the Ocean Through Soundscapes*. Acoustics Today, 14(1):9, 2018.

[4] S. S. Sethi, N. S. Jones, B. D. Fulcher, L. Picinali, D. J. Clink, H. Klinck, C. D. L. Orme, P. H. Wrege, and R. M. Ewers. *Characterizing Soundscapes across Diverse Ecosystems Using a Universal Acoustic Feature Set*. Proceedings of the National Academy of Sciences, 117(29):17049–17055, July 2020.

[5] R. Benocci, G. Brambilla, A. Bisceglie, and G. Zambon. *Eco-Acoustic Indices to Evaluate Soundscape Degradation Due to Human Intrusion*. Sustainability, 12(24):10455, January 2020.

[6] I. Roca and I. Van Opzeeland. *Using Acoustic Metrics to Characterize Underwater Acoustic Biodiversity in the Southern Ocean*. Remote Sensing in Ecology and Conservation, 6, September 2019.

[7] F. Michaud, J. Sueur, M. Le Cesne, and S. Haupert. *Unsupervised Classification to Improve the Quality of a Bird Song Recording Dataset*. Ecological Informatics, 74:101952, May 2023.

[8] J. S. Ulloa, T. Aubin, D. Llusia, C. Bouveyron, and J. Sueur. *Estimating Animal Acoustic Diversity in Tropical Environments Using Unsupervised Multiresolution Analysis*. Ecological Indicators, 90:346–355, July 2018.

[9] L. H. Hilasaca, M. C. Ribeiro, and R. Minghim. *Visual Active Learning for Labeling: A Case for Soundscape Ecology Data*. Information, 12(7):265, July 2021.

[10] Y. F. Phillips, M. Towsey, and P. Roe. *Revealing the Ecological Content of Long-Duration Audio-Recordings of the Environment through Clustering and Visualisation*. PLOS ONE, 13(3):e0193345, March 2018.

[11] P. Linardatos, V. Papastefanopoulos, and S. Kotsiantis. *Explainable AI: A Review of Machine Learning Interpretability Methods*. Entropy, 23(1), January 2021.

[12] Y. Cha, J. Shin, B. Go, D.-S. Lee, Y. Kim, T. Kim, and Y.-S. Park. *An Interpretable Machine Learning Method for Supporting Ecosystem Management: Application to Species Distribution Models of Freshwater Macroinvertebrates*. Journal of Environmental Management, 291:112719, August 2021.

[13] C. Parcerisas. *Lifewatch/Pypam: Pypam, a Package to Process Bioacoustic Data*. Zenodo, February 2022.

[14] S. Robinson, P. Lepper, and R. Hazelwood. *Good Practice Guide for Underwater Noise Measurement*. Technical report, National Measurement Office, Marine Scotland, The Crown Estate, 2014.

[15] M. F. McKenna, S. Baumann-Pickering, A. C. M. Kok, W. K. Oestreich, J. D. Adams, J. Barkowski, K. M. Fristrup, J. A. Goldbogen, J. Joseph, E. B. Kim, A. Kügler, M. O. Lammers, T. Margolina, L. E. Peavey Reeves, T. J. Rowell, J. A. Stanley, A. K. Stimpert, E. J. Zang, B. L. Southall, C. C. Wall, S. V. Parijs, and L. T. Hatch. *Advancing the Interpretation of Shallow Water Marine Soundscapes*. Frontiers in Marine Science, 8, 2021.

[16] J. K. Garrett, Ph. Blondel, B. J. Godley, S. K. Pikesley, M. J. Witt, and L. Johanning. *Long-Term Underwater Sound Measurements in the Shipping Noise Indicator Bands 63Hz and 125Hz from the Port of Falmouth Bay, UK*. Marine Pollution Bulletin, 110(1):438–448, September 2016.

[17] M. van der Schaar, M. A. Ainslie, S. P. Robinson, M. K. Prior, and M. André. *Changes in 63Hz Third-Octave Band Sound Levels over 42months Recorded at Four Deep-Ocean Observatories*. Journal of Marine Systems, 130:4–11, February 2014.

[18] M. S. Prawirasasra, M. Mustonen, and A. Klauson. *The Underwater Soundscape at Gulf of Riga Marine-Protected Areas*. Journal of Marine Science and Engineering, 9(8):915, August 2021.

[19] J. L. Miksis-Olds, P. J. Dugan, S. B. Martin, H. Klinck, D. K. Mellinger, D. A. Mann, D. W. Ponirakis, and O. Boebel. *Ocean Sound Analysis Software for Making Ambient Noise Trends Accessible (MANTA)*. Frontiers in Marine Science, 8:703650, August 2021.

[20] S. B. Martin, B. J. Gaudet, H. Klinck, P. J. Dugan, J. L. Miksis-Olds, D. K. Mellinger, D. A. Mann, O. Boebel, C. C. Wilson, D. W. Ponirakis, and H. Moors-Murphy. *Hybrid Millidecade Spectra: A Practical Format for Exchange of Long-Term Ambient Sound Data*. JASA Express Letters, 1(1):011203, January 2021.

[21] EMODnet Seabed Habitats. *Seabed Habitats*, 2018.

[22] EMODnet Human Activities. *Vessel Density*, 2022.

[23] C. Parcerisas. *Lifewatch/Bpnsdata: First Release of Bpnsdata*. Zenodo, November 2021.

[24] FACET Team at BCG Gamma. *Sklearndf*, 2021.

[25] A. N. Radford, E. Kerridge, and S. D. Simpson. *Acoustic Communication in a Noisy World: Can Fish Compete with Anthropogenic Noise?* Behavioral Ecology, 25(5):1022–1030, September 2014.

[26] EMODnet Bathymetry Consortium. *EMODnet Digital Bathymetry (DTM 2018)*, 2018.

[27] B. Rhodes. *Skyfield: High Precision Research-Grade Positions for Planets and Earth Satellites Generator*, July 2019.

[28] S. Derous, M. Vincx, S. Degraer, K. Deneudt, P. Deckers, D. Cuvelier, J. Mees, W. Courtens, E. W. M. Stienen, H. Hillewaert, K. Hostens, I. Moulaert, V. V. Lancker, and E. Verfaillie. *A Biological Valuation Map for the Belgian Part of the North Sea*. Global Change, page 162, 2007.

[29] MFC ODNature RBINS. *Physical State of the Sea - Belgian Coastal Zone - COHERENS UKMO*, 2023.

[30] Vlaanderen. *Wrakkendatabank 2.0*. https://wrakkendatabank.afdelingkust.be/.

[31] Flanders Marine Institute. *Marine Regions*. https://fairsharing.org/10.25504/FAIRsharing.5164e7, 2005.

[32] MFC ODNature RBINS. *Wave Field - North Sea - WAM ECMWF*, 2023.

[33] L. McInnes, J. Healy, and J. Melville. *UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction*, September 2020.

[34] L. van der Maaten and G. Hinton. *Visualizing Non-Metric Similarities in Multiple Maps*. Machine Learning, 87(1):33–55, April 2012.

[35] E. Becht, L. McInnes, J. Healy, C.-A. Dutertre, I. W. H. Kwok, L. G. Ng, F. Ginhoux, and E. W. Newell. *Dimensionality Reduction for Visualizing Single-Cell Data Using UMAP*. Nature Biotechnology, 37(1):38–44, January 2019.

[36] P. Best, S. Paris, H. Glotin, and R. Marxer. *Deep Audio Embeddings for Vocalisation Clustering*. PLOS ONE, 18(7):e0283396, July 2023.

[37] T. Sainburg, M. Thielk, and T. Q. Gentner. *Finding, Visualizing, and Quantifying Latent Structure across Diverse Animal Vocal Repertoires*. PLOS Computational Biology, 16(10):e1008228, October 2020.

[38] M. Thomas, F. H. Jensen, B. Averly, V. Demartsev, M. B. Manser, T. Sainburg, M. A. Roch, and A. Strandburg-Peshkin. *A Practical Guide for Generating Unsupervised, Spectrogram-Based Latent Space Representations of Animal Vocalizations*. Journal of Animal Ecology, 91(8):1567–1581, 2022.

[39] Team A. Audacity. *Team A. Audacity(R): Free Audio Editor and Recorder [Computer Application];1999-2019.*, 2023.

[40] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and É. Duchesnay. *Scikit-Learn: Machine Learning in Python*. Journal of Machine Learning Research, 12(85):2825–2830, 2011.

[41] S. M. Lundberg, G. Erion, H. Chen, A. DeGrave, J. M. Prutkin, B. Nair, R. Katz, J. Himmelfarb, N. Bansal, and S.-I. Lee. *From Local Explanations to Global Understanding with Explainable AI for Trees*. Nature Machine Intelligence, 2(1):56–67, January 2020.

[42] C. Molnar. *Interpretable Machine Learning. A Guide for Making Black Box Models Explainable*. 2 edition, 2022.

[43] J. Ittner, L. Bolikowski, K. Hemker, and R. Kennedy. *Feature Synergy, Redundancy, and Independence in Global Model Explanations Using SHAP Vector Decomposition*, July 2021.

[44] M. B. Kursa and W. R. Rudnicki. *Feature Selection with the Boruta Package*. Journal of Statistical Software, 36:1–13, September 2010.

[45] F. T. a. B. GAMMA. *FACET*, 2021.

[46] *Commission Decision (EU) 2017/848 of 17 May 2017 Laying down Criteria and Methodological Standards on Good Environmental Status of Marine Waters and Specifications and Standardised Methods for Monitoring and Assessment, and Repealing Decision 2010/477/EU (Text with EEA Relevance.)*, May 2017.

[47] M. Mustonen, A. Klauson, M. Andersson, D. Clorennec, T. Folegot, R. Koza, J. Pajala, L. Persson, J. Tegowski, J. Tougaard, M. Wahlberg, and P. Sigray. *Spatial and Temporal Variability of Ambient Underwater Sound in the Baltic Sea*. Scientific Reports, 9(1):13237, September 2019.

[48] N. D. Merchant, K. M. Fristrup, M. P. Johnson, P. L. Tyack, M. J. Witt, P. Blondel, and S. E. Parks. *Measuring Acoustic Habitats*. Methods in Ecology and Evolution, 6(3):257–265, 2015.

[49] D. C. Wilford, J. L. Miksis-Olds, S. B. Martin, D. R. Howard, K. Lowell, A. P. Lyons, and M. J. Smith. *Quantitative Soundscape Analysis to Understand Multidimensional Features*. Frontiers in Marine Science, 8:672336, August 2021.

[50] T.-H. Lin, S.-H. Fang, and Y. Tsao. *Improving Biodiversity Assessment via Unsupervised Separation of Biological Sounds from Long-Duration Recordings*. Scientific Reports, 7(1):4547, December 2017.

[51] M.-N. Havlik, M. Predragovic, and C. M. Duarte. *State of Play in Marine Soundscape Assessments*. Frontiers in Marine Science, 9, 2022.

Part III

Acoustic Event Detection

6

Detecting baleen whale calls using Convolutional Neural Networks (CNN)

Overview

This chapter shows the applicability of using transfer learning from computer vision to detect sound events in a marine environment when the sound sources are known. To this end, we apply an object detector to an open dataset containing Antarctic baleen whale calls. This chapter is used as a conceptual use case for Chapter 7. This chapter has not been published and has been written specifically for this thesis.

6.1 Introduction

Known stereotypical sounds present in a frequency band with little interference with other sounds, can be detected using standard signal processing techniques, or simple mathematical models [1, 2]. In these models, one or several signal processing techniques are applied to extract features, on which a hand-crafted decision tree can be applied. This is the case for cetaceans found in the BPNS, whose vocalizations are known. Moreover, little masking occurs at the frequency of echolocation clicks of harbor porpoises (*Phocoena phocoena*) and dolphins. Therefore, these can be monitored using PAM, which is already happening within the framework of the LifeWatch Cetacean Network [3] with the use of C-PODs and F-PODS. These devices can detect the stereotypical clicks produced by harbor porpoises and dolphins and store the resulting data and metadata of these detections. Dolphins are classified collectively rather than as distinct individual species. The information on the spatio-temporal distribution of these animals can then be utilized in impact assessments, population health status monitoring, and the study of seasonal patterns, among other applications. However, the source code used in the C-PODs to detect the echolocation clicks is not available, and therefore it is a ”black box” which is hard to evaluate. The output produced by these devices are logs of the detected events, and not the waveform of the recorded sound. This does not allow for extracting additional information about the detections and/or re-evaluate them when better software is available. To overcome this, devices that store the entire waveform can be used, recording at a high enough sampling rate to record the high frequency clicks. This solution is highly energy and memory intensive. Other devices such as the SoundTrap (OceanInstruments, New Zealand) have the option to store the waveform only when there is a possible click, which is detected using a signal-to-noise threshold at the frequency band of interest. This module in SoundTrap is called the Click Detector. Open source software such as PAMGuard [4] and D-PorCCA [5] can then be used both on raw acoustic data or on the output of a SoundTrap Click Detector. These algorithms have to be applied after the raw acoustic data are collected because they are not embedded in a sensor. This has the advantage of also having access to the raw data and therefore to more information for performance evaluation, but the disadvantage of a higher battery consumption and storage need. A recent study indicates that despite the different approach of the two systems, they yield similar results when used for presence or absence monitoring [6]. For other regions such as the polar ocean, mathematical models can also be successfully developed for known cetacean vocalizations, such as the detector of fin whale *Balaenoptera physalus* vocalizations listed in Schall & Parcerisas (2022) [1]. However, these mathematical models are e to background noise, and therefore are not always robust across locations. These methods often do not achieve the accuracy of human classification abilities, which is desired for population assessment, as explained in Section 1.2. For this reason growing interest is pointed towards deep learning and other supervised machine learning techniques.

Some publications in the last years have achieved well performing models for whale call detections, but only a small number of these models has been tested on long-term data [7]. Most likely this is because most published approaches were only evaluated on selected subsets of data which do not represent a real-world detection scenario (e.g. [8, 9, 10]). A real-world detection scenario in the marine realm is almost always characterized by a large imbalance between temporal periods when animal vocalizations are present and time periods when “only” environmental (e.g., rain, earthquakes, currents) or anthropogenic noise is present (e.g., shipping). This poses an additional challenge for the development of detection algorithms. Furthermore, the performance of the available models cannot be compared because they are tested on different datasets.

That is why recently a benchmark [11] was published so ML algorithms for baleen whale detection could be evaluated and compared. This benchmark is to be evaluated on the dataset published in Miller et. al (2021) [12]. It proposes new metrics for reporting the performance which are suited for long-term data with sparse-occurring events. Furthermore, it proposes that the model should be tested in a different year-location to mimic a real-world scenario.

Here we developed a model using transfer learning from a computer vision object detection algorithm, YOLOv8 [13], and applied it to the benchmark. Using an object detection instead of the more common approach of segmentation (single or multi-label) solves the problem of half-calls, which is when only a small part of a call is present in a segment but the entire segment is fed to the model as a positive detection. It allows for more flexibility to detect multiple calls at the same time, and for single call count instead of positive windows count. Moreover, it is more similar to the human process of annotating sound events, such as when using Raven Pro [14], where the user draws bounding boxes around the detected calls.

6.2 Methods

6.2.1 Data

The Miller dataset [12], contains annotated blue and fin whale calls from 11 different site-year combinations. This dataset represents a multi-class detection scenario for baleen whale vocalizations in a real-world long-term dataset with a realistic sparseness of calls. When dividing the data into 15-second segments, different stations contain between 73 and 99.9% of background noise segments (average 91%, median 94%). The annotations comprise seven vocalization categories from Antarctic blue whales and fin whales: the Antarctic blue whale Z-call represented by the A-call (A), B-call (B), and the entire Z-call (Z), the Antarctic blue whale D-call (D), the fin whale 40Hz-downsweep (Dswp), and the fin whale 20Hz-pulse represented by the 20Hz-pulse (20Hz) and the 20Hz-pulse plus overtone (20Plus). The representation of these seven vocalization categories within the entire dataset and the different site-year combinations is highly imbalanced (Table 6.1).

To provide the input to the image classification model, we converted all the recordings to spectrograms using an overlapping window. Deciding the parameters to generate spectrograms is critical, not only for the model but also when manually annotating. Depending on the duration and bandwidth of a vocalization different temporal and spectral resolutions in the spectrogram are necessary to properly resolve the specific time-frequency shape of each vocalization. These parameters are call-specific and it is necessary to have previous knowledge and expertise to achieve spectrograms where the foreground sounds are contrasting the background

Table 6.1 Dataset composition from Miller et al. [12] for the 11 site-years and seven vocalization categories ('A', 'B', 'Z', 'D', 'Dswp', '20Hz', '20Plus'). Note the natural imbalance in the dataset. Table from Schall et al. (2024) [11]. Is. stands for Island.

Site-year	Vocalization categories						
	A	**B**	**Z**	**D**	**20Hz**	**20Plus**	**Dswp**
Maud Rise 2014	2,191	37	28	70	23	5	6
Greenwich 2015	827	157	29	66	2	1	46
Kerguelen 2005	812	237	166	435	788	78	444
Kerguelen 2014	2,557	1,177	563	435	1,920	1,826	344
Kerguelen 2015	1,970	542	236	1,180	552	718	344
Casey 2014	3,681	1,398	1,091	679	17	0	0
Casey 2017	1,741	558	119	553	78	214	0
Ross Sea 2014	104	0	0	0	0	0	0
Balleny Is. 2015	923	44	31	46	951	148	78
Elephant Is. 2013	2,447	1,672	141	10,600	3,266	1,599	965
Elephant Is. 2014	6,934	967	100	1,034	4,940	2,912	4,077
Total	24,189	6,791	2,506	15,100	12,539	7,503	6,306

in a well-defined and sharp way. Furthermore, it is also necessary to chose the sampling rate, the frequency bandwidth, the length of each window to analyze and the window overlap. These parameters need to be chosen so the longer vocalizations fit in one window but the shortest are still visible. The chosen parameters for this context situation are specified in Table 6.2. The raw data was first filtered using a IIR bandpass filter of order 20 from 5 to 124 Hz. Once the spectrogram was created it was normalized to the [0, 98] % percentiles, over the entire spectrogram, without converting to dB. Then it was converted to a gray scale, where white is 0 and black 1.

Table 6.2 Settings used to generate the spectrograms for the Miller dataset. Parameters chosen to match those of the benchmark.

Parameter	**Explanation**	**Value**
chunk duration [s]	duration of the chunk to analyze	30
chunk overlap [%]	how much to overlap between the images fed to the model	50%
sampling frequency [Hz]	sampling frequency to resample to (if sampling frequency higher)	250
nfft	number of Fourier Transforms	2048
window length	length of the window to apply the FFT to	256
window overlap	in %, overlap of fft windows	83.6%
window shape	name of the shape used as a window	Hamming

Due to the inability to distinguish between certain calls, we grouped various annotated calls into three distinct categories, as specified in the benchmark: ABZ including the calls A, B and Z; 20Plus20Hz including 20Hz+ Pulse and 20Hz Pulse; and DDswp including the D call and the down-sweep.

6.2.2 Model description and training

The YOLOv8 model incorporates several data augmentation techniques:

- image HSV-Hue augmentation (fraction)
- image HSV-Saturation augmentation (fraction)
- image HSV-Value augmentation (fraction)
- image rotation (+/- deg)
- image translation (+/- fraction)
- image scale (+/- gain)
- image shear (+/- deg)
- image perspective (+/- fraction), range 0-0.001
- image flip up-down (probability)
- image flip left-right (probability)
- image mosaic (probability)
- image mixup (probability)
- segment copy-paste (probability)

The augmentation techniques for mix-up, copy-paste, mosaic, rotation, shear, perspective and scale were disabled because they did not represent any realistic case in the underwater acoustic world, and they might produce outputs which a human would then not label as a detection. All the other parameters were used with the default values provided by ultralytics, except the Intersection over Union (iou) which was set to 0.3. The images were resized to 640 pixels. A list of the values of the parameters is provided in the Supplementary Material S6.1. The pre-trained model YOLOv8 (nano version) was used as an initialization. It is trained on the Common Objects in Context (COCO) dataset [15], which is a large-scale picture dataset containing common objects. The Raven Pro annotations were converted to the YOLOv8 format, and all the samples from the Balleny Islands (2015) were left out as a test set. All the samples from the Casey station (2014 and 2017) were used as a validation set. The model was then trained on all the other available data for 100 epochs, with a batch size of 32.

6.2.3 Model evaluation

To evaluate the model we converted the Yolov8 predictions to Raven format. Each YOLOv8 prediction has a confidence score which expresses how likely the model considers that prediction to be correct. Because of the 50% overlapping windows, every call could be detected two times. Therefore, before evaluating the model we merged all the detections which were overlapping an intersection over union (iou) of 50% or more, keeping the largest boundaries resulting from the union of the two boxes. The confidence of the new box was assigned as the highest confidence of all the merged detections. The pseudo-code for this routine is specified in Algorithm S6.1 (Supplementary Material).

Then all the predictions were evaluated, and if there was at least one ground truth detection of the same class overlapping with the prediction at least 20% (iou of 0.2), it was considered correct. All the ground truths complying with these conditions were marked as detected.

All correct predictions were labeled true positives (TP) while all other predictions were labeled false positives (FP). All the ground truth detections which were not marked as detected were labeled false negatives (FN). The true negatives (TN) of one class were considered the number of images which did not contain a ground truth detection nor a prediction of that class. From the TP, TN, FP and FN the recall and false alarm rate were computed for different confidence values, and the curves were plotted.

The metrics specified by the benchmark were also computed. These are the following:

$$\mathrm{TCR} = \frac{\mathrm{TP}_1}{\mathrm{ABZ}_{\mathrm{total}}} + \frac{\mathrm{TP}_2}{\mathrm{20Plus20Hz}_{\mathrm{total}}} + \frac{\mathrm{TP}_3}{\mathrm{DDswp}_{\mathrm{total}}} \tag{6.1}$$

$$\mathrm{NMR} = \frac{\mathrm{FP}_{n1} + \mathrm{FP}_{n2} + \mathrm{FP}_{n3}}{\mathrm{Noise}_{\mathrm{total}}} \tag{6.2}$$

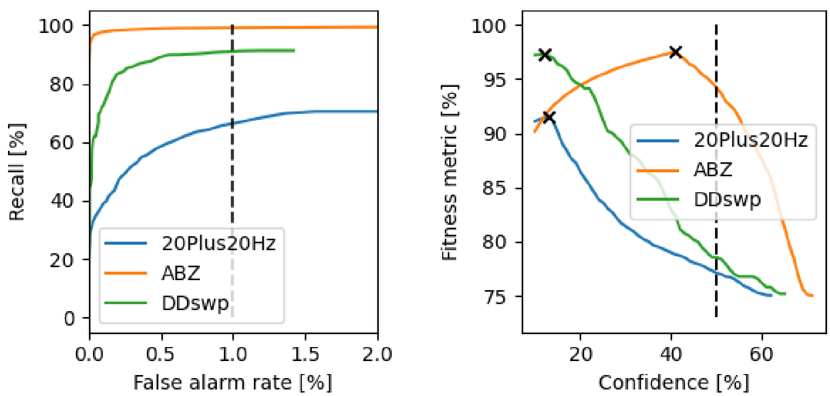
$$\mathrm{CMR} = \frac{\mathrm{FP}_{c1}}{\mathrm{ABZ}_{\mathrm{total}}} + \frac{\mathrm{FP}_{c2}}{\mathrm{20Plus20Hz}_{\mathrm{total}}} + \frac{\mathrm{FP}_{c3}}{\mathrm{DDswp}_{\mathrm{total}}} \tag{6.3}$$

where,
TCR is the True Classification Rate, NMC is the Noise Misclassification Rate, CMR is the Call Misclassification Rate, FP_{ni} are the false positives from class i which did not overlap with any other call more than 20%,
FP_{ci} are the false positives from class i which did overlap with a call from another class more than 20%,

6.3 Results

The training set had 67,491 images of 30 seconds, a total of 281h of audio. The validation set had 91,394 images of 30 seconds, a total of 380h of audio. The test set had 48,952 images of 30 seconds, a total of 204h of audio files. In these images there were a total of 1,099 20Plus20Hz annotated calls, 998 ABZ calls, and 125 DDswp calls. The model results on the test set are shown in Figure 6.1. It can be seen that the model performs best for ABZ and DDswp, and worse for 20Plus20Hz. At a false alarm of 1%, the recall of ABZ is 99.05%, for DDswp is 90.83% and for 20Plus20Hz it is 66.77%. Furthermore, the confidence to achieve a 1% false alarm rate for 20Plus20Hz and DDswp is 0.16 and 0.15, respectively, while for ABZ is 0.6.

Figure 6.1 Results of the object detector model on the Miller dataset. These results are from the test set, where only data from Balleny Islands was used for testing and not included in the training nor the validation dataset. **(a)** These curves represent the false alarm rate and recall at different confidence levels. Dotted vertical line marks 1% false alarm rate. All the results were computed with a iou of 0.2. **(b)** These curves represent the fitness metric specified at the benchmark at different confidence levels. All the results were computed with a iou of 0.2. Black crosses mark the best fitness value (and corresponding selected confidence) and dashed line marks 50 % confidence threshold. Note the different between y-axes of the two figures.



(a) (b)

Table 6.3 Comparison with the Benchmark results. BI stands for Balleny Islands, and represents te results of the Benchmark only when this station was left out for testing. Results for BI are approximately taken from the Figure 4 of Schall et al. (2024) [11]. A-S stands for ANIMAL-SPOT [16]. C. CNN stands for Custom CNN. OD stands for Object Detector. Best results of the benchmark are in **bold**. The results of the object detector are considering an iou of 0.2. The column “OD0.5” are the results obtained when setting the confidence of all the classes to 0.5. The column “ODsel” are the results when setting the confidence to the value giving a 1% false alarm rate for each call type.

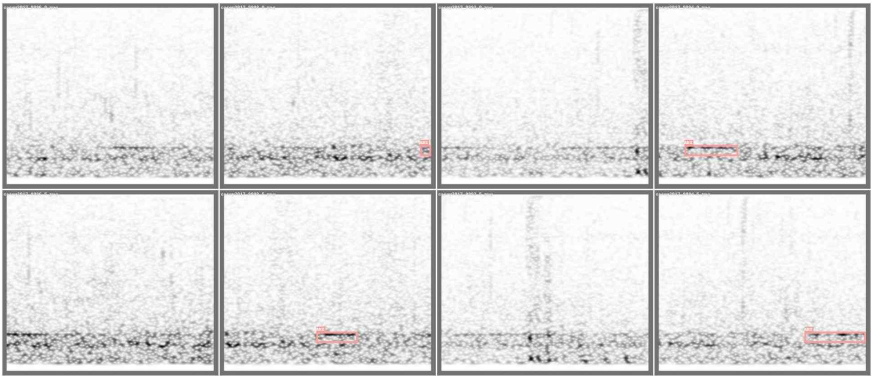
Metric		A-S	C. CNN	OD0.5	ODsel
TCR	average	0.67	**0.73**		
	BI	**0.76**	**0.76**	0.35	0.66
NMR	average	**0.16**	0.24		
	BI	0.06	0.1	**0.01**	**0.01**
CMR	average	**0.04**	0.06		
	BI	0.04	0.05	**0.0003**	0.004
F	average	**0.83**	0.80		
	BI	0.90	0.88	0.83	**0.91**

6.4 Discussion

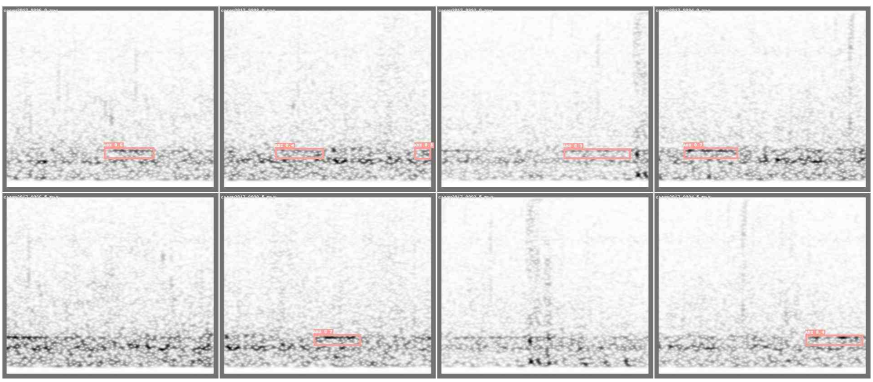
Transfer learning from the YOLOv8 computer vision model was successfully applied to train a model on a long-term dataset comprising data from 7 different locations spanning 4 years (2013-2017). As this chapter was intended as a proof of concept, the model was trained only once leaving one full independent location out. The location to leave out was chosen randomly. The obtained metrics on this independent locaiton outperformed the other models tested in the benchmark. However, according to the benchmark proposed by Schall et al. (2024) [11], evaluation metrics should be reported for multiple models and averaged, with each iteration using one station-year combination as test set. To ensure an adequate comparison with the highest-scoring models from the benchmark, it would be necessary to re-train the model while excluding each station-year combination once.

The model performed well when tested on a fully independent location, with a recall higher than 66% for all the classes at a false alarm rate of 1%. The different classes were learned with different confidence curves by the model. This means that to achieve the 1% false alarm rate, different confidence thresholds need to be specified for each call type. This could be explained by missing annotations on the test location. This is confirmed when looking at the some randomly selected examples where ABZ predictions of the model which are marked as false positives could very well be actual calls which were missed by the annotators (see Figure

Figure 6.2 Example of the comparison between the **(a)** ground truth labels and **(b)** predicted labels. The model might have detected clear calls which were missed during human annotation.



(a) Ground truth annotations from Miller et al. (2021) [12]



(b) Predicted annotations

6.2), a very common problem when manually annotating long-term acoustic data [17]. To accurately assess the real performance of the model, the detections should be manually checked again.

It was previously noted that manual annotations are not consistent among different analysts, and in Leroy et al. 2018 [17] it was shown that the discrepancies between human analysts was more than 50%. Furthermore, human analysts agreed with themselves only between 66% and 88%. It is especially difficult to manually determine single vocalizations within a continuous chorus, and low signal-to-noise calls within a chorus account for a big part of the disagreements. As there is currently no real solution to this, the results of an automatic detector cannot be expected to match at a 100% the ground truth. With recalls ranging from 70% to 85%, the results presented here can be considered within the range of error of human analysts.

6.5 Conclusions

We trained an object detector (YOLOv8) to detect baleen whale calls in an open-source Antarctic dataset, and we evaluated it following the recommendations of the benchmark published in Schall et al. (2024) [11]. This model was intended as a proof of concept for Chapter 7, where instead of labeled events, we aim to detect any acoustic event. An event is defined as any acoustically and visually salient sound on a spectrogram, shorter than half the duration of the exploration window. The obtained performance confirms that YOLOv8 is a valid approach to detect acoustic events in long-term underwater recordings, as the results are comparable to the ones from human annotators. It can also be concluded that the model learned each call type with different confidence levels, and that therefore a call-specific confidence threshold will provide better overall results. Finally, in this chapter we show the importance of using the false and true positive rate as evaluation metrics when faced with highly imbalanced datasets and long-term recordings.

References

[1] E. Schall and C. Parcerisas. *A Robust Method to Automatically Detect Fin Whale Acoustic Presence in Large and Diverse Passive Acoustic Datasets*. Journal of Marine Science and Engineering, 10(12):1831, December 2022.

[2] M. Cosentino, F. Guarato, J. Tougaard, D. Nairn, J. C. Jackson, and J. F. C. Windmill. *Porpoise Click Classifier (PorCC): A High-Accuracy Classifier to Study Harbour Porpoises (Phocoena Phocoena) in the Wild.* The Journal of the Acoustical Society of America, 145(6):3427–3434, 2019.

[3] E. Debusschere, F. Hernandez, and K. Deneudt. *Cpod-Lifewatch: Permanent Cetacean Passive Acoustic Sensor Network in the Belgian Part of the North Sea*, 2015.

[4] D. Gillespie, D. K. Mellinger, J. Gordon, D. McLaren, P. Redmond, R. McHugh, P. Trinder, X.-Y. Deng, and A. Thode. *PAMGUARD: Semiautomated, Open Source Software for Real-time Acoustic Detection and Localization of Cetaceans.* The Journal of the Acoustical Society of America, 125(4):2547–2547, April 2009.

[5] M. Cosentino, P. Schwarzbach, J. Tougaard, D. Nairn, F. Guarato, J. Jackson, and J. F. C. Windmill. *D-PorCCA, a New Tool to Study the Behaviour of Harbour Porpoises*. In World Marine Mammal Conference, ESP, December 2019.

[6] M. Cosentino, C. Parcerisas, K. Brundiers, J. Sarnocinska, D. Botteldooren, P. Devos, M. Dähne, J. Tougaard, M. Wahlberg, and E. Debusschere. *Passive Acoustic Monitoring of Odontocetes: The Devil Is in the Details.*, 2024.

[7] A. N. Allen, M. Harvey, L. Harrell, A. Jansen, K. P. Merkens, C. C. Wall, J. Cattiau, and E. M. Oleson. *A Convolutional Neural Network for Automated Detection of Humpback Whale Song in a Diverse, Long-Term Passive Acoustic Dataset*. Frontiers in Marine Science, 8, 2021.

[8] E. H. Belghith, F. Rioult, and M. Bouzidi. *Acoustic Diversity Classifier for Automated Marine Big Data Analysis*. In 2018 IEEE 30th International Conference on Tools with Artificial Intelligence (ICTAI), pages 130–136, November 2018.

[9] W. Vickers, B. Milner, D. Risch, and R. Lee. *Robust North Atlantic Right Whale Detection Using Deep Learning Models for Denoising*. The Journal of the Acoustical Society of America, 149(6):3797–3812, June 2021.

[10] E. L. White, P. R. White, J. M. Bull, D. Risch, S. Beck, and E. W. J. Edwards. *More than a Whistle: Automated Detection of Marine Sound Sources with a Convolutional Neural Network*. Frontiers in Marine Science, 9, 2022.

[11] E. Schall, I. I. Kaya, E. Debusschere, P. Devos, and C. Parcerisas. *Deep Learning in Marine Bioacoustics: A Benchmark for Baleen Whale Detection*. Remote Sensing in Ecology and Conservation, April 2024.

[12] B. S. Miller, N. Balcazar, S. Nieukirk, E. C. Leroy, M. Aulich, F. W. Shabangu, R. P. Dziak, W. S. Lee, and J. K. Hong. *An Open Access Dataset for Developing Automated Detectors of Antarctic Baleen Whale Sounds and Performance Evaluation of Two Commonly Used Detectors*. Scientific Reports, 11(1):806, January 2021.

[13] G. Jocher, A. Chaurasia, and J. Qiu. *YOLO by Ultralytics*, January 2023.

[14] K. Lisa Yang Center for Conservation Bioacoustics. *Raven Pro: Interactive Sound Analysis Software (Version 1.6.3)*. Ithaca, NY: The Cornell Lab of Ornithology, 2022.

[15] T.-Y. Lin, M. Maire, S. Belongie, L. Bourdev, R. Girshick, J. Hays, P. Perona, D. Ramanan, C. L. Zitnick, and P. Dollár. *Microsoft COCO: Common Objects in Context*, February 2015.

[16] C. Bergler, S. Q. Smeele, S. A. Tyndel, A. Barnhill, S. T. Ortiz, A. K. Kalan, R. X. Cheng, S. Brinkløv, A. N. Osiecka, J. Tougaard, F. Jakobsen, M. Wahlberg, E. Nöth, A. Maier, and B. C. Klump. *ANIMAL-SPOT Enables Animal-Independent Signal Detection and Classification Using Deep Learning*. Scientific Reports, 12(1):21966, December 2022.

[17] E. C. Leroy, K. Thomisch, J.-Y. Royer, O. Boebel, and I. Van Opzeeland. *On the Reliability of Acoustic Annotations and Automatic Detections of Antarctic Blue Whale Calls under Different Acoustic Conditions*. The Journal of the Acoustical Society of America, 144(2):740–754, August 2018.

7

Acoustic Event Detection when the sound sources are not known

This chapter is a reproduction of the publication:

- *Parcerisas, C.; Schall, E.; Te Velde, K.; Botteldooren, D.; Devos, P.; Debusschere, E. Machine learning for efficient segregation and labeling of potential biological sounds in long-term underwater recordings. Frontiers in Remote Sensing - Acoustic Remote Sensing, 2024.* https://doi.org/10.3389/frsen.2024.1390687

For the audio files from the supplementary material, please refer to the online version of this manuscript.

Overview

This chapter provides an example of how to analyze soundscapes using sound event detection and classification. We apply the same technique presented in Chapter 6 to the stationary dataset explained in Chapter 2. The particularity of this method is that the sound sources are not known, and therefore this chapter is intended to answer research question:

Is it possible to disentangle different biological sound events from long-term recordings, and can machine learning assist in detecting and classifying such events?

7.1 Introduction

The technological advances in Passive Acoustic Monitoring (PAM) underwater devices in recent years have enormously increased the amount of marine acoustic data available. Studies carried out using these data typically focus on a single or a limited number of species, mainly concentrating on taxa at the top of the food chain [1, 2]. Archived long-term data, however, contain a great diversity of other sounds, most of which remain to date unidentified. Interest in studying these sounds has grown in recent years, as they can serve as a proxy for biodiversity or ecosystem health [3, 4, 5, 6].

Sound events can inform animals about their surroundings [7]. This can either come in the form of biotic associated sounds from predators, prey, or conspecific, or in the form of geophonic sounds that can contain information about habitat quality or provide navigational cues [8, 9]. Since any sound event could potentially carry information about an organism's environment, characterizing and quantifying unknown sound events can be used to characterize and understand soundscape components. Soundscape characterization has been done by detecting certain acoustic events of relevance, such as animal vocalizations or anthropogenic sounds, and quantifying their temporal patterns, relationships, or proportions [8, 10]. This provides knowledge on the local acoustic community and it can be used as a proxy for biodiversity and ecosystem health. Soundscape characterization by isolating acoustic events can also be done with sounds from an unknown source, as long as they can be detected and classified. Finding, reporting, and understanding the patterns of unidentified sounds is of significant benefit to the assessment of underwater soundscapes. This can then help raise awareness and inform policymakers on the health status of an ecosystem and how best to tackle conservation or noise mitigation measures [6].

Some sound sources present diurnal, celestial, seasonal and annual patterns, especially studied for biological sources [11, 12, 8]. For example, certain fish species are known to vocalize during dusk and dawn [13]. Therefore, some sources could eventually be assigned to certain sound events by an exclusion procedure taking into account their spatio-temporal patterns, but to do this larger scales than the ones that can be covered in one study/area are necessary. Therefore, a database of unidentified sounds is, in some ways, as important as one for known sources [6]; as the field progresses, new unidentified sounds will be collected, and more unidentified sounds can be matched to species. Therefore, documenting these sounds before they are identified provides a baseline for their presence and supporting information for later source identification. This is especially applicable in areas where very little sound sources are clearly described.

However, studying unknown sounds is a challenging task, as it is difficult to find sound events in long-term recordings when one does not know which events to expect. Unidentified sounds might be, or might not be, of importance for the marine fauna. Yet before deciding that a certain sound is relevant, a considerable time investment in the manual screening of the acoustic recordings is necessary, and this can still sometimes be inconclusive [14]. In the marine environment, this task is even more complicated as biological sounds of interest are often sparse-occurring [15], non-continuous or rare [16]. Because of the amount of generated PAM data, there is interest in having this process automated. Several studies suggest that using deep learning is a promising solution [1]. In these studies, the detection and classification are often applied to segmented data, where long recordings are split into equal-sized overlapping windows, and then a binary output algorithm is used to detect the possible sound events [15]. Afterwards, the selected windows are run through a classifier where they are assigned a call type, or further discarded as noise. However useful this approach can be, it has its limitations. For example, it is complicated to detect and classify signals of different lengths, or to deal with signals overlapping in time in different frequency bands. When looking for unidentified sounds, these considerations are key, as there are no predefined frequency bands, frequency patterns or event duration to focus on.

Here we propose a method to detect and categorize sound events in long-term recordings. The method concept is inspired by the analysis process of human annotators when screening for unidentified sounds. Human annotators look first at the temporal-spectral shape in a spectrogram, the duration, and the frequency limits to assign certain sounds to a specific species. The sounds are then usually annotated by drawing bounding boxes around them in the time-frequency domain, namely in a spectrogram. Human annotators first screen a lot of hours of recordings before deciding which sound groups can be considered, and only after that are labels assigned to the selected sounds. Therefore, following the same strategy, we propose to use of one of the newest computer vision algorithms for object detection YOLOv8 [17], to detect all the possible sound events on a spectrogram using transfer learning.

Supervised deep learning models such as the proposed detection model (YOLOv8) are known to need large amounts of annotated data to achieve good performances. Hence, to reduce the human annotation effort needed to generate a first dataset to train the model, we propose an active learning approach, where the model selects the files which could be more beneficial for the model to be manually annotated. We then compare the results with a random selection of files for annotation. To test the robustness of the model to unseen data, and to investigate the possible generalization of the model to detect acoustic events in any underwater environment, the obtained models are tested on two datasets: (1) a test set recorded in the Belgian Part of the North Sea (BPNS) as part of the LifeWatch Broadband Acoustic Network [18], and (2) test set of freshwater acoustic recordings collected in 4 major European rivers.

The detection model can then be used to detect sound events in new data. The obtained detections could already be directly used to speed up manual annotations, but they can also be used to further cluster these sound events into sound types and explore the acoustic environment. To this aim, all the detected events are converted into a multidimensional embedding space using another pre-trained deep learning model, which has been trained on a large dataset of diverse bioacoustic data. We then use these embeddings to cluster all the detections in an unsupervised way. This approach enables an initial analysis of the existing sound types within a specific dataset. Through a manual review of the clusters, sound labels can be assigned to them if considered appropriate. Once these clusters are defined and revised, we analyze the obtained temporal patterns and the new sound categories discovered. We showcase this second part of the methodology in a short deployment spanning 10 days in one location of the BPNS.

7.2 Methods

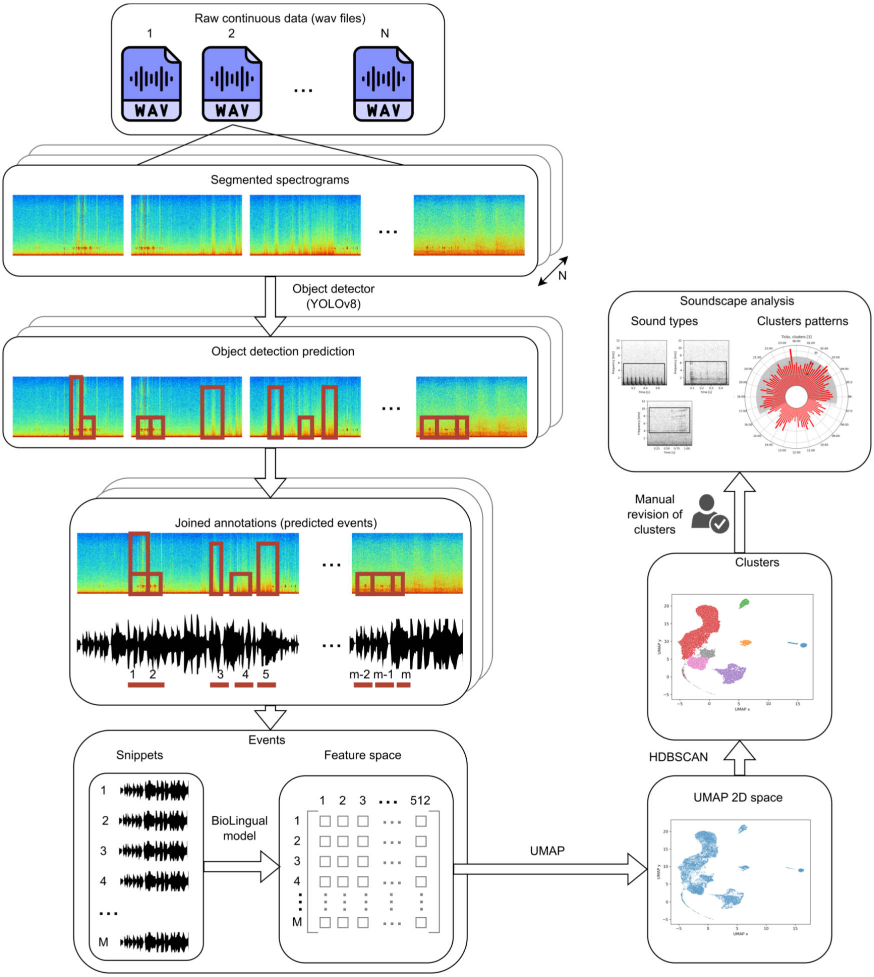
7.2.1 General concept flow

The presented methodology is a new approach to discover sound types of a relatively unstudied environment while reducing human annotation and labeling time, providing insights on the spatio-temporal patterns of the discovered sounds leading to potential clues on sound origins.

The general idea of the proposed methodology is to first detect all the potentially relevant underwater acoustic events, regardless of their sound type, using an automated method. Next, the detected events are converted to a multidimensional embedding space and then clustered into different classes. The clusters are then manually revised by checking 10 events per cluster. Finally, the temporal patterns of the obtained clusters are plotted to assist in the identification of the source of each sound type (cluster) and to already provide insight on the soundscape dyamics. A general schematic of the entire process can be seen in Figure 7.1.

For the event detection, because we want to allow for multiple events happening simultaneously in different frequency bands, this detection is performed in the spectro-temporal space using an object detector (YOLOv8) [17] from computer vision. Therefore, the recordings needed segmentation and transformation into images, in order to be ingested in the model. This process is explained in Section 7.2.4. To account for the continuity of the data, these segments are overlapping. Therefore, the predictions from the object detector need to be merged afterwards to avoid double detections. This is explained in Section 7.2.5.2. We will refer to the model predictions once they are already merged as sound event detections from now on. Although YOLOv8 is pre-trained, specialization for the task at hand is needed, so the model needs to be re-trained (see Section 7.2.5.1). This requires a

Figure 7.1 Flow of the proposed methodology. N is the number of wav files to analyze. m is the number of detections in one wav file after merging. M is then the total number of annotations within the N wav files.



human selection of areas in the spectrogram that are potentially interesting sounds, which is a time-consuming task (see Section 7.2.3 for details on how these areas are selected). We will refer to this process as human annotation throughout the manuscript. To increase efficiency of this process we propose an active learning approach where audio files are selected that could enrich the database of human annotations most. To this end, suitable metrics are proposed. This process is described in Section 7.2.6, and it is compared to the random selection of files. The performance of the object detector is tested on two independent manually annotated datasets: an extensive dataset from the BPNS (see Section 7.2.4.3) and a dataset from freshwater recordings (see Section 7.2.4.4).

Once all the overlapping predictions are merged, using the start and end time of each sound detection the raw waveform snippet is extracted and filtered to the predicted frequency band (frequency limits predicted by the model). Each snippet is then converted to a multidimensional embedding space using the pre-trained model BioLingual [19] (see Section 7.2.7 for more information on the model). The obtained features are next reduced to a smaller feature space using UMAP to deal with the curse of dimensionality, and the reduced feature space is clustered using HDBSCAN. The obtained clusters are then manually revised to assign them a label and a possible source (from now on, labeling), and their temporal patterns are analyzed. This process is explained in Section 7.2.7, and it is performed as a showcase on data from continuous recordings spanning 10 days.

7.2.2 The datasets

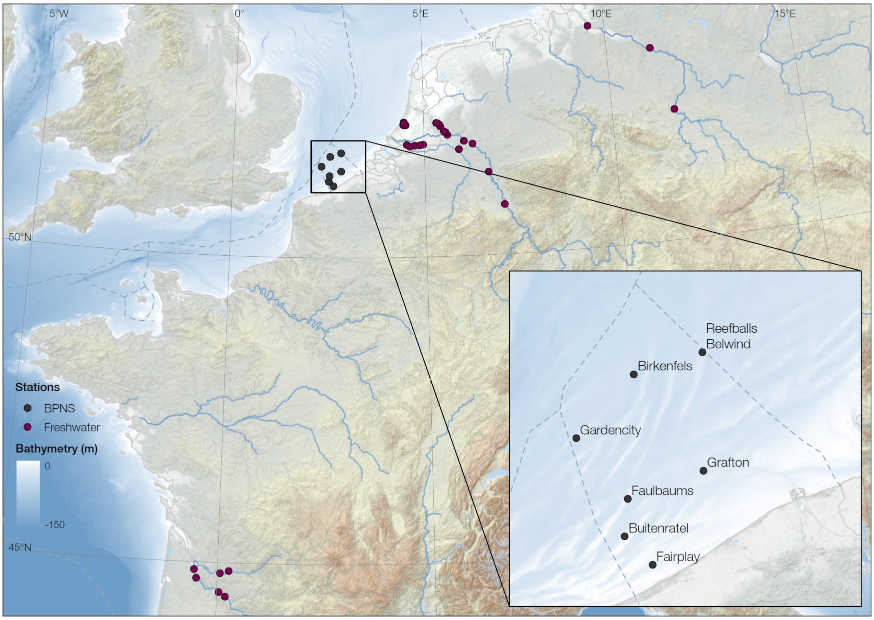
7.2.2.1 BPNS Data

The audio data were selected from the Stationary dataset explained in Section 2.2. The locations of data collection within the BPNS are displayed in Figure 7.2. The considered deployments from the BPNS dataset are the ones marked with a 1 on the HMB column in Table 2.1 (Section 2.2, from Chapter 2). Each deployment was manually screened to decide the period where the data were valid, considering clipping, instrument noise and failure of the recorder. The files considered in this study were all the files falling inside the valid period of a deployment and had less than 1E-6 percentage of data points clipping. All the files were between 5 and 10 minutes long, depending on the deployment.

7.2.2.2 Freshwater Data

To further evaluate the robustness of the detection algorithms the model was tested on an extra test set recorded in a variety of freshwater habitats across Europe. The locations of data collection within Europe are displayed in Figure 7.2. These habitats included ditch, pond, medium river, large river, and 4 very large European rivers with varying characteristics. The total dataset included 42 different deployments,

Figure 7.2 Locations where the sound was acquired, for both the BPNS and the freshwater datasets. Area zoomed to the BPNS with the corresponding station names. Data used from [21] and [22]



each at a different location, recorded with two different instruments. On one side, 28 deployments were conducted using SoundTrap 300 STD hydrophones (Ocean Instruments NZ, sensitivity: -176.6 dB re: 1 μPa V^{-1}, frequency range -3 dB 20 Hz to 60 kHz), suspended between an anchor and sub-surface buoy 50 cm above the sediment. The other 14 deployments were recorded using Hydromoth hydrophones (Open Acoustic Devices, unknown calibration specs) [20], attached to a steel frame 20 cm above the sediment. A detailed summary of all the considered deployments can be found in the Supplementary Material Table S7.1.

7.2.3 Manual annotation on audio files using RavenPro

Manually annotating sounds (drawing bounding boxes around acoustic events in the spectro-temporal space) is a time and human labor-intensive task. This is especially the case when a lot of sounds have to be annotated, and when one does not know what sounds to look for, because the whole bandwidth needs to be screened. In this study we focused on a broad frequency band to include benthic invertebrate, fish and some marine mammal sounds. Invertebrate sounds are characterized by a wide bandwidth and very short duration compared to fish sounds [23]. This difference in duration and bandwidth poses a challenge during annotation and description of the sounds.

As ground truth to train and test the sound event detection model, wav files were manually annotated using Raven Pro version 1.6.4 [24]. The software settings were configured to visualize a window duration of 20 seconds with frequencies ranging from 0 to 12 kHz. To facilitate optimal visual representation, the selected color scale was 'Grayscale', and the spectrograms were generated with a Hann window with 2048 FFT bins and a hop size of 164 samples (same parameters than the ones used afterwards to convert the data into images to input to the model). Spectro-temporal bounding boxes were meticulously hand-drawn to accurately capture the contours of the corresponding audio signals as observed in the spectrogram.

Because of the subjectivity of annotating sound events, some rules were decided about how to label events. The first requirement for an event to be logged was that it was both acoustically and visually salient. Therefore, sounds perceived as 'background' were not annotated. This included ambient sound but also long, continuous sounds not salient according to subjective human perception. Deciding if a sequence of sound events was a 'sentence' or separate individual sounds was done following the subjective criteria of whether the events were perceived to be coming from the same sound source or not, focusing on the continuity of the sound. Abrupt frequency jumps were considered an indication of the start of a different event. Events happening simultaneously at the same time with the same rhythm in different frequency bands were annotated as a single event. In case of doubt, a separate box was always added.

7.2.4 Data preparation for object detector

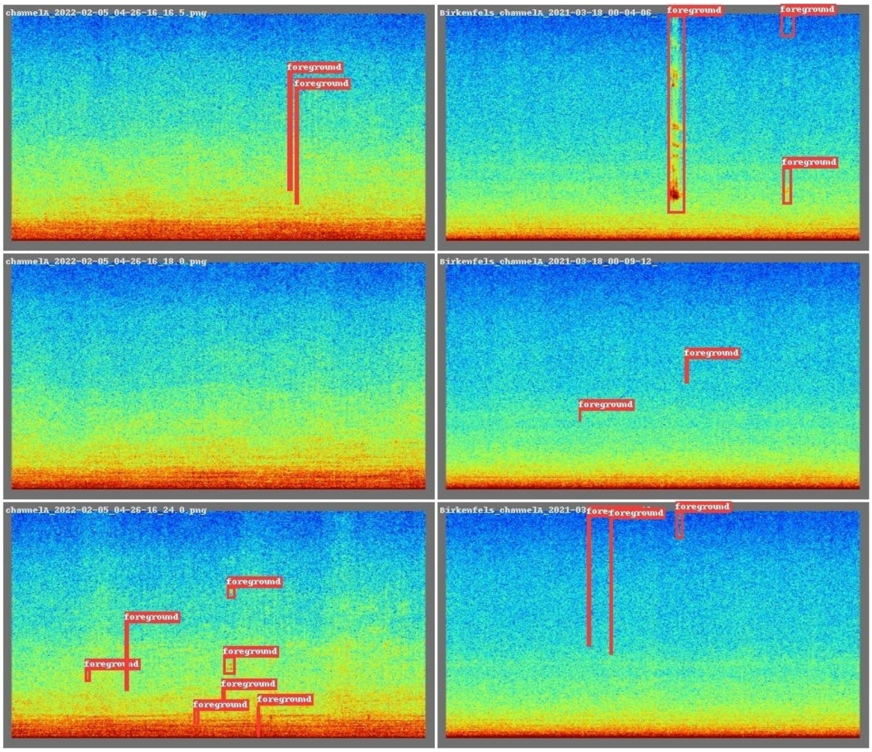
The object detector (YOLOv8) is based on visual detections of the sound events. Therefore, the data were processed into spectrograms using overlapping windows longer than the expected sound events of interest. Deciding the parameters to generate spectrograms is a critical step. All these parameters are context-specific and should be chosen in a way that the foreground sounds are contrasting the background in a well-defined and sharp way. The chosen parameters for this context situation are specified in Table 7.1. Once the spectrograms were generated, a spectral high pass filter at 50 Hz was applied to exclude flow noise. Then they were normalized to the [1, 99] % percentiles after converting to dB. Normalization was done to exclude very loud and short sounds (particles hitting de hydrophone) would not compromise the visualization of other sound events. Finally, the spectrograms were converted first to gray scale, where white is 0 and black 1 and then converted to RGB using the colorscale 'jet' provided by the package matplotlib (Python). 'jet' was selected because it is an RGB color conversion commonly used by bioacousticians. This step was done because the pre-trained YOLOv8 model was trained on RGB images. Further work would be necessary to asses the impact of the color conversion on the model performance. Data were processed using the scripts available at https://github.com/lifewatch/sound-segregation-and-categorization, and Raven annotations were converted into the YOLOv8 format for each spectrogram (segment). An example of several input images with their corresponding annotations is shown on Figure 7.3.

Table 7.1 Settings used to generate the spectrograms for the YOLOv8 model

Parameter	Explanation	Value
chunk duration [s]	duration of the chunk to analyze	20
chunk overlap [%]	how much to overlap between the images fed to the model	50%
sampling frequency [Hz]	sampling frequency to re-sample to (only files with a different sampling frequency)	24000
nfft	number of Fourier Transforms in one chunk	2048
window length	length of the window to apply the FFT to, in samples	2048
window overlap	in %, overlap between the windows	92%
window shape	name of the shape used as a window	Hann
temporal resolution	computed from window length and window overlap, in seconds	0.042
frequency resolution	depending on nfft, in Hz	11.71

For manual annotations (not model detections), the column SNR NIST Quick (db) of Raven Pro was added as a proxy for the signal to noise ratio (SNR) of the event, to provide a more objective threshold to include an event or not. This was done because human annotations of sound events are not always consistent [25, 26], and "the accuracy of a trained model heavily depends on the consistency of the labels provided to it during training" [27]. Therefore, the post-annotation filtering of SNRs provided a less subjective criteria on whether to add or not an annotation to training or test sets. With this idea, box-annotations with a SNR NIST Quick (db) lower than 10 dB were discarded. For both manual annotations and model detections, all events shorter than 1 pixel in temporal resolution (shorter than 0.085 seconds) were also removed before the reshaping.

Figure 7.3 Example of labeled sounds from the unidentified sounds Dataset when colored to RGB values. Each rectangle represents 20 seconds in the x axis and 12000 Hz in the y axis.



7.2.4.1 Initial training set

For the initial training set, approximately 1.5 hours were used from the Birkenfels station, recorded the 18th of March of 2021, from midnight to 1:30am. The annotations were carried out by expert E.S.. After processing the raw data, the training data consisted of 556 images in RGB of a size of 1868 x 1020 pixels, representing 20 seconds each. From the 556 images, 14 had no annotation (background images). In total there were 1595 annotations, from which 1532 complied with the SNR and minimum duration criteria. This was chosen as a representative starting scenario, where the available annotations from a lab are consecutive files from one location.

7.2.4.2 Pool of data for selecting additional training samples using active learning

A pool of the data was created to avoid predicting the entire dataset at every active learning iteration, which was computationally not feasible. The unlabeled pool for active learning was selected in a stratified fashion considering season, station, moment of the day and moon phase. Season included the four different seasons, moment of the day considered twilight, night and day, station consisted of the seven different stations of LifeWatch Broadband Acoustic Network, and moon phase included new, full, growing and decreasing moon states. The python package Skyfield [28] was used to assign the environmental variables to each wav file. Seven files were randomly selected per available combination from all the available wav files of all the recordings, excluding all the files that had been selected for the training set, leading to a total of 1,005 wav file (126.5 hours).

7.2.4.3 BPNS Test set

The BPNS test set consisted of a stratified selection of files from the LifeWatch Broadband Acoustic Network. The selection strategy was the same than the one for the unlabeled pool but selecting 1 file per possible combination of environmental variables instead of 7, and excluding all the files from the unlabeled pool and from the training set. The test set was independent of the training set, but it did overlap with the training set regarding location, season and environmental conditions. The final selection consisted of 145 wav files, a total of approximately 18 hours. The audio files were processed the same way than the training set, leading to a dataset of 6,342 images. The annotations were carried out by independent annotators K.M and O.S., each annotating half of the test set. The annotations were manually checked using a model-assisted approach to speed up the process. We used the model obtained from training using only the initial training set (Model Base) to predict the files. Then, the human annotator went through all the files, adding and removing detections, or modifying the boxes boundaries when necessary. Only the manual annotations complying with the selection criteria were used for evaluation.

7.2.4.4 Freshwater test set

The freshwater test set was also selected for evaluation of the model from all the freshwater data in a stratified way considering water type and moment of day. Moment of day included day, night and civil twilight. The stratified selection was run using the same approach than for the BPNS test set. 24 files of 5 minutes were selected, from which 21 were recorded with SoundTrap and 3 using Hydromoth. The freshwater test set was annotated by expert K.V.

7.2.5 Object detector model

7.2.5.1 Training

The pre-trained YOLOv8n (nano) model was used as an initialization, which was initially trained on the Common Objects in Context (COCO) images dataset [29].

First this model was re-trained on the initial training set of spectrogram images. From now on we will refer to this model as the Model Base.

For all the training runs, the initialization was kept to the YOLOv8 nano weights. The initial training set was split for training and validation using a K-fold strategy with 3 folds (6 different full files were kept for validation for each model). This led to 3 different Model Base.

For each training round, data were fed into the YOLOv8n model and trained for 200 epochs, with batch size 32. The YOLOv8 model incorporates several data augmentation techniques. The augmentation techniques for mix-up, copy-paste, mosaic, rotation, shear, perspective and scale were deactivated for the re-training and the prediction on new data because they did not represent realistic scenarios in the case of object detection in spectrograms of underwater sounds, and therefore they were not expected to create any advantageous for spectrograms, or could even be detrimental. The rest of the augmentation techniques were kept as the default values. The Intersection Over Union (iou) was set to 0.3 for validation evaluation, and the images were resized to 640x640 pixels. The rest of the parameters were kept as the default values. This resizing changed the initial spectrogram resolution, providing a final temporal resolution of 31.25 ms and a spectral resolution of 18.75 Hz.

7.2.5.2 Merging predictions: from segmented images to continuous audio

We used a minimum confidence of 0.1 for all the predictions. This is the default value used by YOLOv8 for validating the model. Because of the 50% overlap between two consecutive images, some model predictions would be repeated when merged as a continuous audio file. Therefore, we first merged all boxes that had a 50% overlap or more, keeping the largest boundaries resulting from the union of the two boxes. The confidence of the resulting box was assigned to the maximum of the box. The pseud-code to merge the boxes is shown in Algorithm S6.1 (Supplementary Material).

7.2.5.3 Evaluation

The evaluation was done once the detections were already merged. When analyzing sounds using an object detector for unknown sounds, the evaluation metrics are not straight forward. The sound events selected in the ground truth are subjectively split into units or merged, according to the best criteria of the human annotator. Sound events occurring simultaneously in different frequency bands can be considered two different sounds or the same sound, and marked accordingly, but all these options should be considered valid when evaluating the model.

To compute the True Positives, each detection d was compared with all the manual annotations starting and ending between ($d_{\text{start\_time}}$ - 5) seconds and ($d_{\text{end\_time}}$

+ 5 seconds). 5 seconds was chosen as the longest detections were set to 10 seconds. This selection was done for computational efficiency. For the comparison, the iou was computed between the detection and all the manual annotations within the respective time window. If any iou was greater than 0.3, the detection was marked as a true positive. Detections without an iou value greater than 0.3 were considered false positives. Manual annotations not exceeding an iou of 0.3 for any prediction were considered false negatives. From true positives, false positives and false negatives, we computed recall, precision and F1 metrics.

To gain more information on the performance (i.e., to evaluate if the errors made by the model were in the time and/or the frequency dimensions), three additional metrics were computed considering the overall area detected:

- detection percentage (time/area): the total percentage of time/area correctly highlighted by the model (detections) divided by the total time/area of all the manual annotations
- true negative percentage (TNP) (time/area): total percentage of time/area correctly not highlighted by the model divided by the total time/area
- false positive percentage (FPP) (time/area): total percentage of time/area incorrectly highlighted by the model (detections) divided by the total time/area

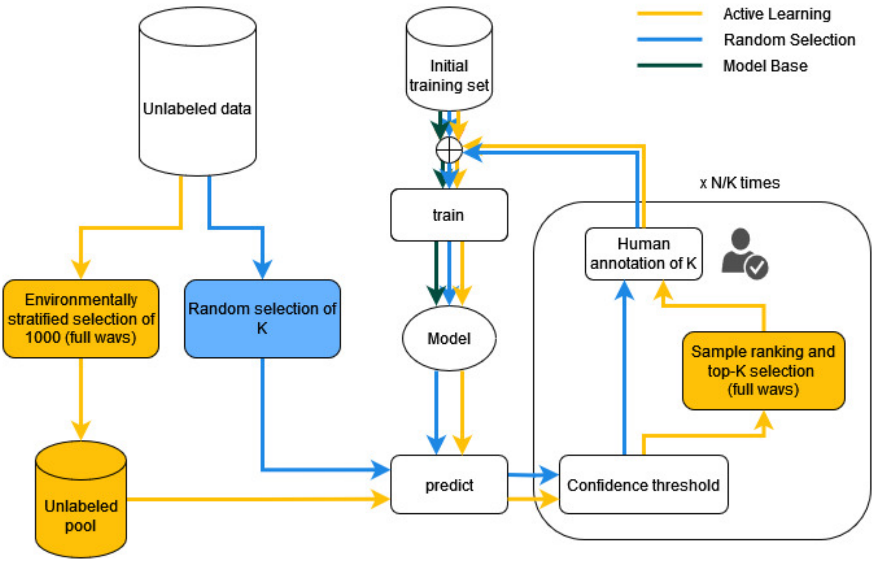
7.2.6 Extending the training dataset

To evaluate the performance of the foreground events detector when adding more annotated data to the training iterations, several approaches were compared:

- Model trained on the training set, without adding any data (Model Base)
- Model trained on all data available from the BPNS (initial training set, added annotations from both the random and active learning selection approaches, and test set) (Model Final)
- Random sampling of the additional files to annotate, with model-assisted annotation
- Active learning annotations, with active selection of the files to annotate, with model-assisted annotation

For the later two approaches where data were added, a maximum annotating budget of 10 wav files was set. All the extra selected files were cut to 5 minutes duration. Each selected file was annotated by one annotator and revised by another to reduce bias on annotations (C.P. and J.A.). The flows of each approach can be seen on Figure 7.4. These two approaches (active learning and random sampling) were run 3 times, each starting with each of the 3 trained Model Base.

Figure 7.4 Flows of the three different compared approaches of the Object detector model to add more data. The Model Base is the model obtained on iteration number 0 on any of the two flow charts (training only on the initial training set).



A last model using all the available annotated data from the BPNS was trained. Its performance was evaluated only on the freshwater dataset.

Annotating boxes is time-consuming, and using pre-annotated boxes has been found to increase annotation speed and improve model performance on other object detection tasks [30]. Hence we used a model-assisted annotation strategy to revise and correct predictions instead of manually adding all the sound events from scratch. Even though from a machine learning perspective it would be more efficient to select individual 20-second snippets from different files rather than full wav files, this is not a common practice for bioacousticians. The process to select more files for each approach is explained in the following sections.

7.2.6.1 Random Sampling

For the random sampling approach, 10 wav files were randomly selected from all the available files, that were not part of the training or test sets. These files were converted into images as explained in Section 7.2.4. The images were then predicted using the Model Base and the output was transformed to a Raven-compatible format as explained in 7.2.5.2. The output of the Model Base was used as initial predictions for model-assisted annotation. The 10 randomly selected files were manually revised and corrected using Raven. Then the 10 selected files were randomly split into 5 groups of 2 to simulate the incremental addition of data. This process was repeated 3 times, one per each Model Base.

7.2.6.2 Active Learning

For the active learning approach, the files to be annotated from the unlabeled pool were determined by the model. This was done by choosing the 2 files scoring the highest following a criteria decided with 3 objectives:

- Find new and rare sounds compared to the training set
- Reduce the uncertainty of the model (i.e., providing more training examples of sounds with high prediction uncertainty)
- Chose a file with a high diversity of sounds

To find rare and new sounds, for each detection in the unlabeled pool we computed the 90th percentile of spectro-temporal overlap with all the detections within the previous training set, as specified in Eqs 7.1 and 7.2. If this value is low, it implies that the overlap with most of the current training dataset is low, and therefore it is a sound event in a new frequency band or a different duration, which is a proxy for the novelty of the sound.

$$\mathrm{iou}_{ij} = \frac{(\min(f_{\text{high},i}, f_{\text{high},j}) - \max(f_{\text{low},i}, f_{\text{low},j}))\min(w_i, w_j)}{A_i} \tag{7.1}$$

$$\mathrm{iou}_i = [\mathrm{iou}_{i,0}, \mathrm{iou}_{i,1}, ..., \mathrm{iou}_{i,n}]; \mathrm{iou}_{i,\text{90th}} = \mathrm{iou}_i(\lceil 0.9n \rceil) \tag{7.2}$$

where,
$f_{\text{high},i}$ is the upper frequency limit of the detection i,
$f_{\text{low},i}$ is the lower frequency limit of the detection i,
w_i is the duration of the detection i,
$A_i = (f_{\text{high},i} - f_{\text{low},i})w_i$ is the area of the detection i,
iou_{ij} is the intersection over union between detection i and j,
n is the number of detections in the training set.

We defined the uncertainty of each detection as $u_i = 1 - c_i$, where c_i is the confidence of the detection i. The number of ’interesting’ sounds in a wav file was then computed considering an uncertainty threshold of 0.75 and an overlap threshold of the 90th percentile, as specified in Eq. 7.3.

$$N_{\text{wav}} = \sum_{i=0}^{m} [u_i > 0.75 \text{ and } \mathrm{iou}_i < \mathrm{iou}_{\text{90th}}] \tag{7.3}$$

where,
iou_{90th} is the 90th percentile considering all the $\text{iou}_{i,\text{90th}}$.

Finally, we computed the diversity of sounds within a file by computing the entropy of the overlapping matrix of all the detections of one wav file (O_{wav} in Eq. 7.4), as specified in Eqs 7.5 and 7.6.

$$O_{\text{wav}} = \begin{bmatrix} d_{0,0} & d_{0,1} & \dots & d_{0,100} \\ \dots & \dots & \dots & \dots \\ \dots & \dots & d_{k,l} & \dots \\ d_{100,0} & \dots & \dots & d_{100,100} \end{bmatrix} \tag{7.4}$$

$$d_{k,l} = \sum_{i=0}^{m} \frac{[f_{\text{low},i} > \frac{f_{\text{nyq}}}{k} > f_{\text{high},i} \text{ and } w_i > \frac{D_{\text{max}}}{l}]}{m} \tag{7.5}$$

$$E_{\text{wav}} = \sum_{k,l} d_{k,l} \ln(d_{k,l}) \text{ if } d_{k,l} \neq 0 \tag{7.6}$$

where,
O_{wav} is the overlap matrix of one wav file with the training set,
$d_{k,l}$ is the overlap computed at frequency index k and time index l,
f_{nyq} is half the sampling frequency (12000 Hz),
D_{max} is the maximum duration of all the detections in the unlabeled pool,
m is the number of detections within the wav file,
E_{wav} is the entropy of the wav file.

Finally, a third component was added to the score to allow the selection of a file based on the presence of one unique sound. We decided to give more weight to adding unseen sounds to the training set as the acoustic richness is reflected by these rare sounds [5]. A rare sound with a very low overlap with the training set and a low confidence in detection, as it should be a shape never seen by the model.

Therefore, the final score of each wav was computed as shown in Eq. 7.7:

$$s_{\text{wav}} = \frac{N_{\text{wav}}}{d_{\text{wav}}} \max\left((1 - \text{iou}_i) u_i \forall i \in \text{wav}\right) E_{\text{wav}} \tag{7.7}$$

where,
d_{wav} is the duration of the file in seconds,
s_{wav} is the score of the wav file.

The overall wav scores were used to select the two files with the highest score at every loop iteration. After 2 iterations, a 30% probability was set of replacing one of the selected files with a randomly selected one. This was done to consider the possibility that not all factors influencing acoustic diversity within the data were considered with this approach, and to not bias the model towards learning on only acoustically diverse files.

The selection of files was done in 5 loop iterations; 2 files were selected each time for annotation, and subsequently removed from the pool of unlabeled data. At each loop iteration, the annotation was carried out using the model-assisted annotation strategy, always using the model obtained at the previous loop iteration for prediction.

7.2.7 Clustering and continuous data analysis

To prove the applicability of the method we applied it in one short deployment from the LifeWatch Broadband Acoustic Network, the deployment from the Grafton station starting the 27th of October of 2022. This deployment consists of a 10 days of recordings with a duty cycle of 50 % (one day on, one day off) at a fixed location at the Grafton station. This is intended as a show case to prove the usability of the proposed methodology, and to illustrate how this pipeline can be used for soundscape characterization.

The deployment's audio data were converted to the YOLOv8 format as explained in Section 7.2.4, and the final model was used to extract a collection of possible sounds events (detections) for subsequent clustering on the 20-seconds images. The predictions were merged as explained in Section 7.2.5.2. A minimum confidence of 0.1 was chosen for predictions being considered as sound event detections.

Using the start and end times of the obtained detections, raw audio snippets were obtained for each detection and converted into a embedding feature space using the pre-trained BioLingual model [19], which is a state-of-the-art model for latent representation for classification of bioacoustics signals across multiple datasets. This model extracts 512 deep embedding acoustic features. The maximum length for a snippet was set to 2 seconds with shorter detections being zero-padded, while longer detections being cut to 2 seconds. Each detection was filtered with a bandpass filter of order 4 to the band of interest of the detected event (between its minimum and maximum frequency).

The BPNS dataset presents a high imbalance between broadband, short, impulsive sounds and other longer, more complex sounds. To avoid obtaining only one big cluster with these events and another one with the rest, all the detections shorter than 0.3 seconds were classified as impulsive sounds and excluded from the clustering. Such short sounds, even though they can be ecologically relevant, are often not classified by their waveform but according to their frequency limits or peak. Therefore acoustic features extracted by the BioLingual model were not expected to provide enough information on further cluster separation for these types of sounds.

For the rest of the detections, the extracted BioLingual features were reduced to a 2D space using UMAP [31] with the number of neighbors set to 10 and a minimum distance set to 0.2. The UMAP dimension reduction was applied to deal with the high-dimensional data resulting from extracting the BioLingual features (512 features), as done previously in Phillips et al. (2018) [32] and Best et al. (2023) [33]. This problem is known as the "curse of dimensionality", and density based clustering algorithms such as HDBSCAN are known to provide low performance in high dimensional spaces. Then the python implementation of HDBSCAN [34] was applied to the resulting 2D embedding space, and the minimum number of samples (events, in this case) per cluster was set to 5 to allow for rare sounds to form a cluster. The epsilon to select the clusters was set to 0.05, and the minimum number of neighbors to 150 based on visual perception of good clusters. All the parameters were selected to get a balance between noise removal and robustness of clusters.

All the obtained clusters were manually revised for possible significance by manually checking a minimum of 10 randomly selected events per cluster. If more than 7 of the revised events were clearly similar sounds, the cluster was assigned a possible source category if previous knowledge was available. The possible categories included pseudo-noise, geophonic, mooring noise, instrument noise, anthropogenic sounds, and biological. When none of these categories could be assigned with certainty, clusters were labeled as unknown. If less than 7 of the revised events per cluster were clearly similar sounds, the cluster was labeled as unclear.

Once all the obtained detections were assigned to a cluster, the different clusters' occurrences were plotted in time to check for diel patterns, adding the sunset and sunrise timestamps to check for dusk/dawn patterns. Furthermore, the temporal patterns were assessed and compared among clusters. This was done by plotting the average percentage of positive detection minutes (minutes where there was at least one detection of that cluster) for each 15 minute bin.

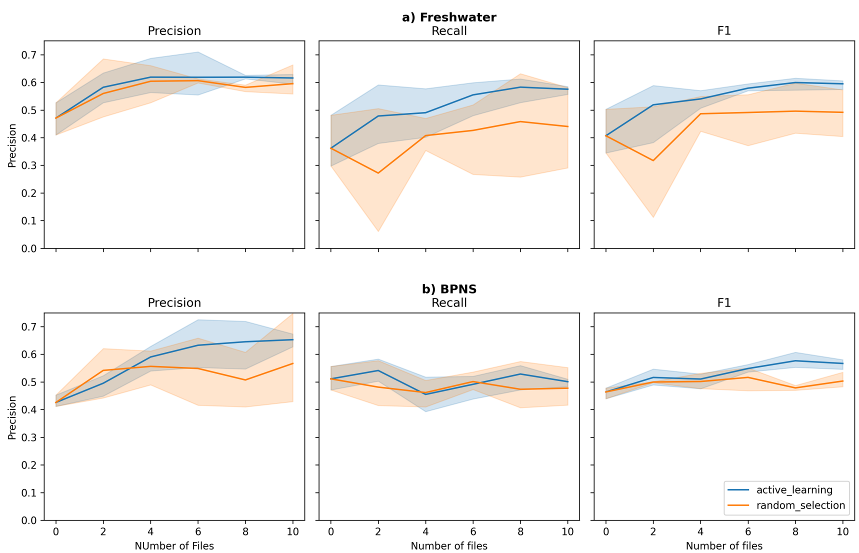
7.3 Results

7.3.1 Detection results

Several models were trained and their performances were compared on the independent test sets: 3 Model Base (MB) using the initial training set, 15 models with incremental training data using random selection (RS), 15 models with incremental training data using active learning (AL) selection, and one final model (MF) trained on all the available annotated data (all the annotations used for training and the test set). Additionally to the model evaluation on the separate BPNS test dataset, obtained models were also evaluated on the freshwater test set to test their robustness to new data.

Regarding the strategy of adding files, the active learning approach presented a faster improvement curve than the random sampling. For the BPNS test set, recall stayed constant for both approaches and did not improve. For the freshwater dataset, precision of the two approaches presented similar values (see Figure 7.5). The performance of the active learning approach converged at the end of file addition, due to the fact that the three repetitions ended up selecting some of the same files. When comparing the average performances of the models from the last training iteration (for active learning, AL5, for random sampling RS5) and the Model Base (MB) on the BPNS test set, AL5 outperformed RS5 and Model Base on 6 metrics, and led to the best F1 score of 56.69% (see Table 7.2).

Figure 7.5 Evaluation metrics on a) freshwater test set and b) BPNS test set. X axis represents the number of additional annotated files using the active learning (AL) and random selection (RS) method for selecting these files. Shaded area represents the minimum and maximum, and the line represents the mean value.



When evaluating models MB, AL5, RS5 and MF on the freshwater test set, the performance is comparable to the metrics obtained in the BPNS test set (see Table 7.2). This proves that the model is robust across datasets and can be used on data from unseen locations. For the freshwater test set, AL5 also outperformed RS5 approach in all the metrics except TNP and FPP (time). It also outperformed MF in several metrics including precision and F1.

The detected percentage for both area and time in the freshwater test set is better than in the BPNS set, but worse when looking TNP and FPP (both time and area). This points out differences between the events distribution in time and frequency between the two test sets. Within each test set, TNP and FPP are best when computed for area, while detection percentage is best when computed for time.

Table 7.2 Average performance of the final models. MB stands for Model Base, AL for Active Learning, RS for Random Sampling and MF for Model Final in percentage. Area metrics are the ones computed considering both frequency and time. Time metrics are the ones considering only times. det area/time stands for percentage of detected area/time, TNP stands for True Negative Percentage, and FPP stands for False Positive Percentage. preci. stands for precision. Detection metrics are computed by counting overlapping boxes, with a iou threshold set to 0.1 to compute precision, recall and F1. Best result per metric and test set is marked in bold. All results are including predictions with a confidence of 0.1 or more. The minimum and maximum values of MB, RS and AL can be seen as the first and last points of the evolution curves on Figure 7.5.

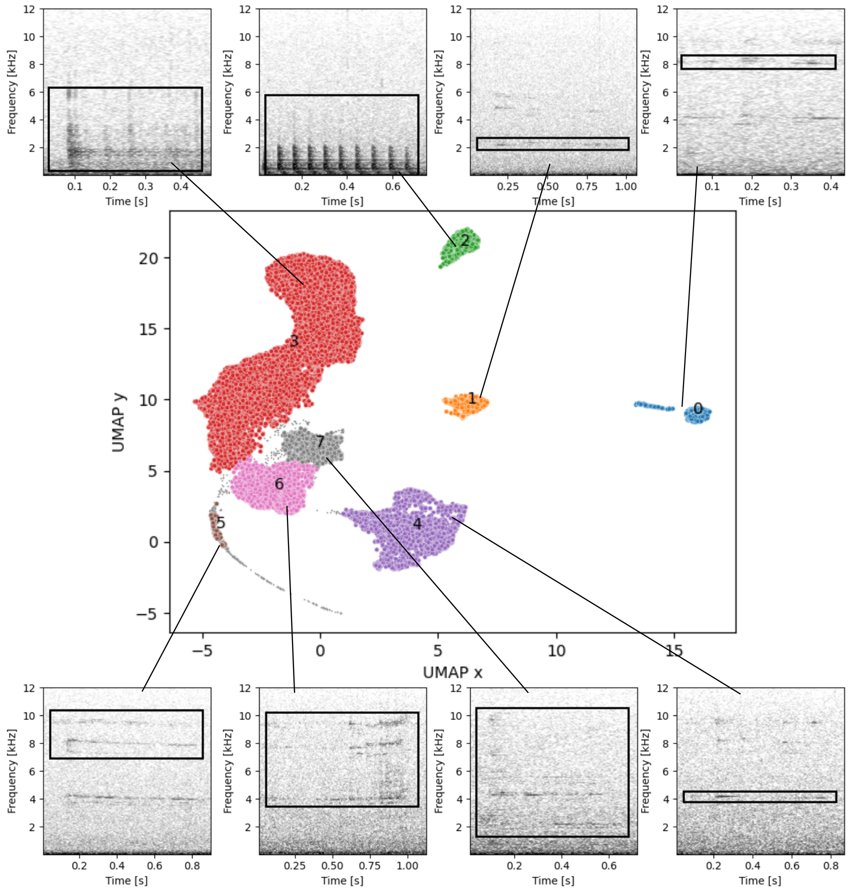
		Area			Time			Detections		
		det. area	TNP area	FPP area	det. time	TNP time	FPP time	preci.	recall	F1
BPNS	MB	52.4	99.98	0.02	60.15	89.21	10.79	42.58	**51.12**	46.38
BPNS	RS	**53.49**	99.99	0.01	**58.95**	93.38	6.62	56.67	47.75	50.33
BPNS	AL	49.12	**99.99**	**0.01**	58.29	**95.76**	**4.24**	**65.28**	50.09	**56.69**
Freshwater	MB	33.45	99.54	0.46	46.86	77.50	22.50	47.07	36.18	40.69
Freshwater	RS	50.39	99.58	0.42	58.27	**81.14**	**18.86**	59.53	44.03	49.16
Freshwater	AL	67.77	**99.71**	**0.29**	**81.87**	77.10	22.90	**61.57**	57.51	**59.47**
Freshwater	MF	**75.23**	99.62	0.38	58.98	76.65	23.35	57.96	**60.03**	58.98

7.3.2 Clustering and continuous data analysis

The final model detected a total of 197,793 events during the 10 days of the deployment (6 days of data, a total of 705 wav files). From all the obtained (merged) detections, 90.91% were detections shorter than 0.3 seconds. These were not converted to the embedding space and were classified as impulsive sounds.

For the rest of the detections, the UMAP 2D space applied to the BioLingual embedded feature space presented a clear cluster structure (see Figure 7.6). When applying HDBSCAN to the 2D space, 8 clusters were obtained (see Figure 7.6). All the clusters were manually revised as explained in Section 7.2.7, and assigned a label and a possible source. The output of these revision is listed in the Supplementary

Figure 7.6 The UMAP 2D reduction colored by obtained clusters and one spectrogram example for each cluster. For the spectrogram generation, number of FFT bins was 512, with an overlapping of 480 samples, Hann window. A black box has been added to show the frequency limits of each detection.



Material Table S7.2, where textual descriptions, mean frequency limits and duration information (10th and 90th percentile) are provided for each cluster.

The clusters were then grouped by source type, and the number of clusters and percentages of each source type were summarized (see Table 7.3). From the 9.09% selected for further clustering, the biological/pseudo-noise were the group with most detections (52.26%), followed by unknown (non-impulsive) sources (42.06%), and finally biological (3.11%).

From all the obtained clusters, only one could be identified as biological with certainty (cluster 2). This cluster was formed by sound events with a high repetition

Table 7.3 Summary of the classification of all the obtained clusters.

Possible Source	Number of clusters	Percentage of events [%]
Unknown	6	3.8
Unknown (impulsive)	1	90.91
Anthropogenic	0	0
Biological	1	0.28
Biological / Pseudo-noise	1	4.75
Mooring noise	0	0
Geophonic	0	0
Instrument noise	0	0
Noise (not clustered)	1	0.23

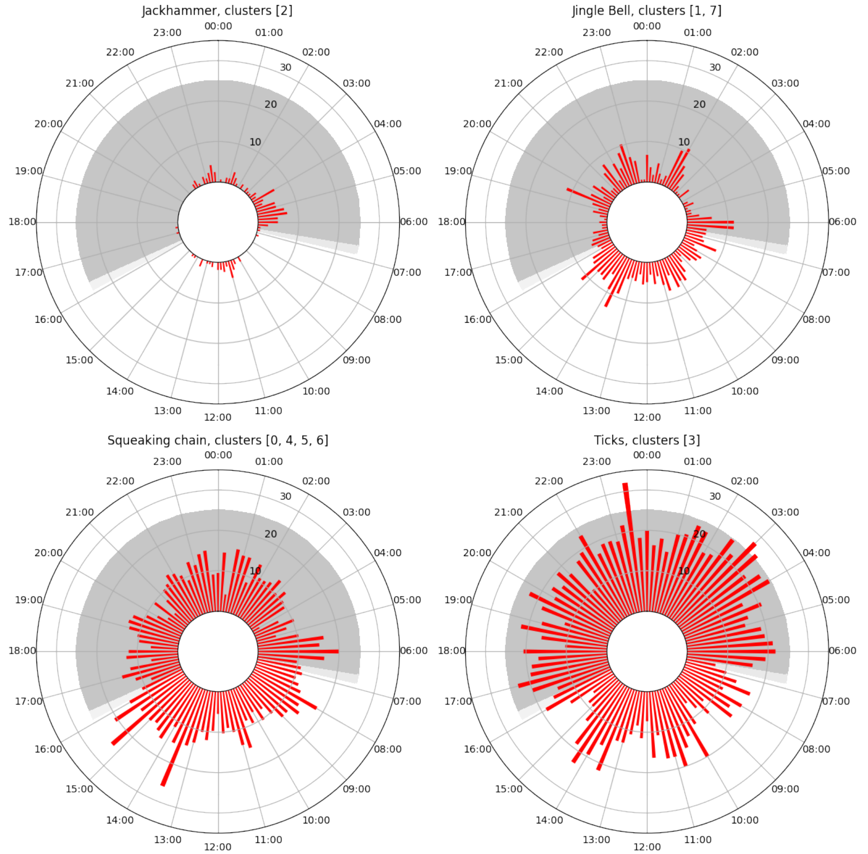
rate, and a frequency spanning from around 100 to 4,000 Hz (see Table S7.2 from Supplementary Material). We refer to it as ‘Jackhammer’. The number of pulses was not constant at each detected event, ranging from 2 to 70 pulses, but with a majority of them around 10-15 pulses. The repetition rate was around 14 pulses/second. When analyzing 15-minute occurrence of this cluster in time (see Figure 7.7), the number of detections seem to be higher at dawn. However, these sound events were mostly detected only on two different days, within a specific time frame lasting between 2 to 5 hours, while absent the rest of the days.

Two ”metallic” sounds were found throughout the analyzed recordings, classified in 6 different clusters due to its variability. From these clusters, two main groups could be extracted, with clusters 6 and 7 being the representation for each group, respectively. Group represented by cluster 7 was defined as a clear jingle-bell-like sound at different frequencies, from 1.5 to 8 kHz, with usually several harmonics and a duration of around one second. Cluster 1 was a single harmonic from this sound, sometimes selected within the full sound, and sometimes found alone probably because of propagation loss. The sound represented by cluster 1 and 7 was named ‘Jingle bell’. On the other hand, group represented by cluster 6 was labeled sounding as a ”squeaking chain” and it was present at higher frequencies, from 4 to 10 kHz. It often looks like a down sweep, and it has no impulsive component at the beginning. When complete (cluster 6), it also presents harmonics and also a duration of around one second. Clusters 0, 4, and 5 were labeled as harmonics from this sound, also sometimes selected within the full sound. This group was labeled as ‘Squeaking chain’.

Cluster 3 was a repetition of impulsive sounds, sounding like a wooden scratch.

Figure 7.7 Daily patterns of the number of detections of the selected classes every 15 minutes. Black lines represent sunset and sunrise.

Figure 7.8 Polar plot of the detection distribution per class depending on the hour of the day for the Grafton deployment. Radius represents the percentage of minutes during which at least one detection was present in the corresponding 15 minute bin. Dark share represents night, light shade represents the twilight and white represents the day.



It presented simultaneously a semi-constant tonal component at around 2 kHz and several impulsive sounds. These impulsive sounds were present both broadband or very narrow band. We named it ‘Ticks’.

When analyzing the temporal distribution of the clusters, no clear patterns could be seen. The polar plot in Figure 7.8 revealed that the ‘jackhammer’ happened mostly at dawn. ‘Jingle bell’ and ‘squeaking chain’ seemed to have similar patterns, with a slight increase during the day. ‘Ticks’ presented a higher density during the night than during the day, but there were also detections during the day.

7.4 Discussion

In this study we show a novel methodology to analyze underwater soundscapes in areas where very few sound sources have been previously described. The method helps to gain insight into the different (recurrent) sound sources in the soundscape, and allows for an automatic detection and categorization of sound events with limited human effort. With this methodology, soundscape analysis could provide meaningful insight even though the sources of the different sound types are not known.

The performance achieved by the object detector model on the test set using any of the three models (base model, random sampling, and active learning) was comparable to human performance. Leroy et al. (2018) [25] found that annotators agreed between 28 % and 86 %. Nguyen Hong Duc et al. (2021) [26] found that inter-annotator agreement varied largely, and all the annotators agreed only between 1 % and 40 % of the cases, depending on the call and the SNR. This is especially true in data scenarios where annotation is challenging, such as when high ambient noise levels mask the sound events of interest [25, 26]. Not knowing which are the sound events of interest adds an extra challenge and inconsistency. The concept of acoustical and visual saliency is subjective. Differentiating foreground events from background noise depends on the human analyst and the selected settings and goals during annotation, as there is no clear separation between foreground and background but rather a continuum of levels of masking. An example of this challenge is shown in Figure S7.1 of Supplementary Material, where model predictions and human annotations are diverging substantially, but the ground truth annotations are very subjective.

The overall obtained F1 values were not high, but TNP and FPP presented an overall good performance, both when looking at the area and the time metrics. This is partly due to the sparseness of the sound events, which makes TNP and FPP suited metrics to evaluate the performance in long-term data. Furthermore, the performance of the detection model was proven to be robust across locations and ecosystems, as it performed better in data from a location that was not used for any training, even from a complete different ecosystem such as freshwater. The fact that the obtained models (AL5, RS5 and MF) performed overall better in the freshwater test set than in the BPNS test set might be because the freshwater recordings are less noisy than those of the BPNS, which is known to be an extremely noisy environment [35]. However, even though the models performed better in the freshwater test set than in the BPNS regarding recall, F1 and detected percentage (area and time), they performed worse when looking at TNP and FPP. This could be due to differences regarding quantity, simultaneity, and the way the events were separated, between the two test sets.

The active learning approach led to better results overall than the random sampling approach. The files selected for active learning presented a higher acoustic complexity than the ones selected randomly. This supports the hypothesis that the metric used to select wav files points to more complex files. The model overall performs better (considering all metrics) when complex files are selected because it can learn how to solve complex situations in a more similar manner to a human annotator. However, the more complex the sounds, the higher is the challenge for the model to find all different sound events (hence the reduction in recall). It is necessary to note that the files selected by the active learning algorithm presenting a higher acoustic diversity might not necessarily represent higher biophonic activity. Furthermore, the files selected by the algorithm are based on the detections from the previous model, which means that totally new sounds could be completely missed (as they are not detected by the model at all). The active learning selection method thus does not assure the addition of unseen new and interesting sounds, but it has been proven to be more effective than the random selection of files. Therefore, if a model has to be re-trained and the available annotation time is limited, the active learning approach can deliver better results while investing less time on annotations. These findings are in line with other studies applying active learning to detect sounds on long-term recordings to extract ecological information [36, 37], pointing out that active learning is an interesting field to explore when human annotations from long-term recordings are necessary to train machine learning models.

The active learning approach also performed better than the Model Final in the freshwater dataset for several metrics. This is counter-intuitive, as MF was trained with a much larger dataset. Our hypothesis to explain this difference is that the AL model is presented with a much harder task, as it contains a high percentage of "complex" files. This forces the model to learn a better generalization compared to the MF, which has a lower percentage of "complex" files and therefore can focus on the easiest annotations, leading to a worse generalization.

The trained detection model can be applied to data from other locations, as it has been proven with the freshwater dataset. The model as it is, provides a performance similar to the human performance, so it can be used right away on other ecosystems. Yet, it might miss some sounds of interest, especially if applied on a different frequency range. A good approach for future fine tuning or re-training of this model would be to first create a base model trained with a balanced annotated dataset containing interesting sounds and a variety of environmental conditions. This way the model can learn from the beginning a good variety of shapes on the spectrogram.

In this study we prove that the BioLingual [19] model together with a UMAP 2D [31] reduction provided enough information to obtain clear and meaningful clusters. The manual revision of the obtained clusters indeed led to the conclusion that the clusters were acoustically meaningful and represented different sound types. Therefore it can be concluded that the BioLingual features contained enough information, and that the reduction to a 2D dimension using UMAP maintained the general density structure. The combination of UMAP reduction on a feature space together with HDBSCAN [34] algorithm applied on the reduced dimension was already successfully applied by Sainburg et al. (2020) [38], Thomas et al. (2022) [39] and Best et al. (2023) [33] to separate different biological sound events, so our results align with their proposal. However, in all these approaches the sound events were manually selected. The novelty of our proposed approach is that the whole process is automatized.

Regarding the application of the clustering to data from other environments, a new clustering algorithm on a new deployment would provide a different set of clusters, not necessarily comparable with the clusters obtained from previously analyzed deployments. To compare soundscapes between different deployments in the same regions, it is sometimes interesting to keep the existing clusters in order to track the changes in sound events. This should be possible if a large and representative enough dataset of detections is first clustered. Then it is possible to query the HDBSCAN model on small amounts of new data [34]. The python implementation of HDBSCAN allows for this, by holding a clustering fixed and then find out where in the condensed tree the new data would fall. The first representative dataset to cluster events can be manually annotated or can also be the outcome of the detection model ran on a selected set of files representing all the possible ecological conditions of interest, and multiple instances of all the expected sound types.

On the presented study we focused on a broadband frequency range, from 0 to 12 kHz. This decision was made so some cetacean sounds would also be included without risking not seeing low frequency sounds in the spectrogram. In the particular case of this study we focused on the entire frequency band equally. This means that when manually annotating a broadband frequency range, very narrow-band sounds can be easily missed, especially the ones in the lower frequency range. As expected then, the model might also miss these sounds as it has not been trained on those. Nonetheless, the same method could be applied to smaller frequency ranges, for example from 0 to 3 kHz if the interest would be focused on fish vocalizations [40]. The model should work regardless of the frequency range as long as the time and frequency resolution are enough to represent the sounds of interest. In future work it would be interesting to train the model with a logarithmic frequency scale to emphasize lower frequency sounds and compare the performance of both models.

In this particular case, due to the abundance of very short impulsive sounds (< 0.3 s), a first separation among detections was applied. This should not be necessary in areas where there is not one sound type dominating and generating this high imbalance, or where impulsive sounds (clicks) are not present. Because of the lack of knowledge and information regarding this impulsive sounds, we just cluster them according to their frequency limits. This is in line with the thesis by

Hardland (2017) [41], where clicks (impulsive sounds) heard in the UK waters were characterized and identified. Future research would be needed to assess the source of this impulsive sounds and their ecological significance, but nothing discards them coming from biological sources [42, 43, 44]. If the analyzed location contains click sequences and the model was not trained to recognize them as sequences but just as individual pulse units, they would not appear as a sequence cluster. However, a posterior analysis from all the impulsive sounds complying with the frequency limits of interest could be analyzed for temporal patterns and join the clicks into sequences.

When analyzing the obtained clusters in the deployment, the obtained sound types are similar to the ones mentioned in Calonge et al. (2024) [45], who did a clustering analysis from labeled manual annotations. This points out the robustness of the model, and highlights the reduction of manual input to reach similar conclusions. Only one sound was found fitting within the known description of fish sounds, as it is within the known vocalization frequency range of fish, and it is a repetitive set of impulse sounds [40, 46]. There have been similar fish sounds reported in literature from the family *Sciaenidae* [47], and an invasive species of this family has been documented in the North Sea [48]. However, these assumptions have to be taken with the most caution as ground truth has not been confirmed. The obtained cluster 'Ticks' could maybe come either from bio-abrasion of the hydrophone [49], invertebrates or fish clicking sounds [41]. Finally two different metallic sounds were found. These could originate from the mooring itself, as it is a steel mounting system (even though it has no moving parts), but could also be related to invertebrate sounds, such as the ones mentioned in Coquereau et al. (2016) [43]. The fact that these two sounds appeared in multiple clusters is because the sound did not present all the harmonics all the time, probably due to propagation loss [50] or sound production inconsistency. With the presented approach, bounding boxes from sound events can overlap as long as iou is less than 0.5, otherwise they are merged and considered the same detection. Therefore, harmonic sounds can potentially be selected in multiple boxes at the same time. This is advantageous because when these harmonics appear by themselves they are clustered together with the boxes that overlap with the full sound, so they can be traced back to their origin. However it can be disadvantageous because it can complicate the counting of sound occurrences.

The shown case study provides an example of the analysis that can be performed with the outcome of the presented model pipeline. This analysis can provide insight on the spatio-temporal patterns of certain sound types, which in the long term can be used to discover their source. With this methodology it is possible to already obtain ecological information at the same time that researchers discover the sources of sounds and gain insight on the soundscape. However, once a sound source is identified, considered of interest, and adequately characterized, other supervised techniques might be more efficient and provide a greater performance for sound event detection and posterior soundscape analysis and description [1, 51]. For this reason it is necessary to create databases of sounds where well-described unidentified sounds can also be added, so in the future they can be used as references [6].

In conclusion, the proposed method is a useful tool to discover unknown sounds in a new environment and can be used as a first analysis tool. Implementing this methodology in already available annotation or exploration software such as PAMGuard [52], Whombat [53] or RavenPro [24] can help addressing some of the challenges encountered when studying underwater soundscapes with little known sound sources. The obtained model is robust to different environments and can be applied directly to new data, even though for higher performance it would be recommended to re-train on a subset of this new data. The principal advantage of this model is that it is not based on previous assumptions of which sounds could be of interest, as all the possible events are detected and classified. Furthermore, it provides a framework for discovering the sound types while already gaining ecological insight of the soundscape. The proposed methodology helps in filling the gap in knowledge on sound types, which is currently major issue for using PAM for ecological assessment of the underwater environment [54, 9, 6].

7.5 Conclusions

This work provides a new methodology to automatically detect any potentially biological salient sound event in underwater long-term recordings, and cluster these detections using state-of-the-art machine learning techniques. To train the detection model, an active learning workflow is proposed to limit the necessary human effort. This approach does not depend on previous knowledge on present sound sources, and it does not build on assumptions on the sound events of interest, so it is especially interesting for unidentified sound events.

The method can be used to discoverer different sound types present in underwater recordings without having to manually annotate them. The detected events, once assigned to a sound type category can be analyzed for temporal and spatial patterns, which could, in the future, lead to identification of the source.

In future work, it would be interesting to assess if the obtained clusters of sound types occur at specific locations, moment or environmental situations. For this, we can assess if the obtained sound types happen more often when the background sound is categorized in one specific soundscape type, such as the ones obtained in Chapter 5. Linking background clusters with foreground clusters could provide further insight on co-occurrence of acoustical phenomena and on possible behavior and ecosystem use.

Acknowledgments

The project was realized as part of the Flemish contribution to LifeWatch. The authors thank the Simon Stevin RV crew for their help with the deployment and retrieval of the acoustic instruments. We would like to thank the maps team at VLIZ for the location map. The authors want to thank Julia Aubach, Svenja Wohle, Quentin Hamard, Zoe Rommelaere, Coline Bedoret, Kseniya Mulhern and Olayemi Saliu who helped with the annotation of the data.

Data Availability Statement

The full dataset from the BPNS can be found on IMIS, and is open and available upon request. https://marineinfo.org/id/dataset/78799
The recordings selected from the full BPNS dataset and used for training and evaluating the model are available in MDA (https://marinedataarchive.org) upon publication. https://doi.org/10.14284/667
The code used for the analysis of this paper can be found on github
https://github.com/lifewatch/sound-segregation-and-categorization

References

[1] D. Stowell. *Computational Bioacoustics with Deep Learning: A Review and Roadmap*. PeerJ, 10:e13152, March 2022.

[2] P. Rubbens, S. Brodie, T. Cordier, D. Destro Barcellos, P. Devos, J. A. Fernandes-Salvador, J. I. Fincham, A. Gomes, N. O. Handegard, K. Howell, C. Jamet, K. H. Kartveit, H. Moustahfid, C. Parcerisas, D. Politikos, R. Sauzède, M. Sokolova, L. Uusitalo, L. Van den Bulcke, A. T. M. van Helmond, J. T. Watson, H. Welch, O. Beltran-Perez, S. Chaffron, D. S. Greenberg, B. Kühn, R. Kiko, M. Lo, R. M. Lopes, K. O. Möller, W. Michaels, A. Pala, J.-B. Romagnan, P. Schuchert, V. Seydi, S. Villasante, K. Malde, and J.-O. Irisson. *Machine Learning in Marine Ecology: An Overview of Techniques and Applications*. ICES Journal of Marine Science, page fsad100, August 2023.

[3] E. Desiderà, P. Guidetti, P. Panzalis, A. Navone, C.-A. Valentini-Poirrier, P. Boissery, C. Gervaise, and L. D. Iorio. *Acoustic Fish Communities: Sound Diversity of Rocky Habitats Reflects Fish Species Diversity*. Marine Ecology Progress Series, 608:183–197, January 2019.

[4] M. Bolgan, C. Gervaise, L. Di Iorio, J. Lossent, P. Lejeune, X. Raick, and E. Parmentier. *Fish Biophony in a Mediterranean Submarine Canyon*. The Journal of the Acoustical Society of America, 147(4):2466–2477, April 2020.

[5] L. Di Iorio, M. Audax, J. Deter, F. Holon, J. Lossent, C. Gervaise, and P. Boissery. *Biogeography of Acoustic Biodiversity of NW Mediterranean Coralligenous Reefs*. Scientific Reports, 11(1):16991, August 2021.

[6] M. J. G. Parsons, T.-H. Lin, T. A. Mooney, C. Erbe, F. Juanes, M. Lammers, S. Li, S. Linke, A. Looby, S. L. Nedelec, I. Van Opzeeland, C. Radford, A. N. Rice, L. Sayigh, J. Stanley, E. Urban, and L. Di Iorio. *Sounding the Call for a Global Library of Underwater Biological Sounds*. Frontiers in Ecology and Evolution, 10, 2022.

[7] W. W. L. Au and M. C. Hastings. *Principles of Marine Bioacoustics*. Springer US, New York, NY, 2008.

[8] R. P. Schoeman, C. Erbe, G. Pavan, R. Righini, and J. A. Thomas. *Analysis of Soundscapes as an Ecological Tool*. In C. Erbe and J. A. Thomas, editors, Exploring Animal Behavior Through Sound: Volume 1: Methods, pages 217–267. Springer International Publishing, Cham, 2022.

[9] T. A. Mooney, L. Di Iorio, M. Lammers, T.-H. Lin, S. L. Nedelec, M. Parsons, C. Radford, E. Urban, and J. Stanley. *Listening Forward: Approaching*

Marine Biodiversity Assessments Using Acoustic Methods. Royal Society Open Science, 7(8):201287, 2020.

[10] M.-N. Havlik, M. Predragovic, and C. M. Duarte. *State of Play in Marine Soundscape Assessments*. Frontiers in Marine Science, 9, 2022.

[11] E. Staaterman, C. B. Paris, H. A. DeFerrari, D. A. Mann, A. N. Rice, and E. K. D'Alessandro. *Celestial Patterns in Marine Soundscapes*. Marine Ecology Progress Series, 508:17–32, August 2014.

[12] S. L. Nedelec, S. D. Simpson, M. Holderied, A. N. Radford, G. Lecellier, C. Radford, and D. Lecchini. *Soundscapes and Living Communities in Coral Reefs: Temporal and Spatial Variation*. Marine Ecology Progress Series, 524:125–135, March 2015.

[13] M. J. G. Parsons, C. P. Salgado-Kent, S. A. Marley, A. N. Gavrilov, and R. D. McCauley. *Characterizing Diversity and Variation in Fish Choruses in Darwin Harbour*. ICES Journal of Marine Science, 73(8):2058–2074, September 2016.

[14] C. C. Wall, R. A. Rountree, C. Pomerleau, and F. Juanes. *An Exploration for Deep-Sea Fish Sounds off Vancouver Island from the NEPTUNE Canada Ocean Observing System*. Deep Sea Research Part I: Oceanographic Research Papers, 83:57–64, January 2014.

[15] D. Stowell, D. Giannoulis, E. Benetos, M. Lagrange, and M. D. Plumbley. *Detection and Classification of Acoustic Scenes and Events*. IEEE Transactions on Multimedia, 17(10):1733–1746, October 2015.

[16] A. Looby, K. Cox, S. Bravo, R. Rountree, F. Juanes, L. K. Reynolds, and C. W. Martin. *A Quantitative Inventory of Global Soniferous Fish Diversity*. Reviews in Fish Biology and Fisheries, 32(2):581–595, June 2022.

[17] G. Jocher, A. Chaurasia, and J. Qiu. *YOLO by Ultralytics*, January 2023.

[18] C. Parcerisas, D. Botteldooren, P. Devos, and E. Debusschere. *Broadband Acoustic Network Dataset*, 2021.

[19] D. Robinson, A. Robinson, and L. Akrapongpisak. *Transferable Models for Bioacoustics with Human Language Supervision*, August 2023.

[20] T. A. C. Lamont, L. Chapuis, B. Williams, S. Dines, T. Gridley, G. Frainer, J. Fearey, P. B. Maulana, M. E. Prasetya, J. Jompa, D. J. Smith, and S. D. Simpson. *HydroMoth: Testing a Prototype Low-Cost Acoustic Recorder for Aquatic Environments*. Remote Sensing in Ecology and Conservation, 8(3):362–378, 2022.

[21] EMODnet Bathymetry Consortium. *EMODnet Digital Bathymetry (DTM 2018)*, 2018.

[22] GEBCO Compilation Group. *GEBCO_2022 Grid*, 2022.

[23] L. Minier, F. Bertucci, X. Raick, E. Gairin, H. Bischoff, V. Waqalevu, T. Maueau, V. Sturny, E. Blin, E. Parmentier, and D. Lecchini. *Characterization of the Different Sound Sources within the Soundscape of Coastline Reef Habitats (Bora Bora, French Polynesia)*. Estuarine, Coastal and Shelf Science, 294:108551, November 2023.

[24] K. Lisa Yang Center for Conservation Bioacoustics at the Cornell Lab of Ornithology. *Raven Pro: Interactive Sound Analysis Software (Version 1.6.4)*. The Cornell Lab of Ornithology, 2023.

[25] E. C. Leroy, K. Thomisch, J.-Y. Royer, O. Boebel, and I. Van Opzeeland. *On the Reliability of Acoustic Annotations and Automatic Detections of Antarctic Blue Whale Calls under Different Acoustic Conditions*. The Journal of the Acoustical Society of America, 144(2):740–754, August 2018.

[26] P. Nguyen Hong Duc, M. Torterotot, F. Samaran, P. R. White, O. Gérard, O. Adam, and D. Cazau. *Assessing Inter-Annotator Agreement from Collaborative Annotation Campaign in Marine Bioacoustics*. Ecological Informatics, 61:101185, March 2021.

[27] H. Bagherinezhad, M. Horton, M. Rastegari, and A. Farhadi. *Label Refinery: Improving ImageNet Classification through Label Progression*, May 2018.

[28] B. Rhodes. *Skyfield: High Precision Research-Grade Positions for Planets and Earth Satellites Generator*, July 2019.

[29] T.-Y. Lin, M. Maire, S. Belongie, L. Bourdev, R. Girshick, J. Hays, P. Perona, D. Ramanan, C. L. Zitnick, and P. Dollár. *Microsoft COCO: Common Objects in Context*, February 2015.

[30] M. Fennell, C. Beirne, and A. C. Burton. *Use of Object Detection in Camera Trap Image Identification: Assessing a Method to Rapidly and Accurately Classify Human and Animal Detections for Research and Application in Recreation Ecology*. Global Ecology and Conservation, 35:e02104, June 2022.

[31] L. McInnes, J. Healy, and J. Melville. *UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction*, September 2020.

[32] Y. F. Phillips, M. Towsey, and P. Roe. *Revealing the Ecological Content of Long-Duration Audio-Recordings of the Environment through Clustering and Visualisation*. PLOS ONE, 13(3):e0193345, March 2018.

[33] P. Best, S. Paris, H. Glotin, and R. Marxer. *Deep Audio Embeddings for Vocalisation Clustering*. PLOS ONE, 18(7):e0283396, July 2023.

[34] L. McInnes, J. Healy, and S. Astels. *Hdbscan: Hierarchical Density Based Clustering*. Journal of Open Source Software, 2(11):205, March 2017.

[35] C. Parcerisas, D. Botteldooren, P. Devos, Q. Hamard, and E. Debusschere. *Studying the Soundscape of Shallow and Heavy Used Marine Areas: Belgian Part of the North Sea*. In A. N. Popper, J. Sisneros, A. D. Hawkins, and F. Thomsen, editors, The Effects of Noise on Aquatic Life, pages 1–27. Springer International Publishing, Cham, 2023.

[36] L. H. Hilasaca, M. C. Ribeiro, and R. Minghim. *Visual Active Learning for Labeling: A Case for Soundscape Ecology Data*. Information, 12(7):265, July 2021.

[37] M. Kholghi, Y. Phillips, M. Towsey, L. Sitbon, and P. Roe. *Active Learning for Classifying Long-Duration Audio Recordings of the Environment*. Methods in Ecology and Evolution, 9(9):1948–1958, 2018.

[38] T. Sainburg, M. Thielk, and T. Q. Gentner. *Finding, Visualizing, and Quantifying Latent Structure across Diverse Animal Vocal Repertoires*. PLOS Computational Biology, 16(10):e1008228, October 2020.

[39] M. Thomas, F. H. Jensen, B. Averly, V. Demartsev, M. B. Manser, T. Sainburg, M. A. Roch, and A. Strandburg-Peshkin. *A Practical Guide for Generating Unsupervised, Spectrogram-Based Latent Space Representations of Animal Vocalizations*. Journal of Animal Ecology, 91(8):1567–1581, 2022.

[40] M. C. Amorim. *Diversity of Sound Production in Fish*. Communication in Fishes, 1:71–105, January 2006.

[41] E. J. Harland. *An Investigation of Underwater Click Sounds of Biological Origin in UK Shallow Waters*. PhD thesis, University of Southampton, UK, February 2017.

[42] B.-N. Kim, J. Hahn, B. Choi, and B.-C. Kim. *Acoustic Characteristics of Pure Snapping Shrimp Noise Measured under Laboratory Conditions*. In Proceedings of Symposium on Ultrasonic Electronics, volume 30, pages 167–168, 2009.

[43] L. Coquereau, J. Grall, L. Chauvaud, C. Gervaise, J. Clavier, A. Jolivet, and L. Di Iorio. *Sound Production and Associated Behaviours of Benthic Invertebrates from a Coastal Habitat in the North-East Atlantic*. Marine Biology, 163(5):127, May 2016.

[44] K. S. Cole, editor. *Reproduction and Sexuality in Marine Fishes: Patterns and Processes*. Univ. of California Press, Berkeley, 2010.

[45] A. Calonge, C. Parcerisas, E. Schall, and E. Debusschere. *Revised Clusters of Annotated Unknown Sounds in the Belgian Part of the North Sea*. Frontiers in Remote Sensing, 5, June 2024.

[46] R. Carriço, M. A. Silva, G. M. Menezes, P. J. Fonseca, and M. C. P. Amorim. *Characterization of the Acoustic Community of Vocal Fishes in the Azores*. PeerJ, 7:e7772, November 2019.

[47] M. C. P. Amorim, J. A. Wanjala, M. Vieira, M. Bolgan, M. A. Connaughton, B. P. Pereira, P. J. Fonseca, and F. Ribeiro. *Detection of Invasive Fish Species with Passive Acoustics: Discriminating between Native and Non-Indigenous Sciaenids*. Marine Environmental Research, 188:106017, June 2023.

[48] P. Morais, I. Cerveira, and M. A. Teodósio. *An Update on the Invasion of Weakfish Cynoscion Regalis (Bloch & Schneider, 1801) (Actinopterygii: Sciaenidae) into Europe*. Diversity, 9(4):47, December 2017.

[49] J. P. Ryan, J. E. Joseph, T. Margolina, L. T. Hatch, A. Azzara, A. Reyes, B. L. Southall, A. DeVogelaere, L. E. Peavey Reeves, Y. Zhang, D. E. Cline, B. Jones, P. McGill, S. Baumann-Pickering, and A. K. Stimpert. *Reduction of Low-Frequency Vessel Noise in Monterey Bay National Marine Sanctuary During the COVID-19 Pandemic*. Frontiers in Marine Science, 8, 2021.

[50] T. G. Forrest. *From Sender to Receiver: Propagation and Environmental Effects on Acoustic Signals*. American Zoologist, 34(6):644–654, December 1994.

[51] V. R. Barroso, F. C. Xavier, and C. E. L. Ferreira. *Applications of Machine Learning to Identify and Characterize the Sounds Produced by Fish*. ICES Journal of Marine Science, 80(7):1854–1867, September 2023.

[52] D. Gillespie, D. K. Mellinger, J. Gordon, D. McLaren, P. Redmond, R. McHugh, P. Trinder, X.-Y. Deng, and A. Thode. *PAMGUARD: Semiautomated, Open Source Software for Real-time Acoustic Detection and Localization of Cetaceans*. The Journal of the Acoustical Society of America, 125(4):2547–2547, April 2009.

[53] S. M. Balvanera, O. Mac Aodha, M. J. Weldy, H. Pringle, E. Browning, and K. E. Jones. *Whombat: An Open-Source Annotation Tool for Machine Learning Development in Bioacoustics*, November 2023.

[54] R. A. Rountree, M. Bolgan, and F. Juanes. *How Can We Understand Freshwater Soundscapes Without Fish Sound Descriptions?* Fisheries, 44(3):137–143, 2019.

8

Revised clusters of annotated unknown sounds

This chapter is a reproduction of the publication:

- *Calonge, A., Parcerisas, C. Schall, E., and Debusschere. E.. Revised clusters of annotated unknown sounds in the Belgian part of the North Sea. Frontiers in Remote Sensing, 2024.* https://doi.org/10.3389/frsen.2024.1384562

As a second author in this paper, I significantly contributed to the conceptualization of the methodology and conducted all feature extraction processes, including training the models. Additionally, I supervised and motivated the creation of the annotations dataset. I actively participated in the writing of all sections of the paper, providing substantial input and revisions. All co-authors have given explicit consent to incorporate this paper as a chapter of this thesis.

For the audio files from the supplementary material, please refer to the online version of this manuscript.

Overview

This chapter explores further strategies to categorize unknown sounds (as done in Chapter 7), but in this case (limited) human labels are available. However, labels from unidentified sounds are highly inconsistent, due to the fact that these sounds have not been properly characterized in literature and therefore it is difficult to decide whether two sounds are from the same source. Therefore, a strategy is necessary to decide which classes are meaningful and clusterable. Here, we describe the steps and considerations to take when clustering acoustic events, and how to evaluate the outcome if human labels are available. We find out that two factors largely influenced the formation of clusters: (1) obtaining a relevant feature representation of the annotated dataset and (2) tuning the hyper-parameters of the chosen algorithms. This chapter is thus intended to build further knowledge on the research question:

Is it possible to disentangle different biological sound events from long-term recordings, and can machine learning assist in detecting and classifying such events?

8.1 Introduction

Sounds in the environment can convey ecologically relevant information and have been used to investigate animal diversity, abundance, behavior and population dynamics [1, 2]. Especially in the marine environment where sound travels faster and further compared to in air, underwater sound is a key component in the life of marine fauna. Multitudes of animals including mammals, fish and invertebrates produce and listen to sounds linked to communication, foraging, navigation, reproduction and social and behavioral interactions [3, 4]. Marine animals also have a widely varying hearing capacity, ranging from lower frequencies ($<$ 5 kHz) in invertebrates, fish and reptiles, to higher frequencies (up to 200 kHz) in cetaceans [5, 6]. Sounds serve as signals that allow these animals to relate to their environment, making the changing ocean soundscape of the Anthropocene an added stressor to life underwater. Adverse effects in the physiology and behavior of various marine animals were reported due to noise from anthropogenic activities such as vessel traffic, active sonar, acoustic deterrent devices, construction and seismic surveys.

Continuous monitoring of ocean soundscapes using passive acoustics has led to a wealth of underwater recordings containing vocalizations of marine mammals [7], feeding of sea urchins [8], stridulations of crustaceans and fish [9], bivalve movements [10] and fish sounds that can, in some cases, form choruses [11, 12], along with numerous unidentified (biological) sounds. Several sounds have been validated and associated with almost all marine mammals, fewer than a hundred aquatic invertebrate species and about a thousand fish species, which has led to the discovery of the soniferous behavior of more species each year [13, 14]. Simultaneously, passive acoustics has been used to assess biodiversity, ecological states and corresponding environmental changes, encompassing the more recent fields of Soundscape Ecology or Ecoacoustics [15, 16].

Biological underwater sound libraries already exist in the web, such as Fish-Sounds (https://fishsounds.net), FishBase (https://www.fishbase.se), Watkins Marine Mammal Sound Database (https://whoicf2.whoi.edu/science/B/whalesounds/index.cfm) and the British Library Sound Archive (https://sounds.bl.uk/Environment), to facilitate working with the growing number of acoustic recordings. Recently, a call for a Global Library of Underwater Biological Sounds (GLUBS) was published by Parsons et al. (2022) [13], to allow a better integrated manner of sharing and confirming underwater biological sounds. As manual annotation becomes an almost unattainable task, especially if focal sounds are poorly known, the growing wealth of acoustic data requires new methods of machine learning (ML) and unsupervised classification algorithms [17, 18]. Automatic detection of different acoustic signals, characterized by distinct features such as frequency, amplitude, duration and repetition rate, lies on the premise that would lead to an efficient and objective identification of species and animal behaviors based on specific vocal repertoires [16].

The recordings from the dataset explained in Section 2.2, were listened to and annotated for unknown, acoustically salient sounds of any target event. No discrimination was done depending on the source since this is uncertain. Sounds of known origin, which were clearly not biological, were excluded from this study. No pre-defined sound classification scheme nor strategy existed. Moreover, human annotations are inherently inconsistent, varying between analysts and between separate periods of annotation due to annotator personality or acoustic event type and its SNR [19, 20]. Therefore, their reliability is often questioned, especially when used to evaluate or train models [21]. This leads to a need for subsequent clustering and revision of the annotations, which highlights the importance of iterative refinement in data analysis during the process of annotation.

Therefore, in the present work, we introduce these annotated unknown sounds with a focus on those that are recurring and potentially of biological origin, and propose steps to derive meaningful clusters from these annotations. We discuss the following related questions: (1) how can we (automatically) identify which of the annotations are from the same source? (2) how can we derive meaningful clusters from annotated unknown sounds in the BPNS? and (3) which of the obtained clusters represent recurring sounds that are likely biological?

8.2 Methods

8.2.1 Data Selection and Annotation

The dataset used was the one presented in Section 2.2 (see Figure 2.2, in Chapter 2 for locations).

Raw audio files used for annotation were manually chosen based on (subjective) recording quality and possible presence of acoustically salient elements within the files, depending on environmental conditions such as period of the year, moment of the day, location, and previously identified sound events. These annotations were part of an initial data exploration, and they were not annotated following a defined strategy. Only some sections of the long-term recordings deemed to contain acoustically salient elements were annotated, with durations ranging from several minutes to several hours per sections. Annotated samples were from four of the seven present stations, namely Belwind, Birkenfels, Buitenratel and Grafton (Figure 2.2, in Chapter 2). The total duration of annotated samples per station is listed in Supplementary Table S8.1. All files were annotated using Raven Pro version 1.6.52 [22], and the settings used during annotation are listed in Supplementary Table S8.2.

Audio events considered to be target events (acoustically salient) were meticulously identified and labeled by drawing boxes around each identified signal. All sounds were manually labeled, both known and unknown. This means that in addition to sounds possibly originating from marine organisms, other sounds tagged included anthropogenic and geophonic sounds.

Label tag names were manually cross-checked with tags available in underwater sound repositories, such as FishSounds (fishsounds.net) and Dosits (dosits.org). Sounds with similar acoustic characteristics to the descriptions found online were named accordingly. If a sound of interest could not be related to a sound from one of these online platforms, another tag name was chosen based on auditory characteristics exhibited by each sound. Sounds with similar shapes within the spectrogram, auditory characteristics, frequency range and duration were assigned the same tag name.

The absence of prior knowledge about the present biological sound sources posed a significant challenge, even when cross-checking with existing libraries. In response, rather than focusing on a time-intensive process of meticulously evaluating and classifying each sound type, an alternative approach was adopted where subjective categories were assigned to any encountered sound, irrespective of repetition or a pre-defined classification scheme, followed by clustering and revision of the identified sounds (see Section 8.2.3).

8.2.2 Automatic feature extraction

We decided to use automatic feature extraction and following statistical clustering of these features to group and describe unknown sounds. As it is not clear which are the best acoustic features to describe and differentiate unknown underwater sounds, we decided to use available published state-of-the-art deep learning algorithms pre-trained and/or tested in bioacoustics data, containing (at least, partly) underwater sounds. Two different options were considered, namely the Animal Vocalization

Encoder based on Self-Supervision (AVES; [23]) and a convolutional autoencoder network (CAE; [24]) to obtain acoustic features. Since the autoencoder approach is unsupervised, we trained it on our own data (for training details see Supplementary Table S8.3). AVES extracts the features directly from the waveform, which has the advantage that no parameters must be chosen manually to create a representative spectrogram. Conversely, CAE uses spectrograms as an input, and all the snippets need to be cut (or zero-padded) to a certain length before generating the spectrogram images. Both models were developed with the intention of being robust across datasets, by evaluating them on datasets which were not used to train the model. AVES was tested on data from birds, terrestrial mammals, marine mammals, insects (mosquitos) and amphibians (frogs), and CAE on data from different birds and marine mammals. Due to their proven generalizability, they were considered appropriate for this study.

Before the extraction of features, we filtered all manual annotations to have a minimum duration of 0.0625 s, a maximum duration of 10 s, a maximum low frequency of 24,000 Hz, and a minimum NIST-Quick SNR of 10 dB. For the annotations whose maximum high frequency exceeded 24,000 Hz the high frequency was adapted to 24,000 Hz. This assured that the characteristics of the remaining sounds complied with the requirements of the two feature extraction algorithms and assured the exclusion of false annotations. All sound files were down-sampled to 48,000 Hz before feature extraction to assure comparability in extracted features.

For each of the models, two different strategies were tested, leading to four different feature sets.

For the AVES-bio-base model, we first extracted snippets from the raw audio files using the start and end time of each annotation. These snippets were then band-pass filtered to the frequency limits of each annotation using a Butterworth filter of order 4 from the scipy Python package [25]. The filtered snippet was then converted to audio representations using the AVES-bio model (1) subjected to mean-pooling (AVES-mean) or (2) subjected to max-pooling (AVES-max) to derive a 768-feature long vector per sound event for succeeding clustering analysis.

The CAE from Best et al. (2023) [24] is applied to spectrogram representations of the sounds instead of raw waveforms. Therefore, Mel spectrograms were created from 3s-windows around the center time of an annotation with an NFFT value of 2048 and 128 Mel filterbanks and passed through the CAE with the bottleneck set to 256, deriving vectors of the same number (CAE-original). This was done following the procedure described in Best et al. (2023) [24]. As a modification to this to better represent the large variability in duration and bandwidth that we observed among the annotated unknown sounds in our data, we also trained the CAE network with cropped spectrograms to their start and end times and low and high frequency limits (CAE-crops). For this, spectrograms were created for the actual duration of the sound event with an NFFT value chosen to deliver at least 128 frequency bins between the minimum and the maximum frequency, and the window overlap was set to deliver at least 128 time slots. The resulting spectrogram matrix was then cropped to the actual frequency limits of the sound event. To achieve compliance with the input format to the CAE, all resulting spectrogram images were then resized to 128x128 bins[1] before they were also passed through the CAE with the bottleneck set to the default number of 256, based on the study of Best et al. (2023) on varying bottleneck sizes. This derived vectors of 256 features at the end. For both CAE approaches (CAE-original and CAE-crops), the audio snippets were filtered to the frequency limits of the annotations using the same filtering strategy as for AVES, a band-pass Butterworth filter of order 4 from the scipy Python package [25].

To all four different types of feature vectors, we added four additional features which were directly extracted from the Raven selection tables, namely low frequency, high frequency, bandwidth and duration.

8.2.3 Statistical clustering

Feature selection is often conducted in large and high-dimensional datasets prior to applying clustering algorithms to get a subset of features which will best discriminate the resulting clusters [27]. This was done because most clustering algorithms do not perform well in high dimensional space, a problem known as "the curse of dimensionality". First, using the scikit-learn Python library [28], the feature sets were centered to the mean and scaled to unit variance. Then, sparse Principal Component Analysis (sPCA) was applied to reduce the four feature sets to the most discriminant features. sPCA forms principal components with sparse loadings—each principal component (PC) resulting to a subset of principal variables, in contrast to ordinary Principal Component Analysis (PCA) wherein each PC is a linear combination of all original variables [29]. sPCA was chosen instead of a non-linear dimension reduction technique so the obtained features would be a subset of the original features. This has the advantage that the dimension reduction does not need to be re-computed if new data are added. Several possible reduction sizes were considered and evaluated. The four additional features—low frequency, high frequency, bandwidth and duration, were retained in addition to the sPCA-selected features.

Upon choosing a clustering algorithm that would give meaningful clusters from our annotations, two aspects of the dataset were of concern due to the manner of annotation done in this study: (1) the variation among cluster densities as the selection of recording files to annotate was not done in a standardized manner, and (2) the presence of noise, or falsely annotated samples in the datasets, due

[1] Resized by interpolation using the torchvision [26] resize function with default parameters, which uses Bilinear interpolation with antialiasing.

to the inherent nature of annotations on underwater sounds, especially when the source is unknown. Therefore, an unsupervised density-based clustering algorithm, HDBSCAN [30], based on hierarchical density estimates was chosen. HDBSCAN partitions the samples according to the most significant clusters with varying density thresholds, excluding samples that are identified as noise by the algorithm itself [31]. The clustering was applied to the sPCA-reduced CAE (CAE-original & CAE-crops) and AVES feature sets (AVES-mean & AVES-max; Figure 8.1).

We coded a grid search function to adjust the hyperparameters of sPCA and HDBSCAN. We searched the hyperparameter space for the most optimized combination of parameter values based on two selected clustering evaluation measures. The homogeneity score from the scikit-learn Python library [28] based on annotations, and the density-based clustering validation (DBCV) score from the DBCV library [32], based on the quality of clusters and not the annotations. Both scores range from 0 to 1, with 1 representing a perfect score. The homogeneity score compares the similarity of original annotations with the predicted clusters, wherein a score of 1 satisfies homogeneity of all predicted clusters [33]. The DBCV score is an index proposed for density-based clusters which are not necessarily spherical. This score is based on the density of samples in a cluster, and the within- and between-cluster distances [34].

We inspected which parameters had a drastic effect on the resulting clusters and must be adjusted, prior to conducting grid search. For sPCA, only the parameter alpha, which controls the sparseness of components, was adjusted. The higher the alpha, the sparser the components, resulting to a lower number of 'relevant' features. We set different values of alpha (Table S8.4, Supplementary) according to the range of features that formed reasonable/acceptable clusters during the data exploration phase of the study. For HDBSCAN, a grid of values based on three parameters–the minimum cluster size (5, 8, 10 and 12), the minimum samples (3, 4 and 5), and epsilon (0.2, 0.5, 0.8)–were specified. While the minimum cluster size specifies the smallest number of samples to form a cluster, the minimum samples parameter is the number of neighboring points to be considered a dense region, therefore restricting the formation of clusters to the denser areas and classifying more samples as noise. The epsilon value is a threshold by which a cluster is split into smaller denser clusters [30].

For selecting the best clustering result from the grid search function, criteria had to be defined as high scores do not directly translate to high clustering performances. Performance is also based on the number of samples classified as noise and the number of resulting clusters. The following criteria were therefore set: (1) the percentage of samples clustered is $> 15\%$; that is, only a maximum of 85% of samples classified as noise by HDBSCAN was acceptable, (2) the number of clusters is $\leq$ the original number of annotation classes, and (3) with the highest average of homogeneity and DBCV scores. The percentage allowed to be classified as noise might seem high, but given the high disagreement between annotators, who often agree less than 50 % of the time when annotating known, stereotyped calls [19, 20], it was deemed necessary to allow for a high percentage of the dataset to not be considered part of any group. For visual comparative analysis, each grid search clustering result was also embedded into Euclidean space using a uniform manifold approximation and projection (UMAP; [35]) with a number of neighbors equal to 15 and a minimum distance of 0.2.

The obtained resulting scores were tested for normality using the Shapiro-Wilk test [36] and homogeneity of variance using Levene's test [37]. The results showed that data were non-parametric, therefore significant differences in the homogeneity and DBCV scores among the 4 feature sets were tested using the non-parametric Kruskal-Wallis test [38]. This test is the non-parametric equivalent to one-way analysis of variance (ANOVA). Pairwise comparisons were performed using a Wilcoxon rank sum test [39] as a post-hoc. To test the association of the parameters with the two scoring metrics, the parameters were fitted in a generalized linear model (GLM) with a Gamma distribution using the 'stats' R package. Finally, predicted clusters from the best grid search result were then revised by a bioacoustics expert (J.A.) and one representative sound was chosen per cluster, which had a good SNR and the highest resemblance to the other sounds within the same cluster. Intra- and inter-cluster variation were assessed using the 'clv' R package (Nieweglowski, 2023) and visualized using the 'qgraph' R package (Epskamp et al., 2023). Intra-cluster distance was calculated as the distance between the two furthest points within each cluster, while inter-cluster distances was calculated for each possible pair of cluster as the average distance between all samples of two different clusters.

An overview of the methodology from feature extraction up to evaluation of clusters is shown in Figure 8.1. Feature extraction and unsupervised clustering were performed using Python version 3.11.5 [40], while statistical tests were performed using R version 4.3.1 [41], with scripts made available on the GitHub repository: `https://github.com/lifewatch/unknown_underwater_sounds`.

Figure 8.1 Overview of the methodology conducted in this study from two different audio representations up to cluster evaluation measures. After feature extraction using the Animal Vocalization Encoder based on Self-Supervision (AVES) and a convolutional autoencoder network (CAE), the dimensionality of the datasets was reduced by selecting relevant features using sparse Principal Component Analysis (sPCA). Different values of the hyper-parameters of sPCA and HDBSCAN were tested using a coded grid search function. Obtained clusters were evaluated using its homogeneity score and density-based clustering validation (DBCV) score.

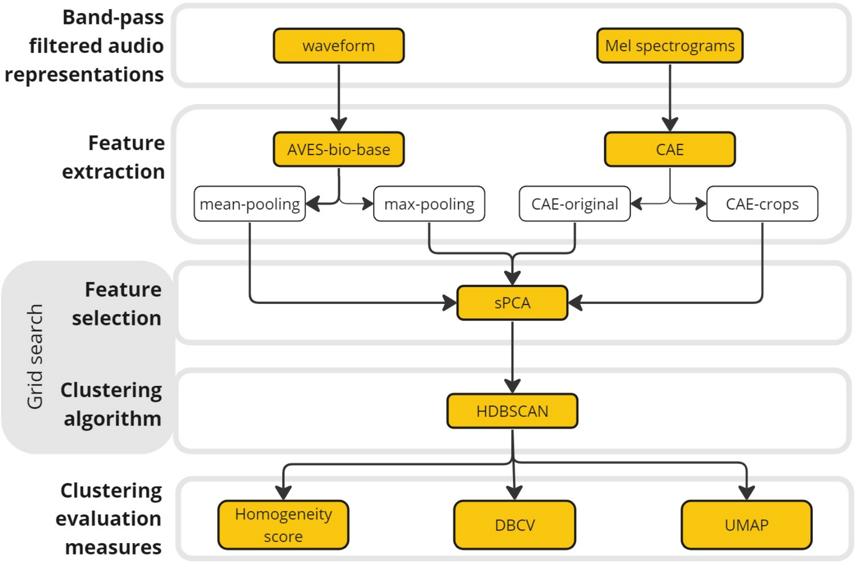
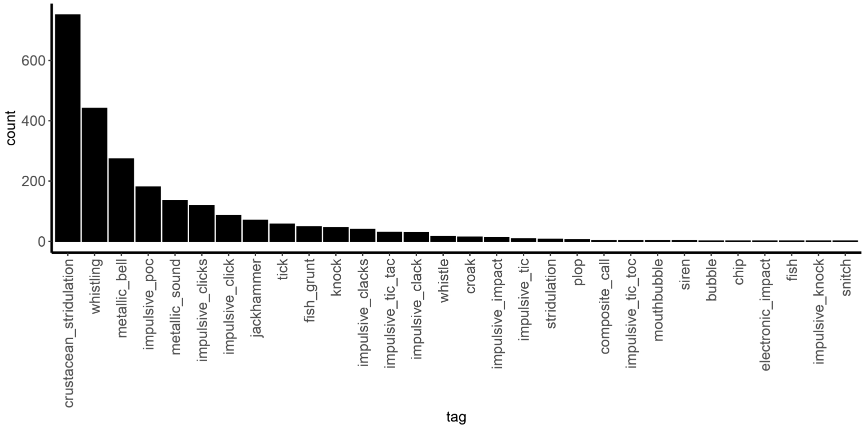


Figure 8.2 Annotated samples with corresponding tags in the Belgian part of the North Sea (BPNS), which excluded audio events related to boat operations, water movements, deployment operations, electronic noises and interference.



8.3 Results

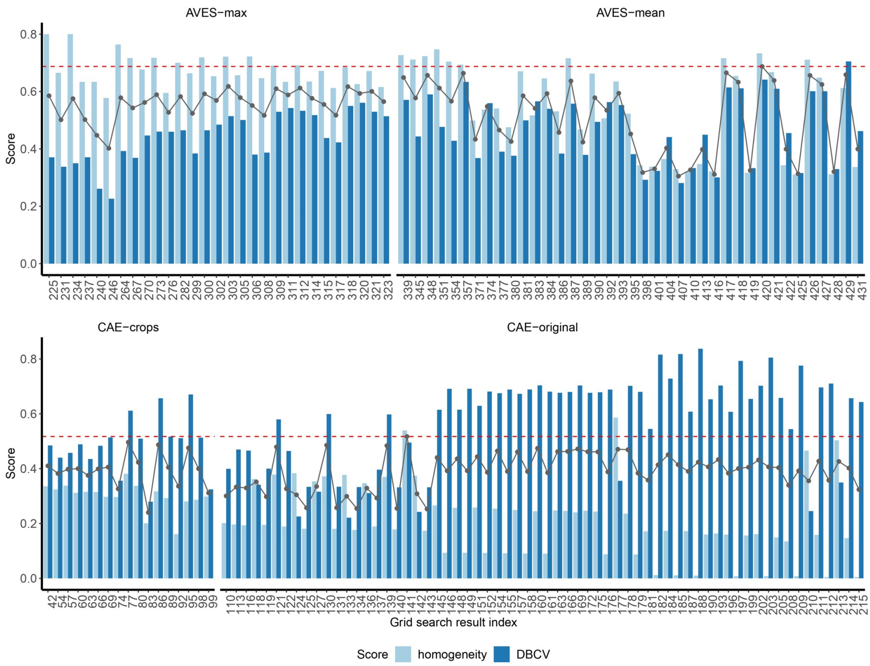
From all the selected raw audio files from the LifeWatch Broadband Acoustic Network in the BPNS, there were 2,874 target sounds of interest, annotated with 30 different tags (Figure 8.2). From all the annotations, those whose source could be identified and were not of biological origin were excluded from the analysis. These included boat noises, out of water sounds, water movements, deployment sounds, electronic noises and interferences. Acoustic features extracted through AVES and CAE were each reduced through sPCA. Different values of alpha were embedded in the grid search giving a similar range (15 - 31) of principal features for each dataset (Supplementary Table S8.4). The different values of sPCA alpha in combination with the different HDBSCAN clustering parameters (epsilon, minimum cluster size and minimum samples) gave a total of 431 grid search results.

Of the total 431 grid search results, only 238 results met the criteria set in this study with percentage of samples clustered $> 15\%$ (> 431 annotated samples) and the number of clusters less than or equal to the number of original annotation tags (= 30). From these 238 grid search results, where the same parameters regardless of the epsilon value gave the same homogeneity and DBCV scores, we only kept the grid search result with the highest epsilon value—further narrowing down the 238 grid search results to 149. Within these 149 grid search results, homogeneity scores ranged from 0.005 to 0.800, DBCV scores from 0.221 to 0.837 and the average of the two scores from 0.240 to 0.687 (Figure 8.3). Seven (CAE-crops & CAE-original) and four (CAE-crops) grid search results had homogeneity and DBCV scores considered outliers, respectively. The means and standard deviations of the scores from the grid search for each feature set are detailed in Supplementary Table S8.5.

Homogeneity and DBCV scores were significantly different among the feature sets (Figure 8.4; Kruskal-Wallis test, p = 2.2-16 [homogeneity], p = 0.0001 [DBCV]). Pairwise comparisons using Wilcoxon rank sum test showed that homogeneity scores among the four feature sets were significantly different from each other ($p < 0.05$). Likewise, the same test showed that DBCV scores were significantly different between the AVES feature sets and CAE-original (all $p < 0.05$). While AVES-max and AVES-mean feature sets had higher homogeneity scores, DBCV scores were higher for CAE-original. Finally, fitting all parameters in a GLM (details in Supplementary Table S8.6 and Supplementary Figure S8.1), the number of features was significantly associated to DBCV (p = 5.960-5), and the minimum cluster size to homogeneity scores (p = 9.864-5).

Among the AVES feature sets, grid search result index # 420 (AVES-mean) had the highest average of homogeneity and DBCV scores (= 0.687), with the number of features reduced to 15. There were 635 samples (22% of the total samples) that were clustered with a minimum cluster size of 10. Among the CAE feature

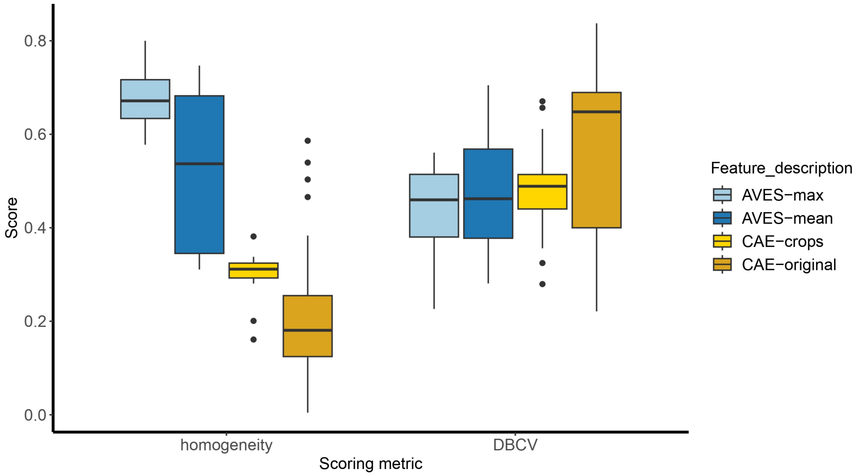
Figure 8.3 Homogeneity and density-based clustering validation (DBCV) scores of the 149 grid search results (indicated by the grid search result indices on the x-axis) grouped by feature set. The black dotted lines indicate the average of the two scores, while the red dashed line indicates the highest average score among the grid search results within AVES and CAE feature sets.



sets, grid search result index # 141 (CAE-original) ranked with the highest average score (= 0.517), with features reduced to 31. There were 1279 samples (45% of the total samples) that were clustered, with a minimum cluster size of 12. The UMAP embeddings of grid search results # 420 and # 141 (best of each approach) show a considerable separation between most of the clusters (Figure 8.5), with bigger clusters more evident from the UMAP embedding of grid search result index # 141 (Figure 8.5B).

The best grid search result index # 420, with a homogeneity score of 0.733 and DBCV score of 0.641, yielded 26 clusters (Figure 8.6). Descriptions of the resulting sound categories are summarized in Table 8.1. Of the 26 clusters, 7 were completely homogeneous (clusters 3, 4, 7, 21, 22, 23 and 24; Figure 8.6) — that is, samples within each of those clusters were labeled with the same tags during annotation. Two additional clusters (15 and 16) were assessed as completely homogeneous after revision. Of the total 9 completely homogeneous clusters, six were represented by 'Whistling', two of 'Tick' and one of 'Metallic Bell'. Two clusters (10 & 20)

Figure 8.4 Boxplots of homogeneity and density-based clustering validation (DBCV) scores for the 4 feature sets (AVES-max, AVES-mean, CAE-crops and CAE-original). Homogeneity scores were significantly different between each feature set (all $p < 0.05$). DBCV scores were significantly different between the AVES feature sets and CAE-original ($p = 0.001$



had less than 50% homogeneity: cluster 10 was composed of tags 'Crustacean Stridulation', 'Impulsive Clack', 'Jackhammer' and 'Fish Grunt', while cluster 20 was composed of 'Impulsive Poc', 'Crustacean Stridulation', 'Fish Grunt', 'Knock' and 'Plop'. Multiple clusters represented by the same sound, such as clusters represented by 'Whistling', 'Impulsive Poc', 'Impulsive Click' and 'Tick', had lower inter-cluster distances, but also relatively high intra-cluster distances (Figure 8.8). Additionally, a clear separation between subgroups of clusters under 'Whistling' and 'Crustacean Stridulation' can be observed. Separation of clusters 3-4 and 21-24, all represented by 'Whistling', and clusters 1, 10, 5 and 6, all represented by 'Crustacean Stridulation', indicates variation in acoustic representations within the same classification of sound.

Figure 8.5 Uniform Manifold Approximation and Projection (UMAP) of the 26 predicted clusters from grid search result # 420 (AVES-mean; A) and the 29 predicted clusters from grid search result # 141 (CAE-original; B), which had the highest average of homogeneity and DBCV scores among the AVES and CAE feature sets, respectively

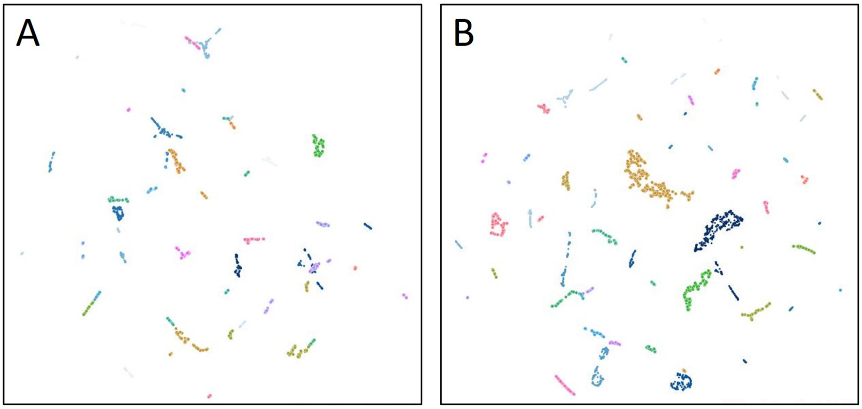


Figure 8.6 Agreement of annotations with predicted clusters of grid search result # 420. The grids are shaded by the percentage of occurrence of a tag within a predicted cluster.

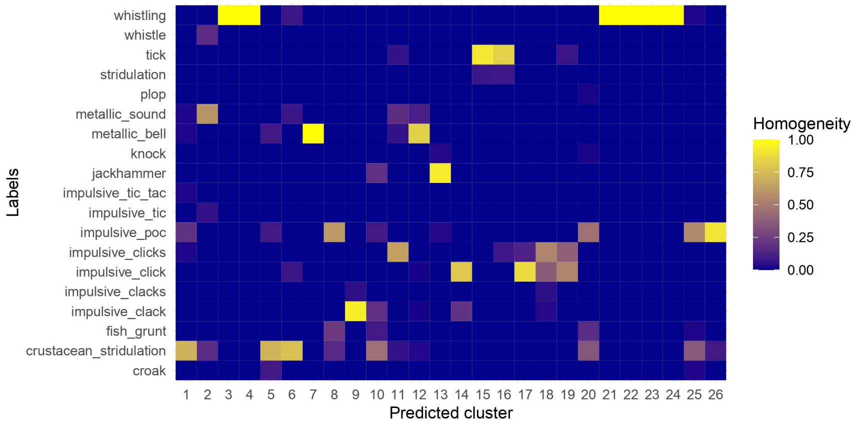


Figure 8.7 Spectrograms of revised clusters from the best grid search result which had the highest average of homogeneity and DBCV scores. Each spectrogram is labeled by the cluster # and the representative sound tag which had the highest resemblance to the other sounds from the same cluster.

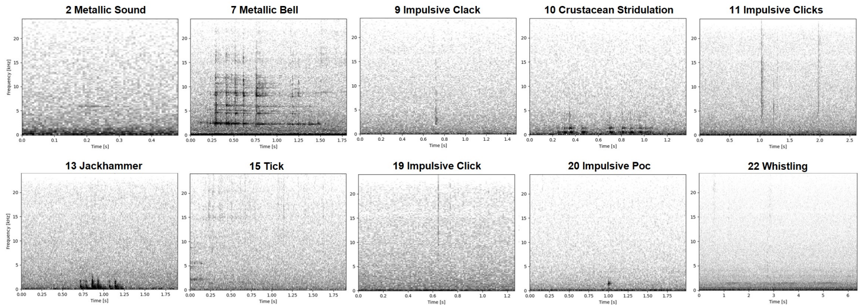


Figure 8.8 Evaluation of intra- and inter-cluster distances between each cluster. Clusters are separated by inter-cluster distance. Line thickness indicates similarity between each cluster. Clusters are numbered according to the cluster number and colored by the representative sound.

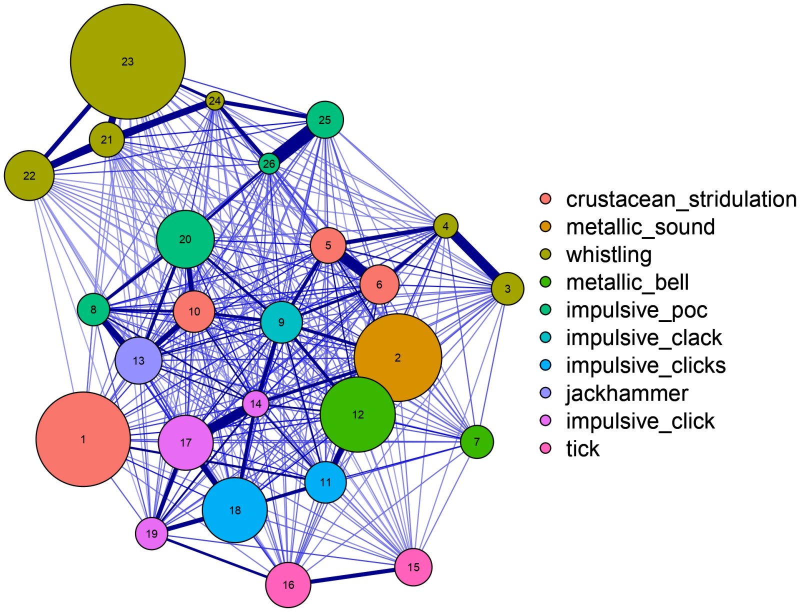


Table 8.1: Summary of descriptions of the obtained sound types after clustering. Spectrogram representations of each cluster are plotted in Supplementary Figure 8.1. Nt refers to the total number of samples. Nc refers to the number of samples inside a cluster. CN refers to the cluster number. Max and Min F are the average maximum and minimum frequency, respectively. Dur. refers to "Duration". Rev. (%) shows the number of samples correctly clustered.

Sound Name	Description	Nt	CN	Nc	Max F (kHz)	Min F (kHz)	Dur (s)	Rev. (%)
Whistling	Constant frequency tone around approximately 1200 Hz lasting up to several seconds	126	3	15	2.6	2.0	0.54	100
			4	10	2.5	2.0	0.34	100
			21	16	1.6	1.1	1.44	100
			22	25	1.7	1.2	1.08	100
			23	50	1.6	1.0	2.18	100
			24	10	1.6	1.0	0.74	100
Crustacean Stridulation	Semi-tonal component at around 1.3 kHz with multiple simultaneous impulsive sounds with energy up to 4 kHz, lasting up to 0.3 s	80	1	45	23.9	0.6	0.31	80
			5	11	2.9	1.8	0.19	91
			6	13	3.3	2.1	0.20	92
			10	11	4.7	0.1	0.25	64
Impulsive Poc	Short impulsive sound with a peak frequency around 2 kHz lasting around 40 ms	117	8	13	0.9	0.0	0.12	54
			20	55	2.0	0.2	0.10	44
			25	38	1.8	1.0	0.09	50
			26	11	1.8	0.9	0.10	91
Impulsive Click	Very short ($<$ 10 ms), broadband click between 3 and $>$ 24 kHz	39	14	10	12.7	4.8	0.11	90
			17	16	16.5	5.7	0.10	94
			19	13	22.8	8.8	0.08	62
Impulsive Clicks	Series of 'Impulsive Click' lasting up to a second and containing from 2 to 5 repetitions	74	11	17	22.3	3.0	0.38	65
			18	57	23.4	4.7	0.12	77

Metallic Bell	Fundamental frequency around approximately 2.2 kHz with higher frequency components up to 24 kHz lasting up to almost 2 seconds	78	7	10	23.5	1.3	1.25	100
			12	68	15.5	1.7	0.43	88
Tick	Series of very short (< 10 ms), high frequency clicks between 12 and 20 kHz	37	15	13	24.0	14.9	1.26	100
			16	24	24.0	14.4	0.30	100
Jackhammer	Series of short (< 30 ms), low frequency impulsive sounds between 300 and 2500 Hz	30	13	30	3.2	0.0	0.30	93
Impulsive Clack	Impulsive sound of about 30 ms and a peak frequency around 8 kHz	19	9	19	8.0	1.7	0.08	84
Metallic Sound	Very low in SNR presenting a tonal component at 6 kHz lasting 0.1 s	35	2	35	6.2	4.9	0.16	63

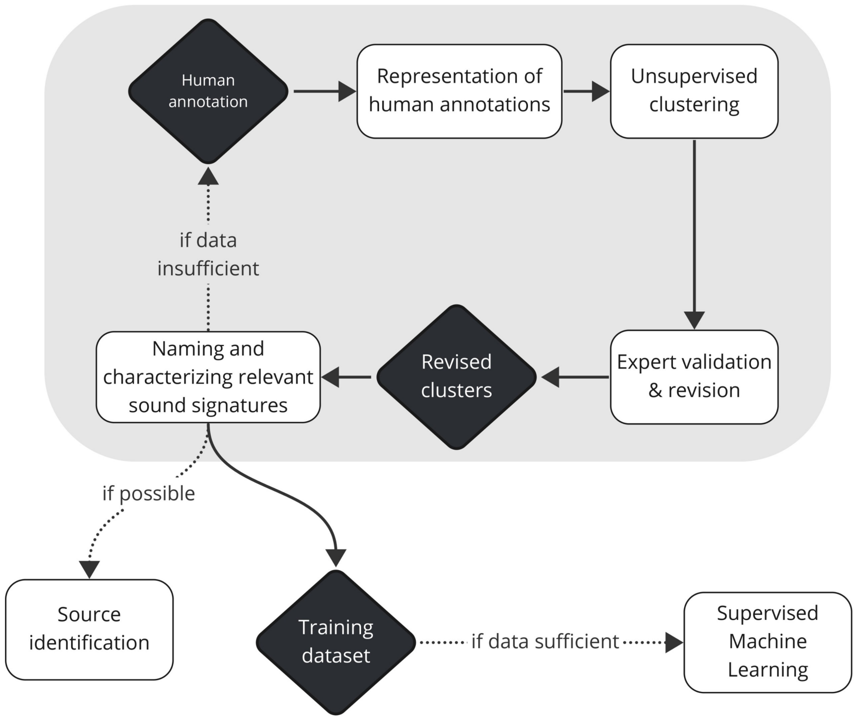
8.4 Discussion

The lack of reliable annotated training datasets and sound libraries is a critical methodological gap in studying soundscapes where sound sources are unknown. Our study demonstrated that unsupervised clustering of annotated unknown sounds eases revision and validation of annotated datasets. The revised clusters (Figure 8.7) we detailed already define a few groups of recurring distinct sounds that could serve as a preliminary component of an annotated training dataset. Annotation and labeling, in the absence of a reference dataset with validated annotations, are arduous and subjective especially for an underwater soundscape such as the BPNS, where sound signatures are unknown and the inherent acoustic scene is complex [42]. Although unsupervised clustering is conventional in ecological research [43, 44, 45], we highlight its practical use in revising clusters of annotated unknown sounds. Unsupervised clustering and subsequent revision of obtained clusters are therefore proposed steps to systematically reduce annotations to distinct and recurring sound events deemed relevant (Figure 8.9). With the proposed approach, labeling efforts only become a requisite for unclassified sounds of interest from newly gathered data, which do not fall under the same classification as the obtained revised clusters when clustered together with the old data, potentially forming new clusters. Sounds falling under the obtained clusters would not need to be manually labeled then. This approach speeds up the entire process of future human annotations and labeling efforts. The obtained datasets can be used for supervised ML (Figure 8.9), provided that the built training dataset is sufficient. Supervised ML models would render automatic detections and classification of already named and characterized sound signatures, whether the source is known or unknown. Human labeling and classifying efforts of these distinct and recurring sound events would then become less necessary in the future.

The most crucial step to achieve a meaningful characterization of recurring unknown sound signatures with our approach is the formation of relevant clusters. Two factors largely influenced the formation of clusters: (1) obtaining a relevant feature representation of the annotated dataset and (2) adjusting the hyperparameters of the chosen algorithms. Slight changes in any configuration altered the quality of clusters obtained. For instance, the choice of subjecting AVES feature sets to mean- or max-pooling (AVES-mean vs. AVES-max) or cropping spectrogram representations to the actual frequency limits and duration of the sound event (CAE-crops vs. CAE-original), accounted for significant differences in the homogeneity and shape of formed clusters (Figure 8.4). Selecting different minimum sizes of clusters significantly affected homogeneity scores, and the number of relevant features significantly affected DBCV scores.

While selecting the best representation model to extract features and applying the most appropriate clustering method have been obvious factors to consider in

Figure 8.9 Proposed steps to build a validated training dataset that can feed supervised machine learning (ML) models for unknown soundscapes. Unsupervised clustering is applied to representations of human annotations. Robust clusters, revised and validated by an expert, are named and characterized, and when possible, their sources are identified. The steps from human annotation to naming and characterization of relevant sound signatures (enclosed in a gray rectangle) are repeated in a cycle until the training dataset is sufficient to feed supervised ML models.



bioacoustics research, we highlight the performance variations brought by the large search space of hyperparameter configurations [24] which have remained obscure in the literature. As these configurations, mostly related to hyperparameters of algorithms, are often ambiguous and dataset-specific, grid search is therefore a step that should be considered when applying any algorithm. Though deep learning features, such as AVES and CAE, can be used efficiently by neural networks to classify sounds, not all extracted features are directly representative for any dataset. Varying numbers of selected discriminant features in the grid search using sparse PCA contributed to performance variation. Although, due to time-constraint, we only revised the best clustering result with the highest average of homogeneity and DBCV scores, it is also possible to revise other clustering results which performed as well, such as grid search result index # 141 (CAE-original).

In evaluating clustering performance, understanding the reliability of annotations determines the type of scoring metrics. As manual annotations made by a single individual are not fully reliable, additionally scoring the clusters through an unsupervised metric (DBCV) allowed for a reasonable evaluation of cluster quality. Scoring metrics must be cautiously interpreted, however, as high scores do not necessarily translate to relevant clusters. We excluded 43% of our grid search results from subsequent analyses, although some of these had higher homogeneity or DBCV scores, since these either gave too many clusters of the smallest size possible or very few clusters of the largest size possible, with more than 8% of the samples rejected by HDBSCAN. With highly conservative clusters, higher homogeneity and DBCV scores are obviously easier to achieve but would defeat the purpose of grouping vocalization repertoires per species or sounds derived from the same source in the same cluster. As a consequence of underwater sound variability (both in sound production and reception), sounds could possibly be originating from the same source yet either grouped into several smaller clusters, or classified as noise by HDBSCAN, due to slight differences in selected acoustic representation. The similarity of multiple clusters within the same classification of sound was evident in clusters represented by ‘Whistling’, ‘Impulsive Poc’, ‘Impulsive Click’ and ‘Tick’ (Figure 8.8). However, in some cases, variation within a cluster could be just as large as the variation between clusters represented by different sounds such as ‘Crustacean Stridulation’ (clusters 1, 10, 5, 6) and ‘Whistling’ (clusters 3-4 and clusters 21-24). This highlights the difficulty of manually categorizing and naming unidentified sound types. Multiple clusters of the same sound could arise from signals of varying SNRs, including the environmental noise of a recording station [46], overlapping in time and frequency [47]. Reception of signals can also differ due to differences in sound propagation and frequency-dependent loss brought by largely varying distances from the source [48]. This is common in harmonic sounds such as the ‘Metallic Bell’ where some clusters have less harmonics visible in the spectrogram. However, multiple clusters of the same sound could also be partly due to our imbalanced annotations, which is highly dominated by 'Crustacean Stridulation', 'Whistling', 'Impulsive Poc', 'Metallic Bell', 'Impulsive Click', 'Impulsive Clicks' and 'Tick' (Figure 8.2). Thus, small numbers of minimum cluster size and minimum samples were necessary to consider tags in the annotations with only a few samples. Since vocal repertoires remain unknown in the BPNS to date, we rely on this growing dataset to detect apparent and consistently similar sounds which are likely of the same origin. In the future, with a bigger dataset composed of annotations that are more representative of the BPNS soundscape, an ample number of bigger and denser clusters is plausibly easier to obtain.

Presently, there is no standard in annotating datasets, or which acoustic features should be used in bioacoustics research, when deciding whether two sounds are from the same source [49, 44]. In the present work, AVES and CAE performed differently depending on model configurations and hyperparameters. Meaningful clusters were obtainable from both models, although AVES-mean consistently ranked with the best scores since the exploration phase of this study. AVES feature sets resulted in more solid, homogeneous clusters with relatively lower intra-cluster distances, appearing to be more advantageous for the purpose of obtaining distinct clusters of recurring sounds with precision in this dataset. However, AVES clusters were disadvantaged by the relatively higher percentage of samples classified as noise by HDBSCAN. With CAE feature sets, a higher number of samples (68% on average) were clustered, but this resulted in more scattered and less homogeneous clusters which entail greater effort in manual revision. This could be due to the diversity of sound duration and frequency bands, hence no spectrogram parameters were found which could visually represent all different sound types appropriately. Furthermore, CAE was re-trained on our annotations containing only 2,875 examples, which might not be enough for a deep learning model. CAE might be more effective in representing datasets with a more uniform distribution in both sound duration and frequency boundaries, because then it is possible to ensure that all the spectrograms fed to the model are meaningful for all the annotations using the same parameters. CAE might also be more effective when trained on bigger datasets, such as those from biologically rich environments with a large amount of similar fish sounds. A possible approach to deal with the data limitation would be to use data augmentation techniques when training the model [50]. This could include, for example, training on both the original and the cropped spectrograms. The rigor of these models and the influence of hyperparameters must be continuously explored.

Though the purpose of this study was to obtain clusters of recurring sounds, a density-based clustering algorithm such as the HDBSCAN could potentially omit extremely infrequent sound types with insufficient data by classifying these as noise. Extremely rare sounds, which occurred less than five times in the dataset (e.g., snitch, impulsive knock, fish, electronic impact, bubble, siren, mouthbubble, impulsive tic toc and composite call; Figure 8.2), were classified as noise by the clustering algorithm as we expected. Separating rare sounds, which were not encountered on multiple occasions, from actual noise is a common clustering limitation that consequently underrepresents what is essentially a more biologically diverse soundscape. Denoising and processing signals prior to clustering can improve unsupervised learning, such as source separation techniques [51, 52]. However, there is currently no algorithm that satisfactorily addresses a broad spectrum of conditions for bioacoustics data in general, and marine bioacoustics in particular [53, 54]. For this reason, during a preliminary exploration of the feature extraction algorithms, we compared the results obtained when (1) applying a general non-stationary noise reduction algorithm (*noisereduce*; [55]), (2) no filter is applied, and (3) applying a band-pass filtering to each snippet. The results showed that for this dataset with such a variety of sounds, the band-pass filtering yielded the best results which could be due to several reasons. Firstly, for AVES and CAE-original, it allows for distinguishing between sounds occurring simultaneously at different frequency bands, which otherwise would be confused by the models. This is not the case for non-stationary noise reduction strategies, as both events happening simultaneously would be enhanced. Secondly, for low SNR annotations, a clearer signal can be obtained. Noise reduction algorithms such as *noisereduce* can enhance the SNR of signals, but they can perform badly for very short sounds such as 'Impulsive Click'.

While the complex acoustic scene of the BPNS has been previously described [42], sound signatures, especially of biological sources, remain unclear and unidentified. Of the final revised clusters in the present work, we were able to name and describe 10 unique sounds (Figs. 8.7 & 8.8). However, only cluster 13 with the 'Jackhammer' sounds and clusters 15 and 16 with the 'Tick' sounds can be interpreted as biological with some certainty. The 'Jackhammer' sounds fit within the known vocalization frequency range of fish ($<$ 3 kHz) and is a repetitive set of impulse sounds [11, 56]. They resemble sounds produced by fish species from the family Sciaenidae [57] and occurrences of an invasive species of this family have been also documented for the North Sea [58]. The short duration, high frequency 'Tick' sounds are similar to crustacean acoustic signals, which are known to span a large range of frequencies [59, 10]. While 'Metallic Bell' is largely supposed as mooring noise, we do not fully dismiss the possibility that it is biological. It has high resemblance to recorded sounds (although from a glass tank) of a spider crab Maja brachydactyla [60], which is present in the BPNS and often found in our moorings. However, these are just speculations as no ground-truth has been confirmed and further research would be necessary to assign the species to the sound with certainty. All other clustered sounds can only be interpreted with caution (e.g., 'Metallic Sound' could be a chirp of a mammal, 'Whistling' could be an anthropogenic sound originating from boats or geophony caused by wind, and 'Crustacean Stridulation' could be something called bio-abrasion, namely the mechanical disturbance of the recorder by an animal – [61]) as, to our knowledge, there are no similar sounds described in the literature so far. Spatiotemporal analyses of sound occurrence (which was not achievable in this study due to biases during data selection), alongside detailed comparisons of the acoustic characteristics of encountered sounds, are therefore necessary to further infer the possible biological sources of identified sounds.

Apart from AVES and CAE, there are numerous feature extraction models in the literature which perform differently from case-to-case. For example, Ozanich et al. (2021) [62] adds an extra clustering layer to an autoencoder similar to the CAE model to penalize points that are distant from cluster centers. If licenses of required MATLAB toolboxes are present, the software CASE (Cluster and Analyze Sound Events; [44]) can be freely downloaded for the purpose of selecting an appropriate clustering algorithm among 4 methods (community detection, affinity propagation, HDBSCAN and fuzzy clustering) and 3 classifiers (k-nearest neighbor, dynamic time-warping and cross-correlation) iterated over different values of parameters. Results are then subsequently evaluated using normalized mutual information (NMI), a scoring metric similar to the homogeneity score which relies on the level of agreement with pre-labeled data. If pre-labeled data are absent or highly unreliable due to the unidentified nature of the labels, we recommend coding a grid search function to easily compare results of differently tuned algorithms. Since incorporating gained information by different research groups is often difficult, we echo the pressing need for GLUBS [13] which will highly benefit bioacoustics research of unknown soundscapes such as the BPNS.

8.5 Conclusions

The proposed analysis evaluates the influence of certain hyperparameters on the clustering performance of acoustic events. We find out that even though they give similar results, AVES features perform better than the features from the Auto Encoder (CAE) for the unidentified sounds from the BPNS. Furthermore, we put emphasis on the importance of hyper-parameter tuning when clustering acoustic events, as different parameters will lead to very different clusters. To be able to evaluate these clusters it is preferable to have human annotations to compare them to. However, because of the unreliability of human labels and annotations the evaluation metrics need to be complemented with non-label dependent metrics. In any case, manual revision of sounds and clustering results are synergistic approaches which can be used by researchers simultaneously when screening for unknown sounds.

Acknowledgments

This work demonstrated LifeWatch observation data, which made use of infrastructure (the broadband acoustic network) provided by VLIZ and funded by Research Foundation - Flanders (FWO) as part of the Belgian contribution to LifeWatch. A lot of hands contributed to the maintenance of the network—we thank the Marine Observation Center and Infrastructure department of VLIZ and the crew of RV Simon Stevin for their continuous involvement and support at sea. Last but not the least, we thank Colien Bedoret for her dedication in annotating the recordings over the summer of 2023.

Data Availability Statement

Data will be made available at Marine Data Archive (MDA; https://marinedataarchive.org) upon publication.

Scripts are available at: https://github.com/lifewatch/unknown_underwater_sounds

References

[1] S. H. Gage and A. Farina. *Ecoacoustics Challenges*. In Ecoacoustics, chapter 18, pages 313–319. John Wiley & Sons, Ltd, 2017.

[2] A. V. Lindseth and P. S. Lobel. *Underwater Soundscape Monitoring and Fish Bioacoustics: A Review*. Fishes, 3(3):36, September 2018.

[3] A. J. R. Cotter. *The "Soundscape" of the Sea, Underwater Navigation, and Why We Should Be Listening More*. In Advances in Fisheries Science, chapter 19, pages 451–471. John Wiley & Sons, Ltd, 2008.

[4] N. Rako-Gospić and M. Picciulin. *Chapter 20 - Underwater Noise: Sources and Effects on Marine Life*. In C. Sheppard, editor, World Seas: An Environmental Evaluation (Second Edition), pages 367–389. Academic Press, January 2019.

[5] C. M. Duarte, L. Chapuis, S. P. Collin, D. P. Costa, R. P. Devassy, V. M. Eguiluz, C. Erbe, T. A. C. Gordon, B. S. Halpern, H. R. Harding, M. N. Havlik, M. Meekan, N. D. Merchant, J. L. Miksis-Olds, M. Parsons, M. Predragovic, A. N. Radford, C. A. Radford, S. D. Simpson, H. Slabbekoorn, E. Staaterman, I. C. Van Opzeeland, J. Winderen, X. Zhang, and F. Juanes. *The Soundscape of the Anthropocene Ocean*. Science, 371(6529):eaba4658, February 2021.

[6] A. Looby, K. Cox, S. Bravo, R. Rountree, F. Juanes, L. K. Reynolds, and C. W. Martin. *A Quantitative Inventory of Global Soniferous Fish Diversity*. Reviews in Fish Biology and Fisheries, 32(2):581–595, June 2022.

[7] R. Sousa-Lima. *A Review and Inventory of Fixed Autonomous Recorders for Passive Acoustic Monitoring of Marine Mammals*. Aquatic Mammals, 39(1):23–53, March 2013.

[8] D. Cato, R. McCauley, T. Rogers, and M. Noad. *Passive Acoustics for Monitoring Marine Animals - Progress and Challenges*. In New Zealand, 2006.

[9] J. C. Montgomery and C. A. Radford. *Marine Bioacoustics*. Current Biology, 27(11):R502–R507, June 2017.

[10] M. Solé, K. Kaifu, T. A. Mooney, S. L. Nedelec, F. Olivier, A. N. Radford, M. Vazzana, M. A. Wale, J. M. Semmens, S. D. Simpson, G. Buscaino, A. Hawkins, N. Aguilar de Soto, T. Akamatsu, L. Chauvaud, R. D. Day, Q. Fitzgibbon, R. D. McCauley, and M. André. *Marine Invertebrates and Noise*. Frontiers in Marine Science, 10, 2023.

[11] M. C. Amorim. *Diversity of Sound Production in Fish*. Communication in Fishes, 1:71–105, January 2006.

[12] M. J. G. Parsons, C. P. Salgado-Kent, S. A. Marley, A. N. Gavrilov, and R. D. McCauley. *Characterizing Diversity and Variation in Fish Choruses in Darwin Harbour*. ICES Journal of Marine Science, 73(8):2058–2074, September 2016.

[13] M. J. G. Parsons, T.-H. Lin, T. A. Mooney, C. Erbe, F. Juanes, M. Lammers, S. Li, S. Linke, A. Looby, S. L. Nedelec, I. Van Opzeeland, C. Radford, A. N. Rice, L. Sayigh, J. Stanley, E. Urban, and L. Di Iorio. *Sounding the Call for a Global Library of Underwater Biological Sounds*. Frontiers in Ecology and Evolution, 10, 2022.

[14] A. N. Rice, S. C. Farina, A. J. Makowski, I. M. Kaatz, P. S. Lobel, W. E. Bemis, and A. H. Bass. *Evolutionary Patterns in Sound Production across Fishes*. Ichthyology & Herpetology, 110(1):1–12, January 2022.

[15] J. Sueur and A. Farina. *Ecoacoustics: The Ecological Investigation and Interpretation of Environmental Sound*. Biosemiotics, 8(3):493–502, December 2015.

[16] T. A. Mooney, L. Di Iorio, M. Lammers, T.-H. Lin, S. L. Nedelec, M. Parsons, C. Radford, E. Urban, and J. Stanley. *Listening Forward: Approaching Marine Biodiversity Assessments Using Acoustic Methods*. Royal Society Open Science, 7(8):201287, 2020.

[17] S. Ness and G. Tzanetakis. *Human and Machine Annotation in the Orchive, a Large Scale Bioacoustic Archive*. In 2014 IEEE Global Conference on Signal and Information Processing (GlobalSIP), pages 1136–1140, December 2014.

[18] D. Stowell. *Computational Bioacoustics with Deep Learning: A Review and Roadmap*. PeerJ, 10:e13152, March 2022.

[19] E. C. Leroy, K. Thomisch, J.-Y. Royer, O. Boebel, and I. Van Opzeeland. *On the Reliability of Acoustic Annotations and Automatic Detections of Antarctic Blue Whale Calls under Different Acoustic Conditions*. The Journal of the Acoustical Society of America, 144(2):740–754, August 2018.

[20] P. Nguyen Hong Duc, M. Torterotot, F. Samaran, P. R. White, O. Gérard, O. Adam, and D. Cazau. *Assessing Inter-Annotator Agreement from Collaborative Annotation Campaign in Marine Bioacoustics*. Ecological Informatics, 61:101185, March 2021.

[21] J. M. van Osta, B. Dreis, E. Meyer, L. F. Grogan, and J. G. Castley. *An Active Learning Framework and Assessment of Inter-Annotator Agreement Facilitate Automated Recogniser Development for Vocalisations of a Rare Species, the Southern Black-Throated Finch (*Poephila Cincta Cincta*)*. Ecological Informatics, 77:102233, November 2023.

[22] K. Lisa Yang Center for Conservation Bioacoustics at the Cornell Lab of Ornithology. *Raven Pro: Interactive Sound Analysis Software (Version 1.6.4)*. The Cornell Lab of Ornithology, 2023.

[23] M. Hagiwara. *AVES: Animal Vocalization Encoder Based on Self-Supervision*. In ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pages 1–5, June 2023.

[24] P. Best, S. Paris, H. Glotin, and R. Marxer. *Deep Audio Embeddings for Vocalisation Clustering*. PLOS ONE, 18(7):e0283396, July 2023.

[25] P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J. van der Walt, M. Brett, J. Wilson, K. J. Millman, N. Mayorov, A. R. J. Nelson, E. Jones, R. Kern, E. Larson, C. J. Carey, İ. Polat, Y. Feng, E. W. Moore, J. VanderPlas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E. A. Quintero, C. R. Harris, A. M. Archibald, A. H. Ribeiro, F. Pedregosa, and P. van Mulbregt. *SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python*. Nature Methods, 17(3):261–272, March 2020.

[26] T. maintainers and contributors. *TorchVision: PyTorch's Computer Vision Library*. GitHub, 2016.

[27] M. Dash and H. Liu. *Feature Selection for Clustering*. In T. Terano, H. Liu, and A. L. P. Chen, editors, Knowledge Discovery and Data Mining. Current Issues and New Applications, Lecture Notes in Computer Science, pages 110–121, Berlin, Heidelberg, 2000. Springer.

[28] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and É. Duchesnay. *Scikit-Learn: Machine Learning in Python*. Journal of Machine Learning Research, 12(85):2825–2830, 2011.

[29] H. Zou, T. Hastie, and R. Tibshirani. *Sparse Principal Component Analysis*. Journal of Computational and Graphical Statistics, 15(2):265–286, June 2006.

[30] L. McInnes, J. Healy, and S. Astels. *Hdbscan: Hierarchical Density Based Clustering*. Journal of Open Source Software, 2(11):205, March 2017.

[31] R. J. G. B. Campello, D. Moulavi, and J. Sander. *Density-Based Clustering Based on Hierarchical Density Estimates*. In J. Pei, V. S. Tseng, L. Cao, H. Motoda, and G. Xu, editors, Advances in Knowledge Discovery and Data Mining, Lecture Notes in Computer Science, pages 160–172, Berlin, Heidelberg, 2013. Springer.

[32] C. Jenness. *DBCV*, 2017.

[33] A. Rosenberg and J. Hirschberg. *V-Measure: A Conditional Entropy-Based External Cluster Evaluation Measure*. In J. Eisner, editor, Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL), pages 410–420, Prague, Czech Republic, June 2007. Association for Computational Linguistics.

[34] D. Moulavi, P. A. Jaskowiak, R. J. G. B. Campello, A. Zimek, and J. Sander. *Density-Based Clustering Validation*. In Proceedings of the 2014 SIAM International Conference on Data Mining (SDM), Proceedings, pages 839–847. Society for Industrial and Applied Mathematics, April 2014.

[35] L. McInnes, J. Healy, and J. Melville. *UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction*, September 2020.

[36] S. S. SHAPIRO and M. B. WILK. *An Analysis of Variance Test for Normality (Complete Samples)†*. Biometrika, 52(3-4):591–611, December 1965.

[37] H. Levene. *Robust Tests for Equality of Variances*. Contributions to probability and statistics, pages 278–292, 1960.

[38] W. H. Kruskal and W. A. Wallis. *Use of Ranks in One-Criterion Variance Analysis*. Journal of the American Statistical Association, 47(260):583–621, December 1952.

[39] F. Wilcoxon. *Individual Comparisons by Ranking Methods*. Biometrics Bulletin, 1(6):80–83, 1945.

[40] Python developers. *Python Language Reference*, 2023.

[41] R Core Team (2023). *R: A Language and Environment for Statistical Computing*.

[42] C. Parcerisas, I. T. Roca, D. Botteldooren, P. Devos, and E. Debusschere. *Categorizing Shallow Marine Soundscapes Using Explained Clusters*. Journal of Marine Science and Engineering, 11(3):550, March 2023.

[43] T. Sainburg, M. Thielk, and T. Q. Gentner. *Finding, Visualizing, and Quantifying Latent Structure across Diverse Animal Vocal Repertoires*. PLOS Computational Biology, 16(10):e1008228, October 2020.

[44] S. Schneider, K. Hammerschmidt, and P. W. Dierkes. *Introducing the Software CASE (Cluster and Analyze Sound Events) by Comparing Different Clustering Methods and Audio Transformation Techniques Using Animal Vocalizations*. Animals, 12(16):2020, January 2022.

[45] M. J. Guerrero, C. L. Bedoya, J. D. López, J. M. Daza, and C. Isaza. *Acoustic Animal Identification Using Unsupervised Learning*. Methods in Ecology and Evolution, 14(6):1500–1514, 2023.

[46] M. C. Baker and D. M. Logue. *Population Differentiation in a Complex Bird Sound: A Comparison of Three Bioacoustical Analysis Procedures*. Ethology, 109(3):223–242, 2003.

[47] M. R. Izadi, R. Stevenson, and L. N. Kloepper. *Separation of Overlapping Sources in Bioacoustic Mixtures*. The Journal of the Acoustical Society of America, 147(3):1688–1696, March 2020.

[48] T. G. Forrest. *From Sender to Receiver: Propagation and Environmental Effects on Acoustic Signals*. American Zoologist, 34(6):644–654, December 1994.

[49] K. J. Odom, M. Araya-Salas, J. L. Morano, R. A. Ligon, G. M. Leighton, C. C. Taff, A. H. Dalziell, A. C. Billings, R. R. Germain, M. Pardo, L. G. De Andrade, D. Hedwig, S. C. Keen, Y. Shiu, R. A. Charif, M. S. Webster, and A. N. Rice. *Comparative Bioacoustics: A Roadmap for Quantifying and Comparing Animal Sounds across Diverse Taxa*. Biological Reviews, 96(4):1135–1159, August 2021.

[50] L. Nanni, G. Maguolo, and M. Paci. *Data Augmentation Approaches for Improving Animal Audio Classification*. Ecological Informatics, 57:101084, May 2020.

[51] Y.-J. Sun, S.-C. Yen, and T.-H. Lin. *soundscape_IR: A Source Separation Toolbox for Exploring Acoustic Diversity in Soundscapes*. Methods in Ecology and Evolution, 13(11):2347–2355, 2022.

[52] T.-H. Lin and S. Kawagucci. *Acoustic Twilight: A Year-Long Seafloor Monitoring Unveils Phenological Patterns in the Abyssal Soundscape*. Limnology and Oceanography Letters, 9(1):23–32, 2024.

[53] J. Xie, J. G. Colonna, and J. Zhang. *Bioacoustic Signal Denoising: A Review*. Artificial Intelligence Review, 54(5):3575–3597, June 2021.

[54] J. Juodakis and S. Marsland. *Wind-Robust Sound Event Detection and Denoising for Bioacoustics*. Methods in Ecology and Evolution, 13(9):2005–2017, 2022.

[55] T. Sainburg, A. Zorea, A. Saghiran, K. Amr, Lff5, fardage, and rjfarber. *Timsainb/Noisereduce: 3.0.2*. Zenodo, March 2024.

[56] R. Carriço, M. A. Silva, G. M. Menezes, P. J. Fonseca, and M. C. P. Amorim. *Characterization of the Acoustic Community of Vocal Fishes in the Azores*. PeerJ, 7:e7772, November 2019.

[57] M. C. P. Amorim, J. A. Wanjala, M. Vieira, M. Bolgan, M. A. Connaughton, B. P. Pereira, P. J. Fonseca, and F. Ribeiro. *Detection of Invasive Fish Species with Passive Acoustics: Discriminating between Native and Non-Indigenous Sciaenids*. Marine Environmental Research, 188:106017, June 2023.

[58] P. Morais, I. Cerveira, and M. A. Teodósio. *An Update on the Invasion of Weakfish Cynoscion Regalis (Bloch & Schneider, 1801) (Actinopterygii: Sciaenidae) into Europe*. Diversity, 9(4):47, December 2017.

[59] N. J. Edmonds, C. J. Firmin, D. Goldsmith, R. C. Faulkner, and D. T. Wood. *A Review of Crustacean Sensitivity to High Amplitude Underwater Noise: Data Needs for Effective Risk Assessment in Relation to UK Commercial Species*. Marine Pollution Bulletin, 108(1):5–11, July 2016.

[60] L. Coquereau, J. Grall, L. Chauvaud, C. Gervaise, J. Clavier, A. Jolivet, and L. Di Iorio. *Sound Production and Associated Behaviours of Benthic Invertebrates from a Coastal Habitat in the North-East Atlantic*. Marine Biology, 163(5):127, May 2016.

[61] J. P. Ryan, J. E. Joseph, T. Margolina, L. T. Hatch, A. Azzara, A. Reyes, B. L. Southall, A. DeVogelaere, L. E. Peavey Reeves, Y. Zhang, D. E. Cline, B. Jones, P. McGill, S. Baumann-Pickering, and A. K. Stimpert. *Reduction of Low-Frequency Vessel Noise in Monterey Bay National Marine Sanctuary During the COVID-19 Pandemic*. Frontiers in Marine Science, 8, 2021.

[62] E. Ozanich, A. Thode, P. Gerstoft, L. A. Freeman, and S. Freeman. *Deep Embedded Clustering of Coral Reef Bioacoustics*. The Journal of the Acoustical Society of America, 149(4):2587–2601, April 2021.

Part IV

Applications of Understanding Soundscapes

9 An example of the practical use of soundscape understanding

This chapter is a reproduction of the publication:

- *Schmidlin, S. & Parcerisas C., Hubert, J., Watson, M.S., Mees, J., Botteldooren, D., Devos. P., Debusschere, E., & Hablützel, P.I. Comparison of the effects of reef and anthropogenic soundscapes on oyster larvae settlement. Scientific Reports. 2024*

Overview

In this chapter we provide an example of why it is important to understand soundscapes, distinguish them and characterize them. In a laboratory experiment we exposed larvae of true oyster *Magallana gigas* to different real-world soundscapes including reef, off-reef and boat recordings. A fourth treatment was added where boat sounds were artificially added to the reef sounds. The results show that the soundscape to which they were exposed to did have an effect on their settlement rate. Reef recordings the soundscapes triggering the highest settlement, and masking boat sounds would decrease this settlement rate, possible having an effect on recruitment. This chapter answers research question:

Are there examples where the soundscape plays a crucial role in the development of local fauna? Specifically, does the natural soundscape impact the

settlement behavior of oyster larvae?

9.1 Introduction

Identifying a suitable habitat prior to permanently transitioning to a benthic life stage is critical for future survival, growth, and reproduction in many marine invertebrates with planktonic larvae. These species therefore utilize of a variety of environmental cues, enabling them to identify promising settlement locations[1]. Experimental research has shown that in some species, a single cue can induce settlement and subsequent metamorphosis [1, 2, 3]. But in many other species larvae may respond to more than one cue, *Crassostrea virginica* larvae for example respond similarly to chemicals released by conspecific adults, and chemicals released from mature biofilms [4]. Cues can have chemical and physical origins, and while some types of cues require contact with a prospective settlement location such as cues associated with shells of conspecifics adults or cues from topographical features of a substrate [5, 6], other cues may act over larger distances to guide larvae to their preferred habitat such as those released by waterborne conspecific chemicals [1, 7]. More recently, acoustic cues have been identified as drivers of larval settlement [8, 9]. As sound propagates relatively fast and far underwater, it serves as an efficient signal transmission medium. For many marine species, sounds can convey specific events, such as presences of a predator or a mating opportunity [10, 11]. But, collectively, soundscapes can also convey overall quality and suitability of an environment for a species [12, 13]. Research on acoustic cues informing larvae about optimal habitats has only been established relatively recently [8, 9]. In certain invertebrate species with a settlement/metamorphosis stage, including crabs, corals, and bivalves, acoustic cues have been shown to affect larvae swimming direction [8, 9], settlement rates [9, 14], and amount of time a larvae takes from entering competency to completing metamorphosis [15, 16, 17]. In general, it seems that natural environmental sounds can convey information to invertebrate species in that environment [18]. In coral and bivalve reefs, larvae seem to be attracted to soundscapes from healthier reefs, which produce louder and more acoustically complex sounds compared to less healthy reefs which are much quieter [14, 19]. However, the particular characteristics of reef soundscapes (e.g. sound pressure level (SPL), specific frequencies, complex mixtures of these or other acoustic characteristics) that elicit settlement behaviors remain unclear.

Anthropogenic sounds may interfere with or mask natural marine soundscapes [12]. Vessel noise can mask important sound cues resulting in poorer orientation toward reef sounds for some species of fish [20, 21], and cause coral larvae (planulae) to delay settlement [22]. Anthropogenic noise can not only disrupt or reduce larval settlement but may also be (mis)interpreted as a cue to settle in some taxa [16, 17, 23, 24, 25]. Vessel noises have been shown to increase some larvae settlement, including in mussel *Perna canaliculus* [16] and *Mytilus edulis* [24]. Why anthropogenic noises are interpreted as settlement cues in some taxa but are repulsive to others is unknown. The reaction to anthropogenic sounds may depend on the acoustic profile of a species' preferred habitat and which features of this profile are responsible for attraction [17, 23, 26, 27].

The oviparous true oyster *Magallana gigas* is an important reef-building ecosystem engineer [28] and a valuable species for aquaculture [29]. But in many areas it is invasive and considered a biofouling pest that poses a threat to local species and ecosystems [30, 31]. There is considerable interest in settlement preferences of this species for both bolstering as well as reducing recruitment [32]. In recent years, there has been a global effort to restore oyster reefs, as widespread habitat destruction have left native historical populations decimated [32]. Availability of settlement cues is crucial for reef sustainment, with some reports suggesting that these cues may outweigh other recruitment factors such as local hydrodynamics, and larvae supply [33, 34]. A recent revelation that oyster *Ostrea angasi* not only settle more rapidly but also exhibit horizontal swimming movements toward sound sources underscores the significance of soundscapes as a navigation tool for larvae [9]. So far, larvae of *M. gigas* have not been studied for their response to acoustic settlement cues (but see Stocks et al,. (2012) [23]) for an account of swimming activity in response to natural and vessel sounds). *M. gigas* adults have been studied for their sense of hearing, and were found to react by valve closure to pure tones in the range of 10 to 1000 Hz at minimum energy of 122 dBrms re 1 μPa [35]. While these adults are studied for their pure tone reactions, the range of hearing of these larvae have not yet been identified. Other true oysters with relevant experimental data are the closely related and also oviparous *C. virginica*, and the more distantly related larviparous *O. angasi*. Experimental studies have shown that both *C. virginica* and *O. angasi* larvae prefer louder reef sounds in frequency ranges 1.5-20kHz over quieter off-reef playbacks or no-sound controls [9, 36, 7].

In this study, we present the results of laboratory-playback based settlement experiments on the role of acoustic cues in settlement and metamorphosis of *Magallana gigas*. Firstly, we were interested in the importance of oyster reef sound compared to off-reef sound. Secondly, we wanted to know whether vessel noise attracts or repels pediveliger larvae. To do so, we exposed larvae to different vessel and reef sounds as well as off-reef and no-speaker controls. Finally, we subjected the larvae to vessel and reef sounds simultaneously to find out whether vessel noise modifies, or masks oyster reef sound cues.

9.2 Methods

We conducted laboratory tank-based playback experiments to investigate whether oyster species *Magallana gigas* larvae alter their settlement response in reaction to sounds emitted by conspecific oyster reefs and vessels. Sound recordings were obtained from two regions within the North Sea, and acoustic spectral features were analyzed based on recordings made within experimental tanks.

9.2.1 North Sea soundscape measurements

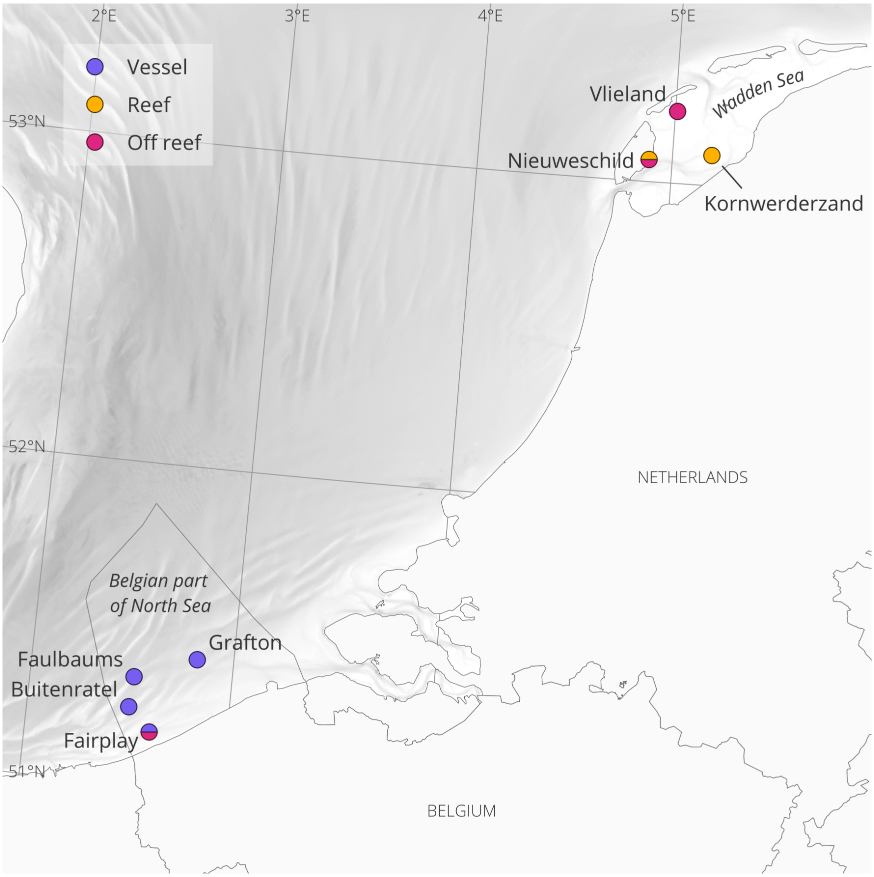
All recordings used during the experiment were recorded in two regions of the North Sea: the Southern Bight near the Belgian coast and in the Dutch Wadden Sea (see Figure 9.1). In the Dutch Wadden Sea, recordings at subtidal reef sites were collected with hydrophones (SoundTrap 300STD; Ocean Instruments, NZ; sampling rate 24 kHz; manufacturer-calibrated; set at low gain; see Table S9.1 for details). Hydrophones were suspended in PVC frames anchored to the seafloor and a subsurface buoy ensured that the hydrophone was positioned approximately 1 m above the seafloor in water depths ranging from 2 to 5 m. These were deployed from a small boat, and left to record continuously for two weeks or the maximum battery life. Reef sound recordings used for this experiment were taken at two subtidal oyster reefs in the Marsdiep tidal basin of the eastern Wadden Sea (Table S9.2). An off-reef sand recording used in this experiment was obtained from a control recording of an artificial reef monitoring in the Eierlandse Gat tidal basin, near the island Vlieland.

For the Belgian part of the North Sea, recordings from the Stationary Dataset (see Section 2.2) were considered.

9.2.2 Sound treatment and playbacks

Suitable recording files for each treatment were manually selected. Only Belgian recordings from spring and summer were considered, in order to correspond to the recording period of the sounds from the Wadden Sea (reef and off-reef). Reef sounds were only selected when they contained no apparent outside influences (e.g. vessel sounds). For vessel sounds, a fair variability of sounds was selected, from short sounds of distant vessels to longer continuous sounds from vessels operating close by, with no other audible background sounds. All vessel sounds were recorded from locations in the Southern Bight (see Table S9.2). For reef treatment, sounds used were recorded from the same location in Texel, NL but sound files used during each day of the experiment were selected from different recording dates (see Table S9.2). For off-reef sounds, two sound files were used recorded from Texel, NL and two sound files were used recorded from non-reef areas in the Southern Bight off the coast of Belgium (see Table S9.2). Treatments where vessel sounds and reef sounds were played together were created artificially overlaying reef sound files and vessel sound files. Treatments where vessel sounds and reef sounds were played together were created artificially overlaying reef sound files and vessel sound files.

Figure 9.1 Distribution of the locations where underwater sound data were collected. Colors represent which treatment was collected there. Sounds acquired in locations with two colors were used for different treatments. Map made by maps@vliz using QGIS version 3.30.0-'s-Hertogenbosch (`https://qgis.org/en/site`)



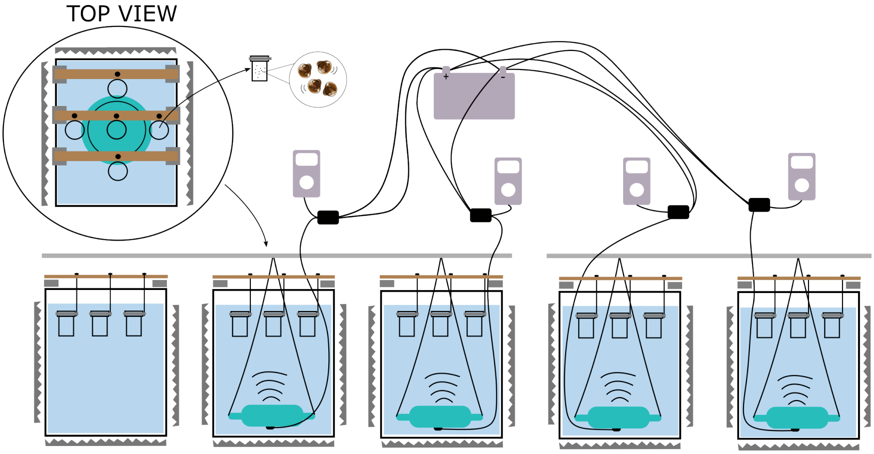
Selected segments were then combined to create one 1 h file per treatment and day. In some cases, the selection led to files shorter than 1 h, so the segments were repeated and combined by applying crossfading with Audacity [37] to create a 1 h file. When enough recordings were available for 1 h or more, segments were not repeated. To deal with differences in sampling rate and minimum recording frequency between the selected files, all files were filtered using a Butterworth bandpass filter (N=4) between 20 Hz and 12 kHz. After filtering, all files were downsampled or upsampled to 48 ksps to match the playback requirements. Details of the selected data are listed in Table S9.2 (Supplementary Material). In total, 3 recordings of reefs from 2 different locations, 4 vessel recordings with several boats on each recording from 4 different locations, and 4 off-reef recordings from 3 different locations were used to represent our treatments.

Throughout the experiment, each treatment group containing sound playback (“reef” “vessel” “reef+vessel” “off reef”) consisted of a separate recording representing the intended environment. The use of multiple sound files of the same treatment was used to strengthen confidence that the sounds were representative of treatments as a whole, and not of a single event. Employing a series of recordings from various sound sources representing the same treatment enhances the extrapolative capacity of a study [38, 39, 40].

The playback set-up consisted of five 100L tanks (49x65x50.5 cm), separated 20 cm from each other on a rack. Each tank sat upon a 4 cm layer of polystyrene to isolate it from the rack and an additional layer of acoustically absorbent foam (25 mm thick) between the polystyrene and the tank bottom. Acoustic foam was also placed at the tank sides. Four Lubell UW30 Underwater Speakers with custom-made amplifiers, battery-powered to avoid 50 Hz noise or electrical interference from the power grid, were used. Each speaker was connected to one TASCAM playback device which played playback files on repeat. No speaker was placed in the no-sound control (see Figure 9.2). This no speaker treatment was added as a second control (aside from the off-reef control) to establish if there were differences between treatments with sound and normal lab conditions. The speakers were hung in the middle of the tank with ropes so they would not touch the tank walls. Larvae were placed inside 100 ml polystyrene jars and these containers were fixed in the same position in the tank for every day of the experiment. These positions were 12.75 cm far from the closest jar (distance between outer jars and center jar), and 33.5 from the speakers (see Figure 9.2).

For each treatment file, a playback volume was chosen so exposure power spectral density (PSD) would match the sound levels recorded in field as closely as possible. Sound levels specified in literature as typical of reefs (at 1 m from the seafloor) and off-reefs (at 2 km from the reef) [41, 9] roughly matched the chosen levels. This was done separately for each sound exposure treatment by adjusting the volume in an iterative fashion until the recorded sound levels in tanks were

Figure 9.2 Schematic depicting five tanks. Four of these tanks are equipped with speakers, each of which is connected to a playback device. All speakers and playback devices are linked to a DC battery as their power source. Five 100ml jars were securely positioned to hang at the same height above speakers. The speakers were suspended within the tanks in a manner ensuring they did not come into contact with the tank's bottom.



similar to the desired ones. For this purpose, each tank setup was recorded using a TASCAM recorder device together with an Aquarian Scientific hydrophone (AS-1). The hydrophone was placed inside a jar without touching the walls, simulating the position where larvae would be during the experiments. The hydrophone cable was taped to the structure supporting the jars. Recorded sound was converted to sound pressure by using the available calibration data for the hydrophone and the TASCAM recorder. This calibration was cross-checked by comparing it to measurements of the hydrophone used in Hubert et al (2022) [42].

To measure sound level and acoustic characteristics of each playback received by larvae during the experiment, each treatment was recorded using the chosen playback volume for 1 h (tank recordings) at 48 ksps. This same procedure was used when selecting the playback volume. When recording these 1 h files, all four different sound treatments of that batch were turned on to record possible acoustic crosstalk from the other treatments. Furthermore, the room noise was also recorded using the same protocol when no speaker was active.

For each treatment, several acoustic features were computed for both 1 h recorded tank files and 1 h compiled field files. Acoustic Complexity Index (ACI), Acoustic Evenness Index (AEI), and Acoustic Diversity Index (ADI) were computed using the maad python package [43], and Power Spectrum Density (PSD) was computed using the scipy python package [44]. The average PSD was computed for three different bands by averaging the spectrum density of all the frequency bins included in the specified frequency band. The parameters used to compute each of the features are summarized in Table S9.3 (Supplementary Material), and the equations to compute the acoustic indices are in the Supplementary Material, equations 11.1, 11.2, 11.3, and 11.4. Both ACI and ADI are proxies to quantify acoustic complexity (the higher the number, the more complex), while low values of AEI represent an even sound and higher values represent more uneven sounds. This is not correlated with the ecological concept of evenness, as acoustic evenness refers to an even distribution of sound energy in different frequency bands, and this can be achieved due to a high biodiversity vocalizing at the same time covering all the frequency bands or by constant broadband sounds such as some anthropogenic sounds [45, 46].

9.2.3 Broodstock and Larvae Culture

Ten mature adult oysters (five females and five males) were purchased from the Guernsey Sea Farms Ltd (Guernsey, UK) and used to produce larvae. Eggs were fertilized by gonad stripping following FAO guidelines [47]. Fertilized eggs were kept undisturbed in flat bottom tanks for 48 hours at 22 °C at a density of ten eggs per ml of filtered seawater (FSW). All seawater used in this experiment was filtered at 0.1 μm and passed through UV light. After 48 hours larvae were sieved over 70 μl nylon mesh sieve, rinsed, and transferred to rearing tanks with FSW. Tanks were aerated and kept at 22 °C for the entire duration of larvae rearing. Every two days larvae were sieved over mesh corresponding to the average size of the larvae and the water in the tanks was changed. Larvae were fed a mixture of fresh microalgae mixture consisting of *Chaetoceros muelleri*, and *Isochrysis galbana* (clone T-ISO). For the first 4 days larvae were fed at 40,000 cells/ml water using only *I. galbana* (clone T-ISO). Days 5-12 larvae were fed *C. muelleri*, and *I. galbana* (clone T-ISO) at 100,000 cells/ml at a volume ratio of 1:1. Days 13+ larvae were fed *C. muelleri*, and *I. galbana* (clone T-ISO) at 100,000 cells/ml at a volume ratio of 3:1. Larvae entered their pediveliger stage and became competent to settle at 29 days and were used in settlement experiments starting on this day. Larvae were determined for competence when they had a prominently displayed eyespot and larval foot and were sized at 320-350 μm in diameter.

9.2.4 Settlement Experiment Design

The experiment aimed at assessing the effect of sound treatment on larvae settlement. In the experimental design, we were constrained by having only four underwater speakers at our disposition. We therefore repeated the experiment four times over five consecutive days. In each of the four trials, different treatments were assigned to a unique combination of speaker and tank to account for possible speaker or tank effects. The sound treatment was applied at the tank level, making tank the experimental unit. As the experiment has a binary outcome (settled vs. not settled), many sample units (larvae) are needed to accurately assess the treatment effect. Therefore, ten larvae were placed together in jars and five such jars were placed in each tank (see Figure 9.2). Larvae were not re-used and for each trial, new pediveliger stage larvae were taken randomly for the same stock. As a consequence, larvae were gradually older through the five day experiment.

9.2.4.1 Settlement Assays

Experiments took place over five consecutive days (03/03/23 – 07/03/23) using larvae from the same batch. To control for larvae size, on each experiment day some larvae from culture tanks were filtered between 260 μm and 300 μm nylon mesh sieves, only larvae retained on the 260 μm sieve were used in the experiment. 10 larvae were gently pipetted randomly into each of the five 100 ml containers per tank and filled with filtered seawater (FSW) and 0.2 grams of oyster shells which could act as a settlement substrate. To get a consistent shell topography, shells were crushed using a hammer and crushed shells were sieved between 1.0 mm and 0.5 mm metal sieve. For each treatment tank, 5 individual containers were used. As all treatments were repeated over 4 consecutive days, 20 jars were used per treatment in total. All trials were conducted in a dark environment at 20 ($\pm$1) °C in a climate-controlled room.

To avoid any air in jars containing larvae, larvae were placed in the jars and the lid was fixed while the jar was fully submerged in FSW. This step was necessary to prevent any distortion of the sounds due to reflection from air bubbles. All FSW used in the experiment had added microalgae *Chaetoceros muelleri*, and *Isochrysis galbana* (clone T-ISO) at 100,000 cells/ml at a volume ratio of 3:1. In a previous study, *M. gigas* larvae increased swimming when exposed to reef sounds, but only if larvae were fed [23], thus microalgae were added to our larvae containers. Microalgae were added at the same concentration as used in larvae rearing tanks and food levels were not limiting for the duration of the experiment.

On top of each tank, jars were attached to a wooden pole sitting horizontally across the tank. Each larvae jar was attached so that it was in a fixed position for the duration of the experiment, the position of the jar was noted so that the effect from placement in the tank could be ruled out. Jars were fully closed so that no water was shared between tanks and jars. The wooden pole was isolated from tank walls with polystyrene to avoid vibration propagation. One jar was located directly above the speaker and the other 4 jars were at the same distance from the center of the speaker (see Figure 9.2).

After 24 hours of exposure, larvae metamorphosis was checked using a dissecting microscope and the number of larvae that had cemented themselves to the substrates were counted. Metamorphosis was confirmed by gently blowing water over the larvae with a pipette to ensure that larvae were fixed to the substrate.

9.2.4.2 Statistical analyses

A generalized linear mixed-effect model was created using the glmer function of the lmer package [48] in R version 4.1.3 (2022-03-10) [49]. As the response variable was binary (settled vs. not settled) we fitted a Bernoulli distribution using a logit link function. The predictor variables that were considered included sound treatment, date, speaker, tank, jar and jar position. First, we established a base model that included treatment and date as fixed effect variables and jars nested in the treatment-tank interaction as a random effect variable. We then included each of the other variables (i.e. speaker, tank and jar position) individually as fixed effects and compared model fit using the Akaike information criterion (AIC) and visually inspected model prediction plots. As the inclusion of any of those variables did not decrease the AIC value and had little to no effect on the effect sizes, we did not consider them in the final analysis. The assumptions of the model were met. See supplementary information for description of all model used (Table S9.5). Post hoc tests were performed using the emmeans function of the lsmeans package [50] to calculate the marginal means adjusting p-values for multiple comparisons with Tukey's method and the pairs function was used to display pairwise comparisons.

9.3 Results

9.3.1 Playback

The recorded sound in the tanks did not perfectly match the spectrum of the sounds recorded in the field due to the technical limitations of the reproduction equipment and the resonances that inevitably occur in tank-based experiments (Figure 9.3). For example, sound levels were amplified at 2 and 7 kHz due to the speaker frequency response, with a dip at 5 kHz. A 50 Hz and its 3rd and 5th harmonic can be observed in the PSD of all the treatments, probably generated in the hydrophone due to electromagnetic interferences from the lab (not coming from the playback system and detectable by the larvae). Despite these limitations, when computing different acoustic metrics in tank and field sound recordings, similar trends were observed (Figure 9.4). For example, ACI was higher for reef treatments compared to all other treatments in both tank and field sound recordings, and the PSD order from loudest to quietest for each batch was the same for field and tank recordings.

When analyzing the tank recordings, some faint cross-talk between tanks could be detected. This occurred only below 200 Hz, and it was most audible in the no speaker tank recording during the loudest moments of the neighboring tank. Nevertheless, the treatments remained very different and recognizable acoustically.

Figure 9.3 Comparison between the field and the tank recorded playback spectrum levels. No speaker refers to the tank recording when all the other playbacks were on (recorded in the tank with no speaker inside)

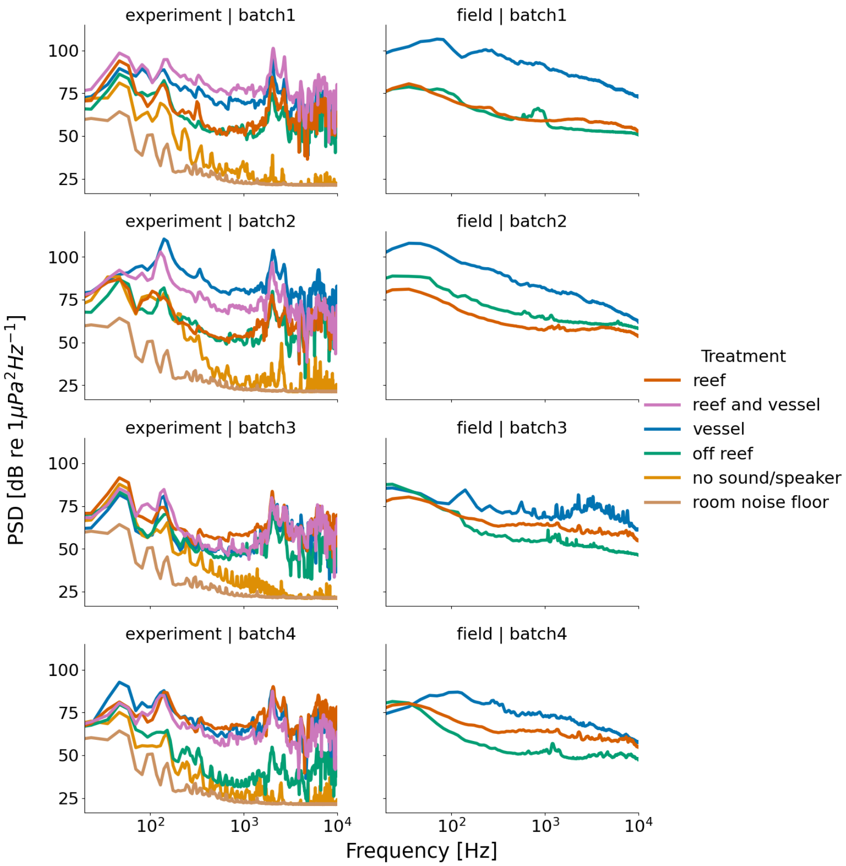
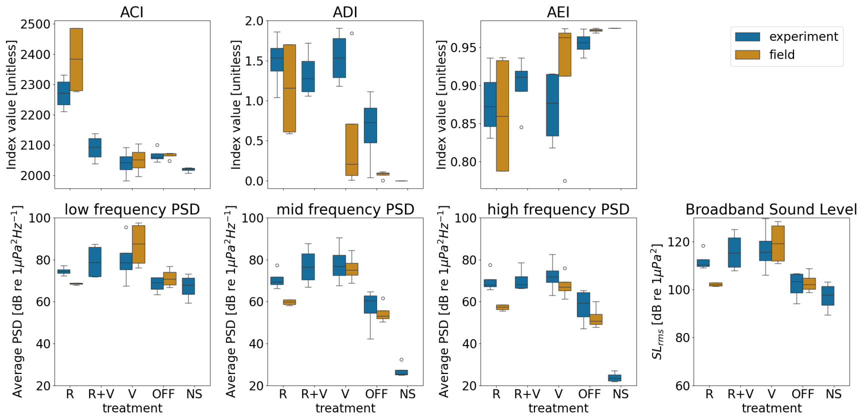


Figure 9.4 Comparison of the obtained acoustic metrics for the tank and the field recordings. The acoustic metrics include the acoustic complexity index (ACI), acoustic diversity index (ADI), and acoustic evenness index (AEI). NS=no speaker, OFF=off reef, R=Reef, R+V=reef and vessel, V=vessel. The number of data points for each treatment is 4. The definition and computation of each of the features are explained in Table S9.3



9.3.2 Settlement rate

Larvae settlement increased significantly in response to reef sound compared to vessel sounds (β = 0.715, SE = 0.260, p = 0.047), compared to off-reef sounds (β = 0.745, SE = 0.261, p = 0.034), and compared to the no speaker treatment (β = 1.015, SE = 0.262, p = 0.0010; Table 9.1). When vessel sound were added to the reef sound, the settlement rate decreased about 1.29 times compared to the pure reef sound (β = 0.560, SE = 0.259, p = 0.193), and was 1.09 times higher than in the vessel-only sound treatment (β = 0.155, SE = 0.261, p = 0.976). Comparisons among other treatments revealed only minor differences (Table 9.1). Vessels and off-reef sounds had very similar effects on settlement. The lowest settlement rates were observed in a no-sound control treatment. Model predictions are plotted in Figure 9.5.

9.4 Discussion

The results of our laboratory experiment showed increased settlement during exposure to sounds of conspecific oyster reefs for larvae of the oyster *Magallana gigas*, with settlement increasing 1.44 and 1.64 times under the oyster reef treatment compared to off-reef and no speaker control treatments, respectively. The settlement of larvae in response to vessel sounds, as well as the combined effects of reef and

Table 9.1 The results of the posthoc of the GLMER model 1 using all data and comparing all treatments. Significant values ($p \leq 0.05$) are in **bold**.

Contrasting treatments	Estimate	SE	df	z.ratio	p.value
reef - off reef	0.7455	0.261	Inf	2.86	**0.0344**
reef - vessel	0.7154	0.260	Inf	2.750	**0.0471**
reef - no speaker	1.0151	0.262	Inf	3.879	**0.0010**
reef - (reef + vessel)	0.5600	0.259	Inf	2.164	0.1934
off reef - vessel	0.0301	0.263	Inf	0.115	1
off reef - no speaker	0.2695	0.263	Inf	1.025	0.8441
off reef - (reef + vessel)	0.1855	0.261	Inf	-0.711	0.9541
no speaker - vessel	0.2997	0.264	Inf	-1.137	0.7869
no speaker - (reef + vessel)	0.4551	0.262	Inf	-1.738	0.4103
vessel - (reef + vessel)	0.1554	0.261	Inf	-0.595	0.9758

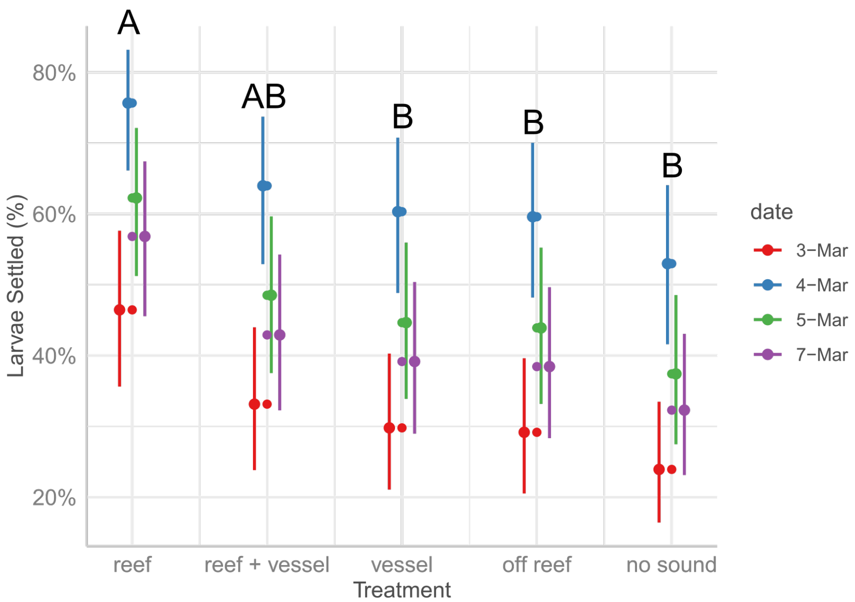
vessel treatments, did not show statistically significant differences compared to off-reef conditions and controls with no speaker.

9.4.1 Preference for oyster reef sounds

Larvae of *M. gigas* consistently settle more readily when exposed to sounds of reefs inhabited by conspecifics. Our finding thus corroborates earlier research in fish, corals, and other oyster species, where larvae were found to increase settlement or orient more readily towards playback of reef sounds [8, 36, 51, 52]. Yet the sound features that trigger this response still remain unidentified. In general, oyster and coral reefs exhibit higher sound levels and greater acoustic diversity than off-reef counterparts, due to increased soniferous biological activity including vocalizations of soniferous fishes and invertebrates, both passive or active, as well as the physical complexity of the reef [14, 22].

It remains undecided in the literature if larvae can distinguish particular sounds from different habitats, or if there is simply a preference for certain acoustic features such as SPL [19]. The spectrum of reef sounds recorded for our study followed patterns similar to other oyster reefs [14, 9], and were louder than off-reefs in two out of four experiments. They also consistently presented a higher acoustic complexity, and higher evenness (lower AEI value) than off-reef areas. Compared to the vessel sounds, our reef sounds tended to have similar or lower PSD (depending on the vessel). Reef sounds were unique amongst the other treatments in their diversity, with consistently higher ACI and ADI values, and lower AEI values. This indicates that loudness (SPL) alone is not responsible for larval attraction, instead spectro-temporal patterns responsible for a high ACI may play a more important role. This conclusion can be corroborated in other marine invertebrates [25, 17]. Pine et al., (2012) [17] found that crab megalopae reduce metamorphosis

Figure 9.5 Prediction plots comparing predicted settlement across sound treatments. The average settlement from each 100 ml jar on each day is represented in circles. Error bars represent 95 % confidence intervals of the model prediction. Letters represent significant differences in the treatment, same letters mean no significant difference between treatments, but different letters indicate a significant difference



(in comparison to natural habitat sounds) when exposed to wind turbine noise, but when the same turbine noises were played back at higher SPL, this did not result in any further changes to crab metamorphosis time. Leading to the conclusion that spectro-temporal characteristics were more relevant feature for megolopae attraction to habitat sounds than volume. Similarly, Gigot et al., (2023) [25] found that scallop larvae reduced metamorphosis rates during drilling sounds, but increased metamorphosis rates when exposed to pile driving sounds. As both sounds were substantially louder than the no control, this further indicates the importance of temporal and spectral composition over simple preference for louder sounds. It would be incorrect to say that louder sounds are not a preferred sound feature for a number of invertebrates. Wilkens et al. (2012) [16] found that when exposed to (the same) vessel sound at increasingly louder SPLs, mussel larvae increased settling at the louder treatments. Lillis et al., (2016) [19] also conclude that louder reefs attract more coral settlers than quieter reefs. Based on the results of this present study and associated literature, a reasonable hypothesized could be made that both of these sound qualities (loudness and spectro-temporal patterns) are perceptible to

M. gigas larvae, and the preferences for each may be highly species-specific and could be based on the preferred habit qualities. In comparison to *M.gigas* adults, who have previously been studied for their sense of hearing [35], larvae appear to be able to detect a larger range of acoustic frequencies. Future research should therefore not only collect species specific data on acoustic feature detection, but also from different life stages.

9.4.2 Vessel noises and larvae settlement

Our results show that exposure to vessel sounds alone did not manifest in any disruptions to settlement compared to off-reef or no speaker controls. This indicates that there may be no intrinsically negative reaction of *M. gigas* larvae to these vessel sounds. In marine invertebrates generally, vessel sounds induce a wide range of physiological and behavioral changes (see [18] and [53] for reviews of vessels on marine invertebrates). Much of the evidence indicates a stress response to vessel noises, but cases where no reaction or a positive reaction to vessel sounds exist [18]. In the few cases where vessel noises are specifically tested on invertebrates during their settlement stage, reactions have varied. While corals show settlement reduction, mussels and sea squirt larvae increase settlement [24, 54].

While there was no significant difference between the reef treatment and the vessel + reef treatment, there appears to be a trend of reduced settlement in the vessel + reef treatment. The effect size is potentially ecologically relevant, with larvae being 1.29 times less likely to settle when vessel noise is added to the oyster reef sound. Treatments of reef + vessel noises did not differentiate significantly from off-reef and no speaker controls. Although this study does not provide conclusive evidence of habitat sounds being masked by vessel noises, it highlights the need for further investigation in this area. While cases of anthropogenic masking in other invertebrate settlement experiments remains unconfirmed, a recent study by McAfee et al. (2022) [7] found that acoustic enriching experimental *Ostrea angasi* oyster reefs with reef sound was effective in low background noise areas, but ineffective in high background noise environments. However, the specific element of the background noise responsible for these results remains uncertain, as the term 'background noise' in the study encompassed all sounds within the soundscape (anthropogenic, geophysical, or biological). Looking past the biological response of overlaying vessels and reef sounds, acoustic characteristics of the reef sounds appeared to change with the addition of vessel noises in the recorded files. We observed a steep decrease in ACI, a mild decrease in ADI, and an increase in AEI, when vessel noises were added to reefs, although statistical confirmation is needed.

9.4.3 Tank experiment limitations and future work

Lab-based sound exposure experiments are key as a first approach to test certain hypotheses, as they can be very controlled. External factors can be isolated, background ambient sound can be removed, and controls can be established from the same batch at the same time. On the other hand, lab-based sound exposure experiments are incomplete in some aspects. For example, when presenting the whole picture of a soundscape. Reefs will change in their sonorific properties from a myriad of factors (time of day, time of year, etc). Vessels, as well, are not likely to produce sound continuously at one location, as experienced in the playbacks. Nevertheless, to understand the acoustic basis for sound discrimination and use as a cue, it is necessary to use defined and identifiable sources so robust and direct conclusions can be extracted. In future work, once the contribution of acoustic characteristics are better understood, longer exposure experiments in the field should be done with more realistic soundscapes and environmental conditions.

Tank experiments also pose technical challenges in keeping playback sounds true to their original field recordings for several reasons. First, aquatic invertebrates, including oysters, sense particle motion rather than sound pressure [18]. Particle motion currently remains challenging to quantify, especially in small tanks. In the field, sound pressure and particle motion levels are strongly correlated, but that is not the case when close to the sound source and reflecting and pressure-relieving surfaces. Therefore, in smaller spaces such as a tank, the sound propagation will not necessarily be related to particle motion because the walls and surface will act as pressure release surfaces [55]. Hence, sound pressure measurements can be a poor indicator of the particle motion levels, especially close to the tank walls. However, the magnitude and direction of the particle motion are expected to differ substantially from the one the larvae would experience in the field.

Second, cross-talk between tanks is possible, as seen in the results Section. In this study, this cross-talk happened mostly at frequencies below 200 Hz, probably due to vibration propagation instead of air propagation, and it was mostly present during loud periods from the neighboring tank. However, tank recordings retained the acoustic characteristics necessary to make them distinguishable, as proven by the obtained results and by the manual analysis of the tank recordings.

Last, in tank sound experiments, the sound field can present great variations at small spatial scales. For this reason, the received sound levels were measured at all the jars when the speaker was playing white noise, giving very similar results, so the levels received at all the jars were considered the same treatment. We acknowledge that if the material would be available, doing simultaneously the tank recordings in all the jars would be of increased value, but still because the larvae were in jars of 100 ml, they still could be exposed to different sound fields (jars had a diameter of 4.4cm and a length of 6.5cm). Still, to account for these possible differences, jar position was included as a possible effect in the GML model.

To make causal inference, all parameters should be kept constant throughout and experiment. In the present study, we could not adhere to this principle in two aspects. Firstly, we used a different sound file for each of the treatments in each of the trials. This approach has the advantage that we make inferences about the effect of sound types, rather than specific sound files. Using multiple sounds offers a more realistic insight to each treatment, which enhances the ability to extrapolate, as explained in Section 'Sound Treatments and Playback'. But it comes at the cost of limited power to detect a causal relationship as there is an additional confounding variation coming from the differences among sound files within sound types. Secondly, we had a limited number of speakers at our disposition. The trials were therefore conducted over several consecutive days and larvae could not be randomly assigned to the different trials and varied by age throughout the study. Our statistical analysis revealed that the date of the experiment had a significant effect on settlement. As the number of trials was limited, our design did not allow us to discriminate between the effect of experiment date and the effect of the variation among different sound files among treatment types. More and more extensive studies are needed to investigate which features of a specific sound type elicit the response in larvae settlement we observed in this study. These limitations do not pose a problem for the current experiment, as our target was a proof-of-concept study into whether the settlement rate presented any differences when exposed to different acoustic stimuli in lab conditions, and to do so a fully controlled environment is necessary. Furthermore, the fluctuations over time at a large enough time scale represented in the ACI values are probably not affected by the dynamics of the tank resonances. Hence this cue remains valid in this experiment. While it necessary for future research to confirm our results under field conditions, proof of concept lab studies such as these, are essential first steps, as in the field, it is currently possible to control added sound, but not possible to remove other background sounds.

9.4.4 Implications for reef restoration

Classic oyster reef mitigation and restoration projects focus on providing new hard substrates for wild larvae to settle, as well as supplying new adults to reefs, however, the importance of acoustic cues may be overlooked. Recently McAfee et al., (2022) [7] used underwater speakers to enhance soundscapes on constructed reefs resulting in greater initial settlement of the oyster *Ostrea angasi*. If similar acoustic preferences are established for other species, these same techniques could be employed elsewhere. Our results indicate that *M. gigas* also responds to acoustic cues, and thus may respond positively to acoustic enrichment as a restoration strategy.

9.4.5 Conclusion

We show that *M. gigas* larvae will settle more readily during playback of oyster reef sounds. The reef sounds were unique in being very acoustically diverse (high ACI), while other acoustic features, such as SPL varied among treatments. This suggests that oyster larvae may be able to detect complex spectro-temporal patterns in the soundscape rather than rely solely on SPL. Furthermore, we find that noise from vessels does not inhibit larvae settlement any more than the effect of off-reefs sounds or no speaker controls. We call for more research to replicate our findings in the laboratory in field experiments. More quantitative evidence is needed to determine if vessel noise (or other anthropogenic sounds) may affect oyster recruitment in ecologically realistic settings in the field.

Acknowledgements

Part of this project was funded by the Research Foundation Flanders (FWO) as part of the Belgian contribution to the LifeWatch ESFRI. We would like to thank the crew of the RV Simon Stevin for their help with acquiring the acoustic data. This study would not have been possible without the lending of the playback equipment from Leiden University. The authors would also like to thank Mattias Bossaer for his help with the lab set-up. Funding for the reef recordings in the Netherlands part of the Swimway Waddenzee project; financial contributions from the Waddenfund, the Ministry of Agriculture, Nature Management and Food Safety, and the Dutch provinces Noord-Holland, Groningen and Friesland.

Data availability statement

All data are available via the marine data archive (MDA) on `https://doi.org/10.14284/670`.

Scripts necessary to reproduce the results can be found in: `https://github.com/sschmidlin/larvae-and-sound`

References

[1] M. Hadfield and V. Paul. *Natural Chemical Cues for Settlement and Metamorphosis of Marine-Invertebrate Larvae*. Marine Chemical Ecology, June 2001.

[2] R. D. Burke. *Pheromonal Control of Metamorphosis in the Pacific Sand Dollar, Dendraster Excentricus*. Science, 225(4660):442–443, July 1984.

[3] C. M. Pearce and R. E. Scheibling. *Induction of Metamorphosis of Larvae of the Green Sea Urchin,* Strongylocentrotus Droebachiensis *, by Coralline Red Algae*. The Biological Bulletin, 179(3):304–311, December 1990.

[4] M. N. Tamburri, R. K. Zimmer-Faust, and M. L. Tamplin. *Natural Sources and Properties of Chemical Inducers Mediating Settlement of Oyster Larvae: A Re-examination*. The Biological Bulletin, 183(2):327–338, October 1992.

[5] H. E. Vasquez, K. Hashimoto, A. Yoshida, K. Hara, C. C. Imai, H. Kitamura, and C. G. Satuito. *A Glycoprotein in Shells of Conspecifics Induces Larval Settlement of the Pacific Oyster Crassostrea Gigas*. PLOS ONE, 8(12):e82358, December 2013.

[6] K. B. Johnson. *Laboratory Settlement of the Eastern Oyster Crassostrea Virginica Influenced by Substratum Concavity, Orientation, and Tertiary Arrangement*. Journal of Shellfish Research, 36(2):315–324, August 2017.

[7] D. McAfee, B. R. Williams, L. McLeod, A. Reuter, Z. Wheaton, and S. D. Connell. *Soundscape Enrichment Enhances Recruitment and Habitat Building on New Oyster Reef Restorations*. Journal of Applied Ecology, 60(1):111–120, October 2022.

[8] M. J. A. Vermeij, K. L. Marhaver, C. M. Huijbers, I. Nagelkerken, and S. D. Simpson. *Coral Larvae Move toward Reef Sounds*. PLOS ONE, 5(5):e10660, May 2010.

[9] B. R. Williams, D. McAfee, and S. D. Connell. *Oyster Larvae Swim along Gradients of Sound*. Journal of Applied Ecology, 59(7):1815–1824, 2022.

[10] M. C. P. Amorim, R. O. Vasconcelos, and P. J. Fonseca. *Fish Sounds and Mate Choice*. In F. Ladich, editor, Sound Communication in Fishes, Animal Signals and Communication, pages 1–33. Springer, Vienna, 2015.

[11] A. R. Hughes, D. A. Mann, and D. L. Kimbro. *Predatory Fish Sounds Can Alter Crab Foraging Behaviour and Influence Bivalve Abundance*. Proceedings of the Royal Society B: Biological Sciences, 281(1788):20140715, August 2014.

[12] C. M. Duarte, L. Chapuis, S. P. Collin, D. P. Costa, R. P. Devassy, V. M. Eguiluz, C. Erbe, T. A. C. Gordon, B. S. Halpern, H. R. Harding, M. N. Havlik, M. Meekan, N. D. Merchant, J. L. Miksis-Olds, M. Parsons, M. Predragovic, A. N. Radford, C. A. Radford, S. D. Simpson, H. Slabbekoorn, E. Staaterman, I. C. Van Opzeeland, J. Winderen, X. Zhang, and F. Juanes. *The Soundscape of the Anthropocene Ocean*. Science, 371(6529):eaba4658, February 2021.

[13] P. J. O. Miller, S. Isojunno, E. Siegal, F.-P. A. Lam, P. H. Kvadsheim, and C. Curé. *Behavioral Responses to Predatory Sounds Predict Sensitivity of Cetaceans to Anthropogenic Noise within a Soundscape of Fear*. Proceedings of the National Academy of Sciences, 119(13):e2114932119, March 2022.

[14] A. Lillis, D. Eggleston, and D. Bohnenstiehl. *Estuarine Soundscapes: Distinct Acoustic Characteristics of Oyster Reefs Compared to Soft-Bottom Habitats*. Marine Ecology Progress Series, 505:1–17, May 2014.

[15] J. A. Stanley, C. A. Radford, and A. G. Jeffs. *Induction of Settlement in Crab Megalopae by Ambient Underwater Reef Sound*. BEHAVIORAL ECOLOGY, 21(1):113–120, 2010-01/2010-02.

[16] S. Wilkens, J. Stanley, and A. Jeffs. *Induction of Settlement in Mussel (Perna Canaliculus) Larvae by Vessel Noise*. Biofouling, 28(1):65–72, January 2012.

[17] M. K. Pine, A. G. Jeffs, and C. A. Radford. *Turbine Sound May Influence the Metamorphosis Behaviour of Estuarine Crab Megalopae*. PLoS ONE, 7(12):e51790, December 2012.

[18] M. Solé, K. Kaifu, T. A. Mooney, S. L. Nedelec, F. Olivier, A. N. Radford, M. Vazzana, M. A. Wale, J. M. Semmens, S. D. Simpson, G. Buscaino, A. Hawkins, N. Aguilar de Soto, T. Akamatsu, L. Chauvaud, R. D. Day, Q. Fitzgibbon, R. D. McCauley, and M. André. *Marine Invertebrates and Noise*. Frontiers in Marine Science, 10, 2023.

[19] A. Lillis, D. Bohnenstiehl, J. W. Peters, and D. Eggleston. *Variation in Habitat Soundscape Characteristics Influences Settlement of a Reef-Building Coral*. PeerJ, 4:e2557, October 2016.

[20] S. Holles, S. Simpson, A. Radford, L. Berten, and D. Lecchini. *Boat Noise Disrupts Orientation Behaviour in a Coral Reef Fish*. Marine Ecology Progress Series, 485:295–300, June 2013.

[21] L. Wilson, R. Constantine, M. K. Pine, A. Farcas, and C. A. Radford. *Impact of Small Boat Sound on the Listening Space of Pempheris Adspersa, Forsterygion Lapillum, Alpheus Richardsoni and Ovalipes Catharus*. Scientific Reports, 13(1):7007, April 2023.

[22] D. Lecchini, F. Bertucci, C. Gache, A. Khalife, M. Besson, N. Roux, C. Berthe, S. Singh, E. Parmentier, M. M. Nugues, R. M. Brooker, D. L. Dixson, and L. Hédouin. *Boat Noise Prevents Soundscape-Based Habitat Selection by Coral Planulae*. Scientific Reports, 8(1):9283, June 2018.

[23] J. Stocks. *Response of Marine Invertebrate Larvae to Natural and Anthropogenic Sound: A Pilot Study*. The Open Marine Biology Journal, 6(1):57–61, June 2012.

[24] A. Jolivet, R. Tremblay, F. Olivier, C. Gervaise, R. Sonier, B. Genard, and L. Chauvaud. *Validation of Trophic and Anthropic Underwater Noise as Settlement Trigger in Blue Mussels*. Scientific Reports, 6(October 2015):1–8, 2016.

[25] M. Gigot, F. Olivier, G. Cervello, R. Tremblay, D. Mathias, T. Meziane, L. Chauvaud, and J. Bonnel. *Pile Driving and Drilling Underwater Sounds Impact the Metamorphosis Dynamics of Pecten Maximus (L., 1758) Larvae*. Marine Pollution Bulletin, 191:114969, June 2023.

[26] C. Erbe. *Critical Ratios of Beluga Whales (* Delphinapterus Leucas *) and Masked Signal Duration*. The Journal of the Acoustical Society of America, 124(4):2216–2223, October 2008.

[27] S. Li, H. Wu, Y. Xu, C. Peng, L. Fang, M. Lin, L. Xing, and P. Zhang. *Mid- to High-Frequency Noise from High-Speed Boats and Its Potential Impacts on Humpback Dolphins*. The Journal of the Acoustical Society of America, 138(2):942–952, August 2015.

[28] B. Walles, K. Troost, D. van den Ende, S. Nieuwhof, A. C. Smaal, and T. Ysebaert. *From Artificial Structures to Self-Sustaining Oyster Reefs*. Journal of Sea Research, 108:1–9, February 2016.

[29] R. Botta, F. Asche, J. S. Borsum, and E. V. Camp. *A Review of Global Oyster Aquaculture Production and Consumption*. Marine Policy, 117:103952, July 2020.

[30] D. McAfee and S. D. Connell. *The Global Fall and Rise of Oyster Reefs*. Frontiers in Ecology and the Environment, 19(2):118–125, 2021.

[31] J. L. Ruesink, H. S. Lenihan, A. C. Trimble, K. W. Heiman, F. Micheli, J. E. Byers, and M. C. Kay. *Introduction of Non-Native Oysters: Ecosystem Effects and Restoration Implications*. Annual Review of Ecology, Evolution, and Systematics, 36(1):643–689, 2005.

[32] M. W. Beck, R. D. Brumbaugh, L. Airoldi, A. Carranza, L. D. Coen, C. Crawford, O. Defeo, G. J. Edgar, B. Hancock, M. C. Kay, H. S. Lenihan, M. W. Luckenbach, C. L. Toropova, G. Zhang, and X. Guo. *Oyster Reefs at Risk and Recommendations for Conservation, Restoration, and Management*. BioScience, 61(2):107–116, February 2011.

[33] J. Pineda, F. Porri, V. Starczak, and J. Blythe. *Causes of Decoupling between Larval Supply and Settlement and Consequences for Understanding Recruitment and Population Connectivity*. Journal of Experimental Marine Biology and Ecology, 392(1-2):9–21, August 2010.

[34] M. A. R. Koehl and T. Cooper. *Swimming in an Unsteady World*. Integrative and Comparative Biology, 55(4):683–697, October 2015.

[35] M. Charifi, M. Sow, P. Ciret, S. Benomar, and J.-C. Massabuau. *The Sense of Hearing in the Pacific Oyster, Magallana Gigas*. PLOS ONE, 12(10):e0185353, October 2017.

[36] A. Lillis, D. B. Eggleston, and D. R. Bohnenstiehl. *Oyster Larvae Settle in Response to Habitat-Associated Underwater Sounds*. PLOS ONE, 8(10):e79337, 30 d'oct. 2013.

[37] Team A. Audacity. *Team A. Audacity(R): Free Audio Editor and Recorder [Computer Application];1999-2019.*, 2023.

[38] J. Hubert, J. J. van Bemmelen, and H. Slabbekoorn. *No Negative Effects of Boat Sound Playbacks on Olfactory-Mediated Food Finding Behaviour of Shore Crabs in a T-maze*. Environmental Pollution, 270:116184, February 2021.

[39] H. Slabbekoorn and N. Bouton. *Soundscape Orientation: A New Field in Need of Sound Investigation*. Animal Behaviour, 76(4):e5–e8, October 2008.

[40] H. Slabbekoorn. *Songs of the City: Noise-Dependent Spectral Plasticity in the Acoustic Phenotype of Urban Birds*. Animal Behaviour, 85(5):1089–1099, May 2013.

[41] A. Lillis, D. R. Bohnenstiehl, and D. B. Eggleston. *Soundscape Manipulation Enhances Larval Recruitment of a Reef-Building Mollusk*. PeerJ, 3:e999, June 2015.

[42] J. Hubert, E. Booms, R. Witbaard, and H. Slabbekoorn. *Responsiveness and Habituation to Repeated Sound Exposures and Pulse Trains in Blue Mussels*. Journal of Experimental Marine Biology and Ecology, 547:151668, February 2022.

[43] J. S. Ulloa, S. Haupert, J. F. Latorre, T. Aubin, and J. Sueur. *Scikit-Maad: An Open-Source and Modular Toolbox for Quantitative Soundscape Analysis in Python*. Methods in Ecology and Evolution, 12(12):2334–2340, 2021.

[44] P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J. van der Walt, M. Brett, J. Wilson, K. J. Millman, N. Mayorov, A. R. J. Nelson, E. Jones, R. Kern, E. Larson, C. J. Carey, İ. Polat, Y. Feng, E. W. Moore, J. VanderPlas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E. A. Quintero, C. R. Harris, A. M. Archibald, A. H. Ribeiro, F. Pedregosa, and P. van Mulbregt. *SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python*. Nature Methods, 17(3):261–272, March 2020.

[45] T. Bradfer-Lawrence, N. Gardner, L. Bunnefeld, N. Bunnefeld, S. G. Willis, and D. H. Dent. *Guidelines for the Use of Acoustic Indices in Environmental Research*. Methods in Ecology and Evolution, 10(10):1796–1807, 2019.

[46] C. Mammides, E. Goodale, S. K. Dayananda, K. Luo, and J. Chen. *On the Use of the Acoustic Evenness Index to Monitor Biodiversity: A Comment on "Rapid Assessment of Avian Species Richness and Abundance Using Acoustic Indices" by Bradfer-Lawrence et al. (2020) [Ecological Indicators, 115, 106400]*. Ecological Indicators, 126:107626, July 2021.

[47] M. M. Helm, N. Bourne, A. c. e. . Lovatelli, and Fisheries and Aquaculture Management Division. *The Hatchery Culture of Bivalves: A Practical Manual*. Number 0429-9345|20-70-7010 in FAO Fisheries Technical Paper. FAO, Rome, Italy, 2004.

[48] D. Bates, M. Mächler, B. Bolker, and S. Walker. *Fitting Linear Mixed-Effects Models Using Lme4*. Journal of Statistical Software, 67:1–48, October 2015.

[49] R Core Team (2023). *R: A Language and Environment for Statistical Computing*.

[50] R. V. Lenth. *Least-Squares Means: The R Package Lsmeans*. Journal of Statistical Software, 69:1–33, January 2016.

[51] J. C. Montgomery, A. Jeffs, S. D. Simpson, M. Meekan, and C. Tindle. *Sound as an Orientation Cue for the Pelagic Larvae of Reef Fishes and Decapod Crustaceans*. In Advances in Marine Biology, volume 51, pages 143–196. Academic Press, January 2006.

[52] C. A. Radford, J. A. Stanley, S. D. Simpson, and A. G. Jeffs. *Juvenile Coral Reef Fish Use Sound to Locate Habitats*. Coral Reefs, 30(2):295–305, June 2011.

[53] K. A. Murchy, H. Davies, H. Shafer, K. Cox, K. Nikolich, and F. Juanes. *Impacts of Noise on the Behavior and Physiology of Marine Invertebrates: A Meta-Analysis*. Proceedings of Meetings on Acoustics, 37(1):040002, March 2020.

[54] J. McDonald, S. Wilkens, J. Stanley, and A. Jeffs. *Vessel Generator Noise as a Settlement Cue for Marine Biofouling Species*. Biofouling, 30(6):741–749, July 2014.

[55] P. H. Rogers, A. D. Hawkins, A. N. Popper, R. R. Fay, and M. D. Gray. *Parvulescu Revisited: Small Tank Acoustics for Bioacousticians*. In A. N. Popper and A. Hawkins, editors, The Effects of Noise on Aquatic Life II, Advances in Experimental Medicine and Biology, pages 933–941, New York, NY, 2016. Springer.

Part V

General Discussion

10
Discussion

10.1 Overview of main contributions

This thesis is centered on developing new methods to analyze shallow marine soundscapes while also examining the outcomes of applying these methods and comparing them to more traditional approaches. The developed methodologies are tailored to analyze the complex acoustic environments in shallow waters. The scope of this thesis goes beyond theoretical frameworks by implementing these methods in data collected in the BPNS, a shallow coastal area, and evaluates the performance and effectiveness of the developed methods through empirical studies. By analyzing the BPNS soundscape, we also offer insights into the acoustic characteristics and dynamics of this area. Finally we also assess in one lab experiment how do oyster larvae perceive this soundscape.

First, this PhD dissertation aimed to investigate different ways to observe which were the main contributors and drivers of the BPNS soundscape. In Part II we demonstrate that machine learning techniques offer deeper insights into these drivers compared to more traditional statistical analyses. The soundscapes within the BPNS cluster in distinct groups, and SHAP is an efficient technique to evaluate the driving factors behind these cluster differences. Combining insights from both traditional and novel methods, we show that the BPNS soundscape is primarily influenced by anthropogenic sounds, as well as current- and wind-related sounds. Furthermore, we find that the soundscape varies across locations based on factors such as water depth, distance to shipping lanes, and proximity to the coast. A summary of all the methods presented in this part and their most relevant features and characteristics can be found in Table 10.1.

Second, we wanted to know if it was possible to disentangle previously not documented biological sound events from long-term recordings. To do so, we propose several Machine Learning approaches that can accelerate the process of detecting and characterizing possible sounds of interest. Results of Chapters 7 and 8 prove that the proposed methodologies are useful tools to analyze the possible biological sounds in the BPNS. Furthermore, we prove that the BPNS soundscape also contains animal vocalizations despite all anthropogenic noise, and therefore there is biological information to be extracted from quantifying these sound events. A summary of the methods used in Part III and their most relevant features and characteristics can be found in Table 10.2.

Both during manual annotations (Chapter 8) and unsupervised detection and clustering (Chapter 7), we have found several interesting sounds in the recordings that we could not directly identify but which appeared repeatedly. Some of these sounds are very similar to fish sounds described in the literature and the library FishSounds [1], and others similar to the crustacean sounds from Coquereau [2]. We also identified seal sounds and dolphin whistles in several locations. Therefore, we can confidently say that the long-term recordings contain biological sounds. These are important findings that support the importance of acoustically monitoring the North Sea. One thing to consider is that fish sounds could have been masked by flow noise or shipping noise, as they typically produce sounds with a fundamental frequency below 1000 Hz. Therefore, some sounds at lower frequencies might have been masked. However, the fish sounds we found are present at higher frequencies than flow noise (100 to 500 Hz). Crustacean sounds are at higher frequencies usually (> 2 kHz), so they should not be masked by flow noise or shipping noise. As we have seen in this thesis, acoustically monitoring the biophony in the BPNS is a challenging but possible task. However, PAM remains as a valid and non-invasive technique.

Lastly, we aimed to assess if soundscapes are important to local fauna, and if the acoustic differences between locations are sensed by the animals living there. To answer this question we exposed oyster larvae to different soundscapes recorded in the BPNS and the Wadden Sea. The recordings of oyster reefs had to be from the Wadden Sea as all the oyster reefs in the BPNS have been destroyed in the last decades. The results show that oyster larvae increase their settlement in the presence of oyster reef sounds compared to sandy areas or vessel sounds, and that loud vessels might mask this effect.

Table 10.1 Summary of the methods and parameters used for each chapter of Part II. Ch. stands for chapter. HMB stands for hybrid millidecade bands. RF for Random Forest.

	Holistic analysis				
Ch.	**Data**	**Input features**	**Window duration**	**Dimension reduction**	**ML technique**
4	BPNS Stationary	HMB	1 min	NA	No ML
5	BPNS Drift	1/3-octave bands	5 sec	UMAP	1. Unsupervised clustering (DBSCAN) 2. Supervised RF 3. SHAP
	BPNS Stationary	HMB	1 min		

Table 10.2 Summary of the methods and parameters used for each chapter of Part III and IV . Ch. stands for chapter. Spec. stands for spectrogram.

	Acoustic Event Detection							
	Event detection				Event analysis			
Ch.	**Data**	**Input features**	**Window duration**	**ML technique**	**Input features**	**Feature extraction**	**Dim. reduction**	**ML technique**
6	Miller dataset	Spec.	30 sec	Supervised DL object detector (YOLOv8)	NA	NA	NA	NA
7	BPNS Stationary + Freshwater	Spec.	20 sec	Supervised DL object detector (YOLOv8)	Mel spec.	BioLingual	UMAP	Clustering (HDBSCAN)
8	BPNS Stationary	NA	NA	Manual annotation	Mel spec.	CAE	sPCA	Clustering (HDBSCAN)
					Waveform	AVES		

10.2 Reporting, comparing and understanding underwater soundscapes

Underwater soundscapes have cultural, ecological and economical value. During this thesis I have seen how underwater recordings can be used to increase Ocean Literacy and environmental awareness. As many others have proven, and as we have seen in Chapter 9, soundscapes also have an ecological value. Disturbing them can cause loss of habitats, or can reduce the chances of recruitment for restoration plans. Therefore, they also have an economical value for fisheries and coastal protection. For this reason, it is necessary to understand and report these soundscapes in a quantifiable way for policymakers. For this, understanding the ecological role of soundscapes is key. This can be done with sound exposure experiments, both in the lab or in the field, or by analyzing correlations between soundscapes and animal behavior. Sound exposure experiments are necessary to be able to assess the impact of soundscape changes in marine life. These changes can be due to noise pollution generated by human activities at sea [3], or by the changes in soundscape driven by changes in the ecosystem itself. For example, reefs become silent when their biodiversity declines [4, 5]. Sound exposure studies need to represent realistic acoustic scenarios that the animals can encounter at sea. For this reason, in the experiment from Chapter 9 we exposed larvae to playbacks of field recordings. When doing this, the observed effects can be more generalized to the treatment concept ('reef' or 'vessels') rather than being attributed solely to a specific sound. This is of special importance in areas such as the BPNS, where soundscapes change on such a small spatial scale. The results obtained in Chapter 9 point out that acoustic complexity could be masked by anthropogenic sound sources. This masking could potentially have an effect on larvae settlement, which is a clear example of the ecological effects that human activities can have if they alter natural soundscapes. The results of our experiment from Chapter 9 also point out the complexity of soundscapes, and how animals use sound in a complex manner to obtain environmental information. There is, therefore, information about the environment encoded in the acoustic footprint of a habitat. Further work is necessary to find ways to report this information from an animal's perspective. A better understanding of the features responsible for larvae (or other species) attraction to reefs' soundscapes (or other habitats) can lead to more effective mitigation and restoration measures.

In order to compare soundscapes between regions, understand changes over time, and to have concrete numbers for policy makers, quantifiable and measurable metrics are needed. However, most of the traditional soundscape characterization methods common in the literature focusing on holistically characterizing soundscapes, although very informative, are mostly done using descriptive rather than quantitative methods [6, 7]. For this reason the MSFD [8] and the Ocean Sound

Essential Ocean Variable Implementation Plan [9] primarily aim to extract quantitative data from soundscape recordings and translate it into policy recommendations. The focus of these variables lies on quantifying noise pollution, and they do not address acoustic diversity or other components of the soundscape.

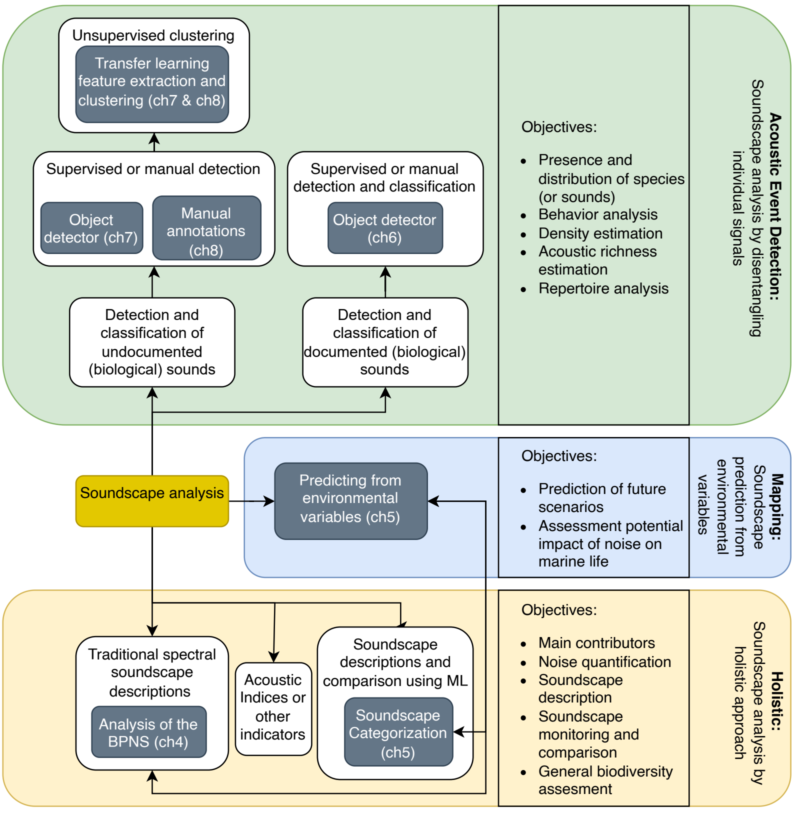
Anthropogenic noise pollution at sea needs to be monitored and reported, both to make sure regulations are followed and to understand the disturbance status of a habitat so it can be protected when necessary. The Good Environmental Status (GES) descriptor number 11 from the European MSFD is described as the following: "Introduction of energy, including underwater noise, is at levels that do not adversely affect the marine environment" [10]. One of this possible adverse effects is the masking of acoustic cues, as seen in Chapter 9, but sound can also have a physical impact on marine fauna and flora [3, 11]. This year (2024) there was a new release on the MSFD regulations affecting the GES 11 indicator, where limits were specified. For impulsive sound: "For short-term exposure (1 day, i.e., daily exposure), the maximum proportion of an assessment/habitat area utilised by a species of interest that is accepted to be exposed to impulsive noise levels higher than the Level of Onset of Biologically adverse Effects (LOBE), over 1 day, is 20 % or lower ($\leq$ 20 %). For long-term exposure (1 year), the average exposure is calculated. The maximum proportion of an assessment/habitat area utilised by a species of interest that is accepted to be exposed to impulsive noise levels higher than LOBE, over 1 year on average, is 10 % or lower ($\leq$ 10 %)". For continuous sound: "20 % of the target species habitat having noise levels above LOBE not to be exceeded in any month of the assessment year, in agreement with the conservation objective of the 80 % of the carrying capacity/habitat size" [12]. Setting these limits is crucial for marine environmental protection.

On the other hand, finding a single acoustic indicator that robustly quantifies biodiversity or ecosystem health across ecosystems remains desirable to using PAM as a monitoring tool. There has been some effort into unifying the most common metrics used for soundscape description into one measure, such as the Soundscape Code proposed by Wilford et al. (2021) [13], or the one proposed by Luypaert et al. (2022) for terrestrial ecosystems [14]. However, even though these methods can unify the acoustical description, they provide little understanding on the drivers of the acoustic properties.

There does not seem to be any set of features common across marine ecosystems that can directly provide information about the biodiversity, but it has been proven that soundscape change consistently indicates community change [15]. Even though this dissertation is a step towards finding reliable and robust metrics to characterize ecosystems in an ecologically meaningful, we did not find a feature set or a specific metric robust across ecosystems and directly linked to biodiversity. Some published approaches have been successful for certain environments, but the results are currently not global [15]. Therefore, further work on universal marine soundscape descriptors remains a challenge for the marine passive analysis community.

10.3 Automating underwater soundscape analysis

Figure 10.1 Overview of the aspects of soundscape analysis. The contributions by this dissertation to the applications are indicated by dark boxes. Chapter numbers are specified between brackets. For more detailed information of the methods, features and parameters used on each chapter see Table 10.1 and 10.2.



The objectives and research questions determine the pathway for soundscape analysis (see Figure 10.1). Holistic analysis gives a better idea of the entire ecosystem change and main soundscape contributors, and it can be used to assess seasonal and long-term changes. On the other hand, Acoustic Event Detection is more appropriate to identify presence, density, or behavior of species, and to quantify acoustic (and eventually biodiversity) richness.

To do holistic soundscape analysis, one can use traditional spectral analysis, novel machine learning approaches or other acoustic indicators (such as acoustic indices). Each of these have different advantages and disadvantages, and each of them is more suited to address different tasks. First, traditional spectral soundscape description is an interesting tool to use as a first analysis of a certain location. This analysis can be utilized to observe trends in noise over time and identify the primary temporal and spatial patterns. If information about potential sound sources is available, this approach can also be used to identify their occurrences, provided the sounds are loud and sustained over time. Examples of such sounds include ice calving sounds, fish choruses, or migrating whales. This type of analysis can also be used to describe and quantify the main drivers of the soundscape, including acoustic anthropogenic pollution, provided one has access to environmental data. These analyses primarily focus on identifying linear relationships. Within these traditional methods, the techniques include, among others SPD, LTSA, or spectral correlations. Furthermore, as done in Chapter 4, analyzing the difference between median and mean values can help identify periods when short transient sounds are present, which can be a very straightforward way to find exactly when certain sources are occurring once it is known which sounds to expect. These methods can also be used to model the sound generated by certain phenomena (e.g. wind or rain) with mathematical approaches. This knowledge can then be used to forecast some components of the soundscape given a set of environmental variables.

Within holistic soundscape analysis, a common proposal in terrestrial environments is to use ecoacoustic indices (or simply acoustic indices). However, as mentioned by Sethi et al. (2022) [15] for terrestrial environments, no soundscape analysis technique has not been proven to be consistently linked to species richness across all the ecosystems. There is increasing evidence that this is especially the case in the marine realm. These acoustic indices were initially designed to describe acoustic diversity in terrestrial environments. They have then been directly applied to the marine realm [16], obtaining mixed and contradictory results. Some studies have found some of the common ecoacoustic indices to be correlated with biodiversity in temperate reefs [17] or abundance and species richness in coral reefs [18]. Others could use these indices to asses daily patterns of fish choruses [19]. Latest reviews on the topic point out that the field of marine ecoacoustic indices still needs to be explored and that indices can only be used with caution [20, 16, 21], as no robust conclusions have been extracted. These indices seem to be useful in marine environments to quantify and report anthropophony, geophony, high-frequency biological sounds, or biological mass acoustic events (choruses). However, Raick et al. (2023) [22] concluded that the acoustic indices currently in use cannot be used as a proxy for fish diversity. Their experiments demonstrated that the indices were correlated with the abundance of sound events but not with the diversity of sound types. They argue that these indices are suited to monitor acoustic mass phenomena such as choruses, but that these choruses are not an appropriate indicator of diversity. A single ecoacoustic index might not be an indicator of species composition, richness or biodiversity robust across ecosystems, but a compound of several acoustic indices have successfully been used as acoustic features to feed ML models [23, 24]. Because of the lack of reliability of acoustic indices in marine environments, the existing acoustic indices were not considered in this dissertation to assess the soundscapes.

On the other hand, the use of ML tools can provide further insight in significant acoustic differences across soundscapes, and can reveal structures that might not be immediately evident through traditional methods. For some of these analysis it is not necessary to have environmental data nor previous knowledge on the expected contributors (even though it will always provide more information). Unlike traditional methods, ML methods such as the one described in Chapter 5 can deal with non-linear relationships between the drivers and the received sound. Using ML techniques which rely on clustering will not quantify the main contributors if those factors are not significantly driving the differences between clusters. Additionally, it may not offer insights into specific transient events when analyzing averaged long-term data. Within these ML holistic approaches, published techniques range from unsupervised separation of sources [25, 26, 27, 28], clustering [29, 30, 31, 32], and ecosystem characterization [23, 24]. All these proposals have in common that the obtained clusters needed to be manually analyzed to make sense of them. The method we propose in this thesis for holistic characterization builds forward on methods such as the one from Phillips et al. (2018) [33] for terrestrial ecosystems, but provides a novel way of analyzing the data instead of having to manually label and describe the clusters.

In conclusion, both ML and traditional statistical methods offer unique strengths and limitations for holistic analysis, and therefore are complementary. Traditional soundscape analysis provides a way to quantify the absolute sound contributors (if environmental data are available) and analyze long-term change, while categorization focuses more on differences between soundscape types. Choosing one or the other will depend on the research question and the available data.

When using acoustic events for soundscape analysis, there is always a detection and a classification part, which can be done separately or in one step. Either of the two parts can be entirely or partially composed of human input. When human input is used for the detection step, the classification step is usually done at the same time (manually) if the sounds are known; otherwise one needs to look twice at the data. The advantage of manual annotation lies in the obtained precision, while the use of ML detection and classification increase the amount of data that can be processed. Up to date, no perfect detector/classifier has been developed, and therefore using the output of the model without manual validation can lead to over or under estimation. It is however still a question if such a perfect model is possible - given the inherent disagreement between human annotators [34, 35] - or necessary. For this reason, very often pipelines combining manual input and ML models are found in literature. For known sounds, sometimes ML models are used to detect (and sometimes also classify) signals of interest and then human annotators validate or classify the output of the model (Figure 10.1). These models are usually designed to have a high recall at an expense of a high false positive rate, which will eventually be corrected by the human in the loop. Both for automatic detection and classification of known sounds, the most common approach is to train supervised models [36, 37]. This requires a big initial human effort to create the ground truth that is needed to train and evaluate models. Some open access datasets exist for some marine mammals, but they are often limited to a certain area and call types.

When working with unknown sounds, or sounds which are known to come from different species but that humans cannot tell apart (e.g. various dolphin whistles), it is also common to do manual annotation for the detection of events and then some kind of unsupervised clustering for the classification, such as we did in Chapter 8 (see Figure 10.1). In Chapter 7 we proposed a double automation of the 2-step pipeline, where we first trained a model to detect "any possible sound of interest" and then we clustered them in an unsupervised way. This is useful when one does not know which sound events to expect. The obtained detections can then be used to describe the soundscape. This is commonly done in other studies with known sound sources (mammals: [38, 39, 40], fish: [41, 42], invertebrates: [43]), but the methodology presented in Chapter 7 is the first time this has been proposed to explore unknown sounds in a broadband context.

The clustering part of our proposals (both for holistic and acoustic event analysis) is a technique which has already been applied by other studies to marine soundscapes, and can be very versatile. Some of these studies used a set of manually selected sound events [30, 32] to segregate animal vocalizations. Ozanich et al. (2021) [31], and Lin et al. (2018, 2021) [44, 25] also combined the clustering with automatic detection of possible events on long-term recordings. Ozanich et al. (2021) [31] used a mathematical detector to find possible events. Lin et al. (2018) [44] used periodicity-coded nonnegative matrix factorization as a source separation technique to denoise the recordings and enhance periodic signals such as fish choruses. Then they clustered the power spectrum of each enhanced 5-minute recording. Lin et al. (2021) [25] used simultaneously the same approach for source separation and the difference between mean and median LTSA to detect transient sounds.

Last, ML, and especially DL, are often regarded as "black boxes", as it is complicated to understand the underlying processes these models use to make a decision. In some cases, this information can be as valuable as the model itself. For this we recommend the use of Explainable Machine Learning tools, such as SHAP [45]. SHAP, which stands for SHapley Additive exPlanations, is a method used to quantify the impact of a feature on the value of the target variable. The total contribution of all the features equates to the value of the target variable for a given set of feature values (i.e., a particular record). The average absolute value of a feature's impact on the target variable serves as a measure of its importance. Using SHAP, as proposed in Chapter 5, supervised models can be described even when presenting non-linear relationships with the parameters causing the differences between them. However, SHAP is not ideal for CNNs, as it analyzes the importance of each feature. To understand CNNs, a recommendation would be to plot saliency maps. These maps color the input images (in this case, probably spectrograms) according to focus or "importance" given to that particular area of the image by the CNN. There are several open-source software to do that, Captum being the one recommended by PyTorch [46]. In conclusion, explainable ML tools can provide information about the model's decision-making process, which can be critical to understand its possible flaws and how to improve it.

10.3.1 Lessons learned and recommendations when using ML for soundscape analysis

The biggest lesson learned when developing ML methods for soundscape analysis is that anything which has been developed for data from another location or instrument might not work on your data. In ML, this is known as domain adaptation [47, 48]. Domain adaptation is the capability of applying an algorithm, which has been trained on one or more "source domains", to a different yet related "target domain". This is particularly applicable to underwater acoustics because of the differences between recorders, between propagation patterns in different locations, the different sounds present depending on the region, and the types of ambient noise [48]. Therefore, there is a different data distribution between the two datasets. This is known as domain shift. A first recommendation when applying an existing model trained on other data, would be to test it extensively on the new data first. This entails human effort to validate the outcome of the model, and is sometimes not possible because of the high human effort required.

Sometimes a model might only be used in one particular location. This is a perfectly realistic scenario where one would train a model on old data from that same location and afterwards use the model to analyze the new data which is being collected. In this case, one does not need to worry about domain adaptation, but the objective should be made clear if the model is published. However, it is common to publish models which are meant to be general and applicable to a large area but are only tested on one dataset or location. For this reason, in Schall et. al (2024) [49] we explained why models published with such purpose always need to be tested in at least one location which has not been used during training (the more, the better).

When evaluating the performance of models it is necessary that the evaluation metrics are representative for the study's objective. For example, if the output of a detector will always be manually validated, recall and false alarm rates are good metrics. If one is evaluating the performance of a classifier which is applied to the output of a detector, accuracy or precision might be good representations of the performance. The final performance, however, should also be reported for the two-step process, not only for the classification. Furthermore, the test dataset must be representative of the objective of the study. If a detector will be used to detect calls on a long-term dataset, the test set should include these long-term recordings instead of an artificially balanced dataset between calls and noise [49].

When training a new model, one recommendation would be to first try existing models, especially the ones pre-trained on large (bio)acosutics dataset(s). Computer vision and speech recognition are very advanced fields of Machine Learning, and underwater acoustics can benefit from both, as most bioacousticians use spectrograms to recognize animal vocalizations. Furthermore, fine-tuning or transfer learning are powerful tools in the particular case of underwater bioacoustics, where available ground truth is limited. This is applicable for event detection, event classification and feature extraction. Computer vision algorithms are suitable for some of these tasks, although they might miss some acoustical information such as the concept of frequency. The pre-trained models can be fine-tuned on the target dataset. Another option would be to use the last embedding layer from the pre-trained model as input to a small and shallow network, and train it on the target dataset.

We would recommend the use of clustering as a tool to discover new sound events and patterns in a dataset. For this task other studies have proposed the use of different features, including autoencoders [30, 31] or directly flattened spectrograms inputted to a dimension reduction algorithm [32]. If the discriminating features are known between the sounds which one expects to find, then deep embeddings are not necessary. As this is usually not the case when dealing with unknown sounds, deep embeddings from pre-trained models are recommended for this particular case. When using deep embeddings, it is necessary to consider different models will perform differently on different datatasets and tasks, as they focus on different aspects of the audio. For our particular case of detecting any possible sound of interest in a broadband context, with a very noisy background, and including sounds in very different frequency bands, BioLingual was found to work the best. AVES was the second one, and last was a re-trained CAE [1]. If the used feature set is large ($\gtrsim$ 20 features), we highly recommend the use of a dimension reduction algorithm before applying the clustering. This can be sPCA when the pipeline

[1]BioLingual was not considered in Chapter 8 because we did not know of its existence when writing that manuscript (despite the order in this thesis, Chapter 8 was written before Chapter 7).

needs to be applied constantly to new data, as this will give a clear idea of which features are being selected and can be traced back. If that is not the case, and the pipeline does not need to constantly be re-applied when adding new data, we would recommend the use of UMAP, as it has been successfully applied in several bioacoustics studies [30, 32, 50] (including Chapters 5 and 7). Furthermore, if the frequency limits of each annotation are available, we would recommend applying a band-pass filter before extracting the DL features, especially when analyzing broadband sound. This is only possible when using manual annotations or object detectors which predict bounding boxes. Both in Chapter 7 and Chapter 8 we found that this approach led to cleaner clusters, but in other cases other noise reduction algorithms might perform better.

The first idea for Chapter 7 was to fine tune the model trained on the Antarctic data to detect sounds in the BPNS. However, the transfer learning did not perform very well. Therefore, we tried fine-tuning on YOLOv8 pre-trained on the COCO dataset, and this seemed to perform better. This seems to indicate that the domain shift for the BPNS is larger for Antarctic low-frequency data than for regular casual images (COCO dataset). In any case, the final results show that, as expected, the performance of the model in the Antarctic dataset is better than in the BPNS dataset. This could be due to the difference in data sizes. For the Miller model, the training set consisted of 67,491 images, while for the BPNS dataset, when using all the available data the training set consisted of 6,712 images. It could also be due to the difference in background noise, as vessel sounds are not as present in the Southern Ocean as in the Northern Hemisphere [51]. Finally, this difference could also be because the whale calls present a typified, well-defined, and a fairly constant, spectral shape. In contrast with the subjective description of acoustic salient events, the image patterns of these calls are a lot easier to detect, both for human analysts and models.

Finally, when employing transfer learning from models trained on images, it may be necessary to convert spectrograms from grayscale to color if the original model was trained on RGB(A) images. This conversion can be accomplished using various “colormaps”. In Chapter 7, we utilized the ’jet’ colormap, which is commonly used by bioacousticians. However, other colormaps such as ’turbo’ or ’parula’ might yield better results due to their linear appearance to human vision. Future research should investigate the impact of different colormap conversions on model performance. Additionally, attention should be given to the resizing of images when using computer vision algorithms, as the resulting resolution may not accurately represent the sound. In Chapter 7, during the exploration phase, we conducted several hyperparameter search runs for YOLOv8 using Ray Tune [52] to determine the optimal values for certain hyperparameters. Although we experimented with larger image sizes, 640x640 was ultimately selected as the best-performing resolution.

Table 10.3 Summary of model approaches and which tasks are they suitable for. OD stands for Object detector. P/A stands for presence/absence. ED stands for event detection. $\sim$ means possible, but not needed (usually requires larger computing times or training sets). ✓means the model approach is suited for the task. "no" means the model approach is not suited for the task. ? means that further tests are necessary.

			Single label		Multi-label		OD
		Task	P/A	ED	P/A	ED	
Individual call detection	Multiple call types	no overlapping sounds	no	no	no	✓	$\sim$
		overlapping sounds different call types	no	no	no	✓	✓
		overlapping sounds same call type	no	no	no	no	✓[2]
		unknown sounds	no	no	no	?	✓
	One single call type	no overlapping sounds	no	✓	no	no	$\sim$
		overlapping sounds	no	no	no	no	✓
P/A	Multiple call types		no	no	✓	$\sim$	$\sim$
	One single call type		✓	$\sim$	no	no	$\sim$

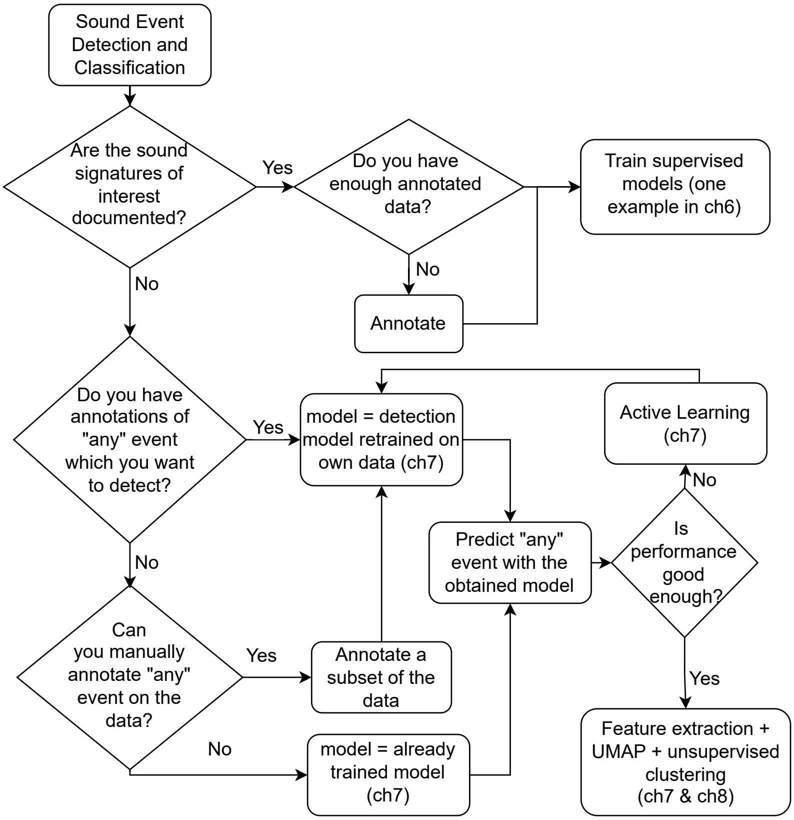
10.3.2 Rules of thumb for sound event detection

Deciding which of the approaches explained in Section 1.2 (see Figure 1.2 in Chapter 1) to use when training a model for detection and/or classification of acoustic events depends on the ultimate objective of the study (see Table 10.3 for summary). There is currently no one-fits-all recipe for deciding which model to choose when starting a bioacoustics event detection task using ML.

Single label presence/absence can be used for hourly or daily presence detection of certain species. However, if individual calls need to be detected and counted for biological and ecological assessments (e.g. for density estimation), one would need to use object detection or single- or multi-label event detection (see Table 10.3). If one wants to detect species which might be happen simultaneously, a multi-label approach will be necessary, or an object detector (Table 10.3). The principal advantage of an object detector compared to a multi-label event detection (see Figure 1.2), is that it provides the frequency limits of each detection. This is not critical when detecting known sounds or sounds on the same frequency range, but it is crucial when studying unknown sounds to classify them afterward.

Even though object detectors have not been widely used in bioacoustics research [53], they have been successfully applied in some terrestrial studies [54, 55, 56, 57]

Figure 10.2 Flowchart of a simplified guideline to decide which of the methods presented in this thesis regarding sound event detection is suitable given a certain task. Chapters where the methods are explained are referred to in between brackets. ”any” event refers to all the possible sounds of interest.



and to detect marine mammal calls [58, 59]. From these studies, Faster-RCNN [56, 57, 54] and YOLO [55, 59] were the most used approaches. The advantage of YOLO is that it is a lot faster and uses a lot less computational resources than Faster-RCNN without compromising accuracy. Most of these studies used the object detector to detect a set of pre-defined calls. Only Romero-Mujalli et al. (2021) [56] also explored the possibility of clustering the obtained detections using their frequency, duration, and contour. The obtained clusters were compared with pre-defined classes and not to explore unknown sound classes, which is what we tackle in Chapter 7.

If one would be interested in exploring all the possible unknown sounds in a location where visual detection and sound recording cannot be performed simultaneously (e.g. because of high turbidity), we would propose to try out the pipeline proposed in Chapter 7. Depending on the application, the trained model to detect "possible sound events of interest" can be used directly. However, even though the detection model presented in Chapter 7 is proven to be robust when used on a totally different location, re-training might be necessary if the focus is on sounds with a different frequency range or duration. In this case, annotating the necessary files to train the YOLOv8 supervised model (the event detection) is still time intensive. For this reason, the active learning approach we developed is an advancement for the field, as it limits the necessary amount of annotated data. In Figure 10.2 we present a flowchart to explain which of the acoustic event detection methods presented in this thesis would be useful for each case. This flowchart aims to help understand how the different methods presented in Chapters 6 to 8 relate to each other.

10.4 Generative aspects of this research

Over the entire PhD trajectory, we have put a big emphasis on publishing modular open-source code suitable to be re-used by other scientists. Generating flexible software that can be used with a limited coding background is important to facilitate acoustic analysis for non-computer scientists or programming experts. Documenting and generalizing code so it can be used by other parties is a time-consuming task. However, open-source code allows for the reproducibility of research, making science more robust. Furthermore, it promotes collaboration and it can help the marine bioacoustics community speed up development without losing resilience.

The python packages resulting from this research path are listed below:

- *pyporcc*. A Python package to detect and classify porpoise clicks. This package facilitates the reading of raw snippets data from the SoundTrap HF (Oceaninstruments, New Zealand) hydrophone and can classify them into porpoise click or not porpoise click using the algorithm of [60]
`https://github.com/lifewatch/pyporcc`
- *bpnsdata*. A Python package to add environmental data to a pandas DataFrame which has time and space information. It retrieves the environmental information from different public datasets.
`https://github.com/lifewatch/bpnsdata`
- *pypam*. A Python package to process long-term underwater acoustic data data in chunks. It deals with the calibration and the concatenation of files, and it allows processing multiple features on one chunk, so data only needs to be loaded once. The output is in netcdf files, and they incorporate all the

necessary self-explanatory data.
`https://github.com/lifewatch/pypam`

- *pyhydrophone*. A Python package to process hydrophone data in a standardized and easy way. Functions are made specifically per hydrophone type so the processing can be standardized and automated using the same function name. All of them include calibrating the output data, either with a flat frequency response or a frequency-dependent calibration.
`https://github.com/lifewatch/pyhydrophone`

Apart from the stand-alone packages, all the scripts developed for specific projects and publications are also provided as reusable scripts and are open and accessible. A list is provided below:

- Baleen whale benchmark. Scripts to run a benchmark of baleen whale calls detection using Machine Learning on the Miller dataset [61]
`https://github.com/cparcerisas/baleen_whale_benchmark`
- SoundCoop project jupyter notebook collection. A collection of jupyter notebooks that people can use to load, visualize, and interpret the output in hybrid millidecade bands sound pressure levels (from pypam or other software like MANTA)
`https://github.com/ioos/soundcoop`
- Fin whale detector. Contribution to the Python version of the code used in Schall & Parcerisas (2024) [62] to detect fin whale choruses and pulses.
`https://gitlab.awi.de/oza-sound-detectors/fin-whale-chorus-and-pulse-detection` (python version)
- Scripts to reproduce publication Parcerisas et al. (2023) [63], copied to Chapter 5. `https://github.com/lifewatch/categorizing_soundscapes`
- Scripts to reproduce publication Parcerisas et al. (2024) [64], copied to Chapter 6. `https://github.com/lifewatch/sound-segregation-and-categorization`
- Scripts to reproduce publication Calonge et al. (2024) [29], copied to Chapter 7. `https://github.com/lifewatch/unknown_underwater_sounds`

We have invested a lot of hours into developing these packages and repositories, but it has already been paid off by other people, as well as ourselves, using it. For example, pypam is now in use by the U.S. SoundCoop project [65] as one of the tools that might become standard in processing the U.S. underwater long-term data. For this project, a wrapper around pypam has been created (pypam-based-processing, https://github.com/mbari-org/pbp) to extend pypam's capabilities for cloud-based processing. pyporcc can be used as the Python alternative to process the output of the SoundTrap Click Detector or to detect high-frequency clicks from continuous data. The scripts from the publication Parcerisas et al. (2023) [63] are currently being used by the University of Kiel to analyze Wadden Sea soundscapes following the same approach, with promising preliminary results [66]. bpnsdata has also been used by several researchers at the Flanders Marine Institute to obtain environmental data from the BPNS [67, 68, 69].

10.5 Societal contribution of this research

Public involvement has become crucial to scientific research. According to the revised roadmap for the United Nations (UN) Decade of Ocean Science for Sustainable Development [70], increasing ocean literacy in the global population has the potential to enable a positive behavior change towards the ocean and its resources. Ocean literacy does not only refer to the understanding and knowledge of ocean-related topics but also connections to, and attitudes and behaviors towards, the ocean [71]. For this reason, (marine) scientists are increasingly collaborating with artists and designers to effectively communicate their findings to diverse audiences [72]. Artists and scientists can work together to reach a wider population and establish a deeper emotional connection and comprehension towards the ocean [73]. This enrichment can ultimately translate to an increase in sustainable action for the good of the ocean [74].

During the course of this PhD trajectory, I have realized that underwater sound is a source of inspiration for artistic projects and citizen science. Humans seem to be attracted to underwater sound. In the same way the work of Payne and McVay (1971) [75] about humpback whales' songs was a driver to impulse the population to push for the protection of whales, marine soundscapes can be used to raise awareness and engagement in society. During the last 4.5 years, I have collaborated with many artists who have done a beautiful job translating underwater sounds into the language of art; feelings, and emotions. A list of all the projects I have participated in is provided below:

- White Noise composition, by Stephanie Haensler (2020). https://stephaniehaensler.com/white-noise
- Sonic Acts x FieldARTS: Maritime Frictions, by Velma Spell (2021). https://sonicacts.com/agenda/maritime-frictions
- Ear to Sea, by Cusk Collective (2022-2023). https://pilar.brussels/en/events/ear-to-sea

- Voice of the North Sea, by Remco de Kluizenaar (2023). https://www.wur.nl/en/newsarticle/voice-of-the-north-sea-update.htm
- Calling in our Corals, by Google Arts & Culture (2023). https://g.co/arts/rqoA5tn3NMvH8FEy7
- De 11e Provincie, by the Instituut voor Onderzoek van de Betovering der Zeeën. \https://iobz.be
- Waves of Resonance, by Elise Guillaume (2024 - ongoing). Project selected for the European Marine Board EMBracingtheOcean programme. https://www.marineboard.eu/elise-guillaume

10.6 Remaining challenges and future work

Soundscape analysis is possible when the sources are unknown, as it has been proven in this dissertation. However, knowing which species produce which sound brings more insight to the PAM analysis of a marine ecosystems. Right now this is not possible for many species, but if more data are gathered on the sound production of fish and invertebrates, access to these sounds through open sound source libraries could lead to the further development of supervised detectors and classifiers of fish (e.g. [76, 77, 78]), and invertebrate sounds or other sound categories. It is necessary to document behavioral contexts of their sound production, record and describe the acoustic characteristics of the sounds [79, 80]. International initiatives such as GLUBS are working towards making this library a possibility [81].

To build these libraries, sound ground truthing mechanisms are necessary. Currently, this is done in lab experiments or by deploying cameras in the wild simultaneously with a hydrophone. Both of these approaches present their disadvantages. In captivity, some animals present a different soniferous behavior than in the wild [79]. In the wild, a camera with a single hydrophone can be inconclusive to determine the source of the recorded sounds. Recently Mouy et al. (2023) [82] have proposed a system to localize sound producers with an array of hydrophones and video cameras. However, solutions entailing cameras are of little use in turbid waters. Therefore, there is a need to develop other sound ground truthing mechanisms. These could be based on the exploration of spatio-temporal patterns of long-term recordings covering large areas. This is only possible if data are open and accessible. Sounds that remain unidentified should also be included in these libraries so they can be identified in the future. Currently, once a sound is detected and selected as relevant, it is common to name it by assigning it a phonetic description (onomatopeia). However, the onomatopeias described in the literature do not always represent the same sounds, which complicates the comparison between areas and studies [1]. This is key for unidentified sounds, as naming agreements will lead to easier comparison of the spatio-temporal patterns of those sound types. Therefore, naming conventions are necessary for unidentified sounds.

Other sound ground truthing existing alternatives in the field in areas where visual surveys are not possible are fish tracking and the use of echosounders. Regarding fish tracking, the European Tracking Network (ETN) aims to track aquatic animals across Europe using Telemetry, which uses acoustic tags to investigate the ecology and movement behavior of aquatic species in relation to their environment. Echousounder data can also be to detect the presence of schools of a certain species, and then check if there is a correlation between the presence of this specie and some of the recorded sounds. These technologies can thus be used to know when a certain fish species was present in the vicinity of the acoustic station. Therefore, the LifeWatch permanent acoustic receiver network in the BPNS [83] combined with the LifeWatch Broadband Acoustic Network [84] makes it possible to reduce the amount of acoustic recordings to be manually annotated to find fish vocalizations, and could be used as a semi ground truthing mechanism. Both these approaches are currently being addressed at VLIZ, but no robust conclusions have been made yet.

If two sound sources occur simultaneously and one is louder it is referred to as masking. This can have serious impacts on populations relying on sound for communication and foraging [85, 86, 87]. When this occurs, one possible way to extract the signal of interest is to apply source separation or denoising algorithms. Source separation has received comparatively less attention than automatic detection and classification of sounds, but it has already been applied to soundscape descriptions [88, 25]. Denoising algorithms are already applied to a variety of acoustic problems, both using traditional signal processing techniques and Machine Learning solutions. In the field of bioacoustics there is no current satisfactory algorithm that can be used in a wide range of conditions [89]. Due to the nature of the possibly biological sounds very often stationary and non-stationary noise reduction techniques such as noisereduce [90] also remove the signal of interest. We saw that for our dataset, it produced worse results and several signals were not identifiable after the noise reduction. Low-, band-, and high-pass filtering are widely used methods. However, these methods are only applicable when the frequency limits of each detection are available. Furthermore, they are limited in their ability to handle recordings with overlapping calls and are highly dependent on the species of animal being studied [91]. Developing robust tools for denoising and source separation and applying them to soundscape description in a standardized way could be an added value to the field.

An additional challenge is that sometimes multiple sound events of interest occur simultaneously and in the same frequency band. Even though our proposal of detecting sounds using an object detector instead of the more common single- or multi-label segmentation approach is an advantage in this situation because it can detect several sound events happening at the same time, it does not provide a solution when multiple individuals from the same species vocalize at the same moment. In this situation, it is not straightforward to detect and classify the different sound sources, an issue known as the cocktail party problem. There have been some successful approaches to solve this for underwater acoustics [92], but it is still a problem when applied to detectors for long-term data.

State-of-the-art machine learning semi-supervised approaches including humans in the loop have great potential to be applied to marine ecosystems [93], to reduce the expensive labeling effort to a minimum by making it more efficient. Even though methods such as active learning make the annotation process more efficient, it is still human resource-intensive. With the advancement of Machine Learning tools, other approaches for marine sound identification can be developed that need very little to no labeling effort. Examples would be few-shot learning, such as the algorithms resulting from the challenge listed in DCASE2023 Task5: few-shot bioacoustic event detection [48] or zero-shot learning [94].

Additionally, there is evidence suggesting that certain marine fauna may be more sensitive to the particle motion component of the sound field rather than the sound pressure [95, 80]. All the results of this thesis are based on the sound pressure component but not the particle motion. This bias is because there are currently no standardized tools to precisely measure, report, and analyze particle motion [96]. The amplitudes of sound pressure and particle motion are linearly correlated under certain conditions, but in the near field (near a sound source) this assumption is not met, and sound pressure and particle motion might scale very differently. Therefore, only measuring one of the two components might not be representative of the acoustic cues received by some fauna. To close this bias and facilitate particle motion measurements, precise particle motion instruments are needed so they can be deployed in the lab and the wild.

The methods presented in this thesis are only focused on the received sound. However, in shallow waters the received sound can greatly differ from the emitted sound due to the complexities of propagation. The sounds received at listening locations might differ from the actual acoustic activity in the surrounding environment [97]. As mentioned in Chapter 3, it would be necessary to produce sound propagation maps for different known sources and for different water depth conditions considering bathymetry and sediment type (and depth). Due to the complexity of this matter, this topic was not addressed during this thesis, and it remains as future work. The propagation conditions should be considered when deciding monitoring locations to be able to capture the variation in soundscapes within a region [97]. Finally, this information should also be considered when comparing acoustic characteristics between sites.

The constant growth and innovation in marine technology have led to larger storage capacities and battery life of underwater acoustic recorders, which allow for long-term monitoring of remote areas. The same technological advancement is now equipping autonomous and semi-autonomous platforms such as gliders, wave gliders, AUVs, USVs or Argo floats with acoustic recorders. This is a very important step towards obtaining real-time PAM data, and therefore future PAM developments should be carried out considering their applicability to edge-processing and their power and data efficiency.

Finally, a very often neglected topic in machine learning is the actual environmental impact of training and deploying these models [98]. To understand the real environmental impact of ML it is necessary to consider ML systems as a whole. This includes examining machine learning pipelines from data collection to model deployment. Optimizing efficiency across software and hardware is crucial, aiming for competitive model accuracy with reduced computational costs. Responsible AI development involves considering the environmental consequences of innovations and adopting sustainable practices. That is, we need ML to be green and environmentally-sustainable [99].

Additionally, data storage is often seen as an immaterial and unlimited resource. This is however not the case, as data storage is also physical [100]. Institutions should determine what data should be stored and ensure accountability and fairness in this decision-making process. This responsibility should not be solely delegated to industry, as it has significant implications for entire societies and local communities [100]. Therefore, only necessary and meaningful data should be collected and stored. Collected data could be reduced to its minimum essential information before being stored. Applied to underwater passive acoustics research, this would imply storing only relevant acoustic features or events extracted from continuous recordings rather than the raw recordings. Currently, as already mentioned, no feature set is known to be general and robust enough across ecosystems, but the work done to implement ocean sound as an essential ocean variable (EOV) [9] or sharing hybrid millidecade bands (sound pressure) levels [101] already points in that direction.

References

[1] A. Looby, S. Vela, K. Cox, A. Riera, S. Bravo, H. L. Davies, R. Rountree, L. K. Reynolds, C. W. Martin, S. Matwin, and F. Juanes. *FishSounds Version 1.0: A Website for the Compilation of Fish Sound Production Information and Recordings*. Ecological Informatics, 74:101953, May 2023.

[2] L. Coquereau, J. Grall, L. Chauvaud, C. Gervaise, J. Clavier, A. Jolivet, and L. Di Iorio. *Sound Production and Associated Behaviours of Benthic Invertebrates from a Coastal Habitat in the North-East Atlantic*. Marine Biology, 163(5):127, May 2016.

[3] F. Thomsen, S. Mendes, F. Bertucci, M. Breitzke, E. Ciappi, A. Cresci, E. Debusschere, C. Ducatel, T. Folegot, C. Juretzek, Frans-Peter Lam, J. O'Brien, and M. E. D. Santos. *Addressing Underwater Noise in Europe: Current State of Knowledge and Future Priorities*. Technical report, Zenodo, October 2021.

[4] J. Butler, M. J. Butler, and H. Gaff. *Snap, Crackle, and Pop: Acoustic-based Model Estimation of Snapping Shrimp Populations in Healthy and Degraded Hard-Bottom Habitats*. Ecological Indicators, 77:377–385, June 2017.

[5] T. A. C. Gordon, H. R. Harding, K. E. Wong, N. D. Merchant, M. G. Meekan, M. I. McCormick, A. N. Radford, and S. D. Simpson. *Habitat Degradation Negatively Affects Auditory Settlement Behavior of Coral Reef Fishes*. Proceedings of the National Academy of Sciences, 115(20):5193–5198, May 2018.

[6] R. P. Schoeman, C. Erbe, G. Pavan, R. Righini, and J. A. Thomas. *Analysis of Soundscapes as an Ecological Tool*. In C. Erbe and J. A. Thomas, editors, Exploring Animal Behavior Through Sound: Volume 1: Methods, pages 217–267. Springer International Publishing, Cham, 2022.

[7] M. McKenna, G. Shannon, and K. Fristrup. *Characterizing Anthropogenic Noise to Improve Understanding and Management of Impacts to Wildlife*. Endangered Species Research, 31:279–291, November 2016.

[8] M. Tasker, M. Amundin, M. Andre, A. Hawkins, W. Lang, T. Merck, A. Scholik-Schlomer, J.Teilmann, F. Thomsen, S. Werner, and M. Zakharia. *Marine Strategy Framework Directive : Task Group 11 Report : Underwater Noise and Other Forms of Energy, April 2010*. Publications Office, LU, 2010.

[9] P. Tyack, T. Akamatsu, O. Boebel, L. Chapuis, E. Debusschere, C. de Jong, C. Erbe, K. Evans, J. Gedamke, T. Gridley, G. Haralabus, R. Jenkins, J.

Miksis-Olds, H. Sagen, F. Thomsen, K. Thomisch, and E. Urban. *Ocean Sound Essential Ocean Variable Implementation Plan*. Technical report, Zenodo, November 2023.

[10] Commission Directive (EU). *Directive 2008/56/EC of the European Parliament and of the Council of 17 June 2008 Establishing a Framework for Community Action in the Field of Marine Environmental Policy (Marine Strategy Framework Directive) (Text with EEA Relevance)*, June 2017.

[11] M. Solé, M. Lenoir, M. Durfort, J.-M. Fortuño, M. van der Schaar, S. De Vreese, and M. André. *Seagrass Posidonia Is Impaired by Human-Generated Noise*. Communications Biology, 4(1):1–11, June 2021.

[12] Commission Directive (EU). *Commission Notice on the Threshold Values Set under the Marine Strategy Framework Directive 2008/56/EC and Commission Decision (EU) 2017/848*, March 2024.

[13] D. C. Wilford, J. L. Miksis-Olds, S. B. Martin, D. R. Howard, K. Lowell, A. P. Lyons, and M. J. Smith. *Quantitative Soundscape Analysis to Understand Multidimensional Features*. Frontiers in Marine Science, 8:672336, August 2021.

[14] T. Luypaert, A. S. Bueno, G. S. Masseli, I. L. Kaefer, M. Campos-Cerqueira, C. A. Peres, and T. Haugaasen. *A Framework for Quantifying Soundscape Diversity Using Hill Numbers*. Methods in Ecology and Evolution, 13(10):2262–2274, 2022.

[15] S. S. Sethi, A. Bick, R. M. Ewers, H. Klinck, V. Ramesh, M.-N. Tuanmu, and D. A. Coomes. *Limits to the Accurate and Generalizable Use of Soundscapes to Monitor Biodiversity*. Nature Ecology & Evolution, pages 1–6, July 2023.

[16] N. Pieretti and R. Danovaro. *Acoustic Indexes for Marine Biodiversity Trends and Ecosystem Health*. Philosophical Transactions of the Royal Society B: Biological Sciences, 375(1814):20190447, December 2020.

[17] S. A. Harris, N. T. Shears, and C. A. Radford. *Ecoacoustic Indices as Proxies for Biodiversity on Temperate Reefs*. Methods in Ecology and Evolution, 7(6):713–724, 2016.

[18] M. Peck, R. F. Tapilatu, E. Kurniati, and C. Rosado. *Rapid Coral Reef Assessment Using 3D Modelling and Acoustics: Acoustic Indices Correlate to Fish Abundance, Diversity and Environmental Indicators in West Papua, Indonesia*. PeerJ, 9:e10761, February 2021.

[19] A. N. Rice, M. S. Soldevilla, and J. A. Quinlan. *Nocturnal Patterns in Fish Chorusing off the Coasts of Georgia and Eastern Florida*. Bulletin of Marine Science, 93(2):455–474, April 2017.

[20] T. A. Mooney, L. Di Iorio, M. Lammers, T.-H. Lin, S. L. Nedelec, M. Parsons, C. Radford, E. Urban, and J. Stanley. *Listening Forward: Approaching Marine Biodiversity Assessments Using Acoustic Methods*. Royal Society Open Science, 7(8):201287, 2020.

[21] M. Minello, L. Calado, and F. C. Xavier. *Ecoacoustic Indices in Marine Ecosystems: A Review on Recent Developments, Challenges, and Future Directions*. ICES Journal of Marine Science, 78(9):3066–3074, November 2021.

[22] X. Raick, L. D. Iorio, D. Lecchini, M. Bolgan, and E. Parmentier. *"To Be, or Not to Be": Critical Assessment of the Use of α-Acoustic Diversity Indices to Evaluate the Richness and Abundance of Coastal Marine Fish Sounds*. JEA, 7(1):0–0, April 2023.

[23] I. Roca and I. Van Opzeeland. *Using Acoustic Metrics to Characterize Underwater Acoustic Biodiversity in the Southern Ocean*. Remote Sensing in Ecology and Conservation, 6, September 2019.

[24] B. Williams, T. A. C. Lamont, L. Chapuis, H. R. Harding, E. B. May, M. E. Prasetya, M. J. Seraphim, J. Jompa, D. J. Smith, N. Janetski, A. N. Radford, and S. D. Simpson. *Enhancing Automated Analysis of Marine Soundscapes Using Ecoacoustic Indices and Machine Learning*. Ecological Indicators, 140:108986, July 2022.

[25] T.-H. Lin, T. Akamatsu, F. Sinniger, and S. Harii. *Exploring Coral Reef Biodiversity via Underwater Soundscapes*. Biological Conservation, 253:108901, January 2021.

[26] T.-H. Lin, T. Akamatsu, and Y. Tsao. *Sensing Ecosystem Dynamics via Audio Source Separation: A Case Study of Marine Soundscapes off Northeastern Taiwan*. PLOS Computational Biology, 17(2):e1008698, 18 de febr. 2021.

[27] Y.-J. Sun, S.-C. Yen, and T.-H. Lin. *soundscape_IR: A Source Separation Toolbox for Exploring Acoustic Diversity in Soundscapes*. Methods in Ecology and Evolution, 13(11):2347–2355, 2022.

[28] E. B. Kim, K. E. Frasier, M. F. McKenna, A. C. M. Kok, L. E. Peavey Reeves, W. K. Oestreich, G. Arrieta, S. Wiggins, and S. Baumann-Pickering. *SoundScape Learning: An Automatic Method for Separating Fish Chorus in Marine Soundscapes*. The Journal of the Acoustical Society of America, 153(3):1710–1722, March 2023.

[29] A. Calonge, C. Parcerisas, E. Schall, and E. Debusschere. *Revised Clusters of Annotated Unknown Sounds in the Belgian Part of the North Sea*. Frontiers in Remote Sensing, 5, June 2024.

[30] P. Best, S. Paris, H. Glotin, and R. Marxer. *Deep Audio Embeddings for Vocalisation Clustering*. PLOS ONE, 18(7):e0283396, July 2023.

[31] E. Ozanich, A. Thode, P. Gerstoft, L. A. Freeman, and S. Freeman. *Deep Embedded Clustering of Coral Reef Bioacoustics*. The Journal of the Acoustical Society of America, 149(4):2587–2601, April 2021.

[32] T. Sainburg, M. Thielk, and T. Q. Gentner. *Finding, Visualizing, and Quantifying Latent Structure across Diverse Animal Vocal Repertoires*. PLOS Computational Biology, 16(10):e1008228, October 2020.

[33] Y. F. Phillips, M. Towsey, and P. Roe. *Revealing the Ecological Content of Long-Duration Audio-Recordings of the Environment through Clustering and Visualisation*. PLOS ONE, 13(3):e0193345, March 2018.

[34] P. Nguyen Hong Duc, M. Torterotot, F. Samaran, P. R. White, O. Gérard, O. Adam, and D. Cazau. *Assessing Inter-Annotator Agreement from Collaborative Annotation Campaign in Marine Bioacoustics*. Ecological Informatics, 61:101185, March 2021.

[35] E. C. Leroy, K. Thomisch, J.-Y. Royer, O. Boebel, and I. Van Opzeeland. *On the Reliability of Acoustic Annotations and Automatic Detections of Antarctic Blue Whale Calls under Different Acoustic Conditions*. The Journal of the Acoustical Society of America, 144(2):740–754, August 2018.

[36] P. Rubbens, S. Brodie, T. Cordier, D. Destro Barcellos, P. Devos, J. A. Fernandes-Salvador, J. I. Fincham, A. Gomes, N. O. Handegard, K. Howell, C. Jamet, K. H. Kartveit, H. Moustahfid, C. Parcerisas, D. Politikos, R. Sauzède, M. Sokolova, L. Uusitalo, L. Van den Bulcke, A. T. M. van Helmond, J. T. Watson, H. Welch, O. Beltran-Perez, S. Chaffron, D. S. Greenberg, B. Kühn, R. Kiko, M. Lo, R. M. Lopes, K. O. Möller, W. Michaels, A. Pala, J.-B. Romagnan, P. Schuchert, V. Seydi, S. Villasante, K. Malde, and J.-O. Irisson. *Machine Learning in Marine Ecology: An Overview of Techniques and Applications*. ICES Journal of Marine Science, page fsad100, August 2023.

[37] M. Bittle and A. Duncan. *A Review of Current Marine Mammal Detection and Classification Algorithms for Use in Automated Passive Acoustic Monitoring*. In Proceedings of Acoustics 2013, Australia, November 2013.

[38] N. Pegg, I. T. Roca, D. Cholewiak, G. E. Davis, and S. M. Van Parijs. *Evaluating the Efficacy of Acoustic Metrics for Understanding Baleen Whale Presence in the Western North Atlantic Ocean*. Frontiers in Marine Science, 8, 2021.

[39] G. Truong and T. L. Rogers. *Seasonal Occurrence of Sympatric Blue Whale Subspecies: The Chilean and Southeast Indian Ocean Pygmy Blue Whales With the Antarctic Blue Whale*. Frontiers in Marine Science, 8, 2021.

[40] V. E. Warren, A. Širović, C. McPherson, K. T. Goetz, C. A. Radford, and R. Constantine. *Passive Acoustic Monitoring Reveals Spatio-Temporal Distributions of Antarctic and Pygmy Blue Whales Around Central New Zealand*. Frontiers in Marine Science, 7, 2021.

[41] J. J. Luczkovich and M. W. Sprague. *Soundscape Maps of Soniferous Fishes Observed From a Mobile Glider*. Frontiers in Marine Science, 9, 2022.

[42] L. R. Brewster, A. K. Ibrahim, J. Locascio, B. C. DeGroot, L. M. Chérubin, and M. J. Ajemian. *Seasonal Dynamics and Environmental Drivers of Goliath Grouper (Epinephelus Itajara) Sound Production*. Fishes, 8(6):293, June 2023.

[43] D. H. Lee, J. W. Choi, S. Shin, and H. C. Song. *Temporal Variability in Acoustic Behavior of Snapping Shrimp in the East China Sea and Its Correlation With Ocean Environments*. Frontiers in Marine Science, 8, 2021.

[44] T.-H. Lin, Y. Tsao, and T. Akamatsu. *Comparison of Passive Acoustic Soniferous Fish Monitoring with Supervised and Unsupervised Approaches*. The Journal of the Acoustical Society of America, 143(4):EL278–EL284, April 2018.

[45] C. Molnar. *Interpretable Machine Learning. A Guide for Making Black Box Models Explainable*. 2 edition, 2022.

[46] N. Kokhlikyan, V. Miglani, M. Martin, E. Wang, B. Alsallakh, J. Reynolds, A. Melnikov, N. Kliushkina, C. Araya, S. Yan, O. Reblitz-Richardson, and F. Ai. *Captum: A Unified and Generic Model Interpretability Library for PyTorch*, 2009.

[47] A. Farahani, S. Voghoei, K. Rasheed, and H. R. Arabnia. *A Brief Review of Domain Adaptation*. In R. Stahlbock, G. M. Weiss, M. Abou-Nasr, C.-Y. Yang, H. R. Arabnia, and L. Deligiannidis, editors, Advances in Data Science and Information Engineering, pages 877–894, Cham, 2021. Springer International Publishing.

[48] J. Liang, I. Nolasco, B. Ghani, H. Phan, E. Benetos, and D. Stowell. *Mind the Domain Gap: A Systematic Analysis on Bioacoustic Sound Event Detection*, March 2024.

[49] E. Schall, I. I. Kaya, E. Debusschere, P. Devos, and C. Parcerisas. *Deep Learning in Marine Bioacoustics: A Benchmark for Baleen Whale Detection*. Remote Sensing in Ecology and Conservation, April 2024.

[50] M. Thomas, F. H. Jensen, B. Averly, V. Demartsev, M. B. Manser, T. Sainburg, M. A. Roch, and A. Strandburg-Peshkin. *A Practical Guide for Generating Unsupervised, Spectrogram-Based Latent Space Representations of Animal Vocalizations*. Journal of Animal Ecology, 91(8):1567–1581, 2022.

[51] J. A. Hildebrand. *Anthropogenic and Natural Sources of Ambient Noise in the Ocean*. Marine Ecology Progress Series, 395:5–20, December 2009.

[52] R. Liaw, E. Liang, R. Nishihara, P. Moritz, J. E. Gonzalez, and I. Stoica. *Tune: A Research Platform for Distributed Model Selection and Training*, July 2018.

[53] D. Stowell. *Computational Bioacoustics with Deep Learning: A Review and Roadmap*. PeerJ, 10:e13152, March 2022.

[54] R. Shrestha, C. Glackin, J. Wall, and N. Cannings. *Bird Audio Diarization with Faster R-CNN*. In I. Farkaš, P. Masulli, S. Otte, and S. Wermter, editors, Artificial Neural Networks and Machine Learning – ICANN 2021, pages 415–426, Cham, 2021. Springer International Publishing.

[55] S. Zsebők, M. F. Nagy-Egri, G. G. Barnaföldi, M. Laczi, G. Nagy, É. Vaskuti, and L. Z. Garamszegi. *Automatic Bird Song and Syllable Segmentation with an Open-Source Deep-Learning Object Detection Method – a Case Study in the Collared Flycatcher*. Ornis Hungarica, 27(2):59–66, December 2019.

[56] D. Romero-Mujalli, T. Bergmann, A. Zimmermann, and M. Scheumann. *Utilizing DeepSqueak for Automatic Detection and Classification of Mammalian Vocalizations: A Case Study on Primate Vocalizations*. Scientific Reports, 11(1):24463, December 2021.

[57] K. R. Coffey, R. E. Marx, and J. F. Neumaier. *DeepSqueak: A Deep Learning-Based System for Detection and Analysis of Ultrasonic Vocalizations*. Neuropsychopharmacology, 44(5):859–868, April 2019.

[58] J. H. Rasmussen and A. Širović. *Automatic Detection and Classification of Baleen Whale Social Calls Using Convolutional Neural Networks*. The Journal of the Acoustical Society of America, 149(5):3635–3644, May 2021.

[59] R. Zhang, Y. Shi, L. Yang, and Z. Zhou. *Classification of Whale Call Signals Based on the Improved YOLOv5 Model*. In 2023 6th International Conference on Information Communication and Signal Processing (ICICSP), pages 429–433, September 2023.

[60] M. Cosentino, F. Guarato, J. Tougaard, D. Nairn, J. C. Jackson, and J. F. C. Windmill. *Porpoise Click Classifier (PorCC): A High-Accuracy Classifier to Study Harbour Porpoises (Phocoena Phocoena) in the Wild*. The Journal of the Acoustical Society of America, 145(6):3427–3434, 2019.

[61] B. S. Miller, N. Balcazar, S. Nieukirk, E. C. Leroy, M. Aulich, F. W. Shabangu, R. P. Dziak, W. S. Lee, and J. K. Hong. *An Open Access Dataset for Developing Automated Detectors of Antarctic Baleen Whale Sounds and Performance Evaluation of Two Commonly Used Detectors*. Scientific Reports, 11(1):806, January 2021.

[62] E. Schall and C. Parcerisas. *A Robust Method to Automatically Detect Fin Whale Acoustic Presence in Large and Diverse Passive Acoustic Datasets*. Journal of Marine Science and Engineering, 10(12):1831, December 2022.

[63] C. Parcerisas, I. T. Roca, D. Botteldooren, P. Devos, and E. Debusschere. *Categorizing Shallow Marine Soundscapes Using Explained Clusters*. Journal of Marine Science and Engineering, 11(3):550, March 2023.

[64] C. Parcerisas, E. Schall, K. te Velde, D. Botteldooren, P. Devos, and E. Debusschere. *Machine Learning for Efficient Segregation and Labeling of Potential Biological Sounds in Long-Term Underwater Recordings*. Frontiers in Remote Sensing, 5, April 2024.

[65] C. Wall, L. T. Hatch, R. Bochenek, G. E. Davis, S. M. Haver, D. W. Ponirakis, A. N. Rice, J. Ryan, S. M. Van Parijs, C. Berchok, K. E. Frasier, M. F. McKenna, J. Miksis-Olds, T. Rowell, and K. Thomisch. *The SoundCoop: Developing a Passive Acoustic Monitoring Cyberinfrastructure Center*. In Ocen Sciences, volume OT23B-08, February 2024.

[66] S. Kühn. *Characterizing Different Sediment Habitats in the Wadden Sea Using Passive Acoustics*, November 2023.

[67] N. Meyers, G. Everaert, . R. Janssen, N. Buhhalko, U. Lips, H. Hostens, and B. De Witte. *Assessing Microplastic Contamination in the Belgian Part of the North Sea: Impacts of Dredging Activities and Innovative Sampling Methods*, 2024.

[68] J. Dujardin, M. Perneel, R. Lagaisse, N. Larriona, J. Mortelmans, H. Theetaert, S. Verbrugge, T. Gkritzalis, S. Maere, and P. I. Hablützel. *From*

Bloom to Foam: Tracing the Role of Phaeocystis in Marine Gel Formation, Manuscript to be submitted September 2024.

[69] M. Perneel, J. Mortelmans, P. I. Hablützel, and Alexander. *Microeukaryotic Plankton Ecosystems over Multiple Seasonal Cycles*, Manuscript to be submitted December 2024.

[70] I. O. Commission. *Revised Roadmap for the UN Decade of Ocean Science for Sustainable Development*, 2018.

[71] R. Kelly, K. Evans, K. Alexander, S. Bettiol, S. Corney, C. Cullen-Knox, C. Cvitanovic, K. de Salas, G. R. Emad, L. Fullbrook, C. Garcia, S. Ison, S. Ling, C. Macleod, A. Meyer, L. Murray, M. Murunga, K. L. Nash, K. Norris, M. Oellermann, J. Scott, J. S. Stark, G. Wood, and G. T. Pecl. *Connecting to the Oceans: Supporting Ocean Literacy and Public Engagement*. Reviews in Fish Biology and Fisheries, 32(1):123–143, March 2022.

[72] Nature editorial. *Collaborations with Artists Go beyond Communicating the Science*. Nature, 590(7847):528–528, February 2021.

[73] V. Gewin. *How to Shape a Productive Scientist–Artist Collaboration*. Nature, 590(7846):515–518, February 2021.

[74] M. I. Severin, L. K. Akpetou, P. Annasawmy, F. E. Asuquo, F. Beckman, M. Benomar, A. Jaya-Ram, M. Malouli, J. Mees, I. Monteiro, J. Ndwiga, P. Neves Silva, O. A. Nubi, Y. K. Sim, Z. Sohou, A. T. Shau-Hwai, S. P. Woo, S. Zizah, A. Buysse, F. Raes, L. A. Krug, S. Seeyave, G. Everaert, E. Mahu, and A. I. Catarino. *Impact of the Citizen Science Project COLLECT on Ocean Literacy and Well-Being within a North/West African and South-East Asian Context*. Frontiers in Psychology, 14, June 2023.

[75] R. S. Payne and S. McVay. *Songs of Humpback Whales*. Science, 173(3997):585–597, August 1971.

[76] I. R. Urazghildiiev and S. M. Van Parijs. *Automatic Grunt Detector and Recognizer for Atlantic Cod (* Gadus Morhua *)*. The Journal of the Acoustical Society of America, 139(5):2532–2540, May 2016.

[77] A. K. Ibrahim, H. Zhuang, L. M. Chérubin, M. T. Schärer-Umpierre, and N. Erdol. *Automatic Classification of Grouper Species by Their Sounds Using Deep Neural Networks*. The Journal of the Acoustical Society of America, 144(3):EL196–EL202, September 2018.

[78] M. Vieira, P. J. Fonseca, M. C. P. Amorim, and C. J. C. Teixeira. *Call Recognition and Individual Identification of Fish Vocalizations Based on Automatic*

Speech Recognition: An Example with the Lusitanian Toadfish. The Journal of the Acoustical Society of America, 138(6):3941–3950, December 2015.

[79] A. Looby, K. Cox, S. Bravo, R. Rountree, F. Juanes, L. K. Reynolds, and C. W. Martin. *A Quantitative Inventory of Global Soniferous Fish Diversity*. Reviews in Fish Biology and Fisheries, 32(2):581–595, June 2022.

[80] M. Solé, K. Kaifu, T. A. Mooney, S. L. Nedelec, F. Olivier, A. N. Radford, M. Vazzana, M. A. Wale, J. M. Semmens, S. D. Simpson, G. Buscaino, A. Hawkins, N. Aguilar de Soto, T. Akamatsu, L. Chauvaud, R. D. Day, Q. Fitzgibbon, R. D. McCauley, and M. André. *Marine Invertebrates and Noise*. Frontiers in Marine Science, 10, 2023.

[81] M. J. G. Parsons, T.-H. Lin, T. A. Mooney, C. Erbe, F. Juanes, M. Lammers, S. Li, S. Linke, A. Looby, S. L. Nedelec, I. Van Opzeeland, C. Radford, A. N. Rice, L. Sayigh, J. Stanley, E. Urban, and L. Di Iorio. *Sounding the Call for a Global Library of Underwater Biological Sounds*. Frontiers in Ecology and Evolution, 10, 2022.

[82] X. Mouy, M. Black, K. Cox, J. Qualley, S. Dosso, and F. Juanes. *Identification of Fish Sounds in the Wild Using a Set of Portable Audio-Video Arrays*. Methods in Ecology and Evolution, 14(8):2165–2186, 2023.

[83] J. Reubens, F. Hernandez, and K. Deneudt. *LifeWatch Observatory - Permanent Acoustic Receiver Network in the Belgian Part of the North Sea*, 2017.

[84] C. Parcerisas, D. Botteldooren, P. Devos, and E. Debusschere. *Broadband Acoustic Network Dataset*, 2021.

[85] A. N. Popper and A. D. Hawkins. *An Overview of Fish Bioacoustics and the Impacts of Anthropogenic Sounds on Fishes*. Journal of Fish Biology, 94(5):692–713, 2019.

[86] A. N. Radford, E. Kerridge, and S. D. Simpson. *Acoustic Communication in a Noisy World: Can Fish Compete with Anthropogenic Noise?* Behavioral Ecology, 25(5):1022–1030, September 2014.

[87] S. Degraer, R. Brabant, B. Rumes, and L. Vigin. *Environmental Impacts of Offshore Wind Farms in the Belgian Part of the North Sea: Attraction, Avoidance and Habitat Use at Various Spatial Scales*. Technical report, Royal Belgian Institute of Natural Sciences, OD Natural Environment, Marine Ecology and Management, Brussels, December 2021.

[88] T.-H. Lin and Y. Tsao. *Source Separation in Ecoacoustics: A Roadmap towards Versatile Soundscape Information Retrieval*. Remote Sensing in Ecology and Conservation, 6(3):236–247, 2020.

[89] J. Juodakis and S. Marsland. *Wind-Robust Sound Event Detection and Denoising for Bioacoustics*. Methods in Ecology and Evolution, 13(9):2005–2017, 2022.

[90] T. Sainburg, A. Zorea, A. Saghiran, K. Amr, Lff5, fardage, and rjfarber. *Timsainb/Noisereduce: 3.0.2*. Zenodo, March 2024.

[91] J. Xie, J. G. Colonna, and J. Zhang. *Bioacoustic Signal Denoising: A Review*. Artificial Intelligence Review, 54(5):3575–3597, June 2021.

[92] P. C. Bermant. *BioCPPNet: Automatic Bioacoustic Source Separation with Deep Neural Networks*. Scientific Reports, 11(1):23502, December 2021.

[93] E. Mosqueira-Rey, E. Hernández-Pereira, D. Alonso-Ríos, J. Bobes-Bascarán, and Á. Fernández-Leal. *Human-in-the-Loop Machine Learning: A State of the Art*. Artificial Intelligence Review, 56(4):3005–3054, April 2023.

[94] D. Robinson, A. Robinson, and L. Akrapongpisak. *Transferable Models for Bioacoustics with Human Language Supervision*, August 2023.

[95] A. N. Popper and A. D. Hawkins. *The Importance of Particle Motion to Fishes and Invertebrates*. The Journal of the Acoustical Society of America, 143(1):470–488, January 2018.

[96] S. L. Nedelec, J. Campbell, A. N. Radford, S. D. Simpson, and N. D. Merchant. *Particle Motion: The Missing Link in Underwater Acoustic Ecology*. Methods in Ecology and Evolution, 7(7):836–842, 2016.

[97] M. F. McKenna, S. Baumann-Pickering, A. C. M. Kok, W. K. Oestreich, J. D. Adams, J. Barkowski, K. M. Fristrup, J. A. Goldbogen, J. Joseph, E. B. Kim, A. Kügler, M. O. Lammers, T. Margolina, L. E. Peavey Reeves, T. J. Rowell, J. A. Stanley, A. K. Stimpert, E. J. Zang, B. L. Southall, C. C. Wall, S. V. Parijs, and L. T. Hatch. *Advancing the Interpretation of Shallow Water Marine Soundscapes*. Frontiers in Marine Science, 8, 2021.

[98] P. Dhar. *The Carbon Impact of Artificial Intelligence*. Nature Machine Intelligence, 2(8):423–425, August 2020.

[99] C.-J. Wu, R. Raghavendra, U. Gupta, B. Acun, N. Ardalani, K. Maeng, G. Chang, F. A. Behram, J. Huang, C. Bai, M. Gschwind, A. Gupta, M. Ott, A. Melnikov, S. Candido, D. Brooks, G. Chauhan, B. Lee, H.-H. S. Lee, B.

Akyildiz, M. Balandat, J. Spisak, R. Jain, M. Rabbat, and K. Hazelwood. *Sustainable AI: Environmental Implications, Challenges and Opportunities*. In Proceedings of Machine Learning and Systems 4 (MLSys 2022), 2022.

[100] F. Lucivero. *Big Data, Big Waste? A Reflection on the Environmental Sustainability of Big Data Initiatives*. Science and Engineering Ethics, 26(2):1009–1030, April 2020.

[101] J. L. Miksis-Olds, P. J. Dugan, S. B. Martin, H. Klinck, D. K. Mellinger, D. A. Mann, D. W. Ponirakis, and O. Boebel. *Ocean Sound Analysis Software for Making Ambient Noise Trends Accessible (MANTA)*. Frontiers in Marine Science, 8:703650, August 2021.

Part VI

Supplementary Material

11
Supplementary Material

S3 Supplementary Material Chapter 3

S3.1 Figures

Here we provide some examples of interesting sounds found while manually scrolling through the data. These examples have been manually selected. The spectrograms were computed with a number of FFT bins of 2048, hop size of 164. A Butterworth band-pass filtered of order 4 between 50 Hz and 1.5 times the maximum frequency seen in the spectrogram (varies depending on the example) was applied. Then data were converted top dB scale, and then normalized to the [5, 99] % range for better visualization.

Figure S3.1 Snippet from the Grafton station the 13th of April of 2021. It lasts for several hours. It could be coming from anthropogenic sources, but it is still to be confirmed.

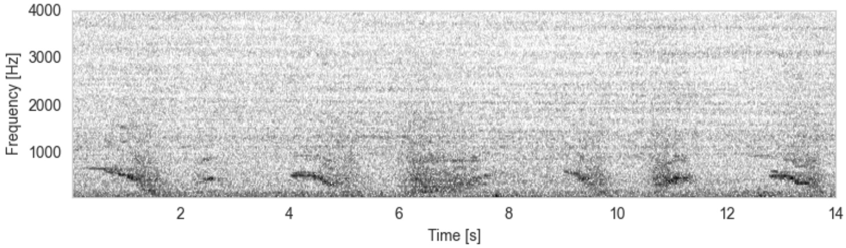


Figure S3.2 Snippet from Grafton station the 7th of November of 2022. It is a drumroll, described in Chapters 7 and 8 as 'Jackhammer'. Possibly coming from fish, but the species is still unknown. The drumroll is often found in groups, with varying number of pulses every occurrence. So far we have only found it in the Grafton station, from 28th October to 7th November 2022.

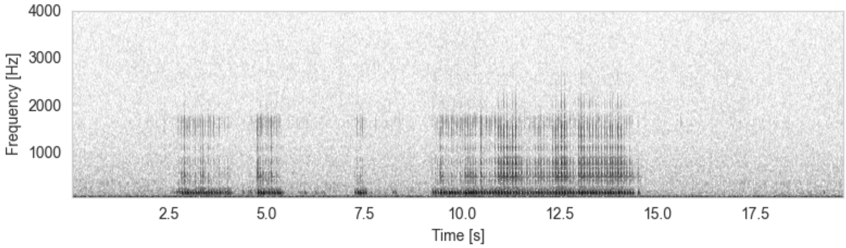


Figure S3.3 Snippet from Fairplay station the 26th of January of 2022. Similar to the Jackhammer from Figure S3.2, but at lower frequencies and not occurring in big groups, only once. Higher number of pulses per occurrence than the Jackhammer, and less variability in the number of pulses. Found in stations Fairplay, Buitenratel, and GardenCity.

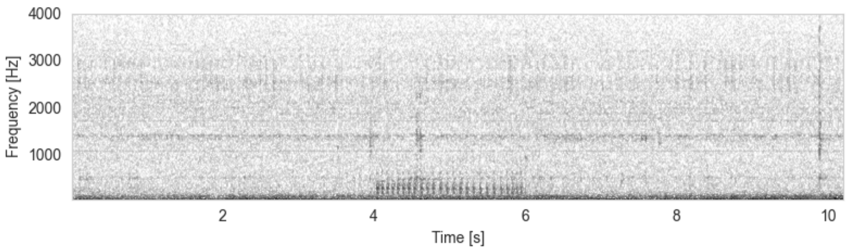


Figure S3.4 Snippet from Grafton station the 1st of September of 2022. These 'rasps' and 'pocs' are commonly found on all the stations and throughout the year. We hypothesize that they have a biological origin, even though it is not clear if it is via contact with the hydrophone or real propagated sound.

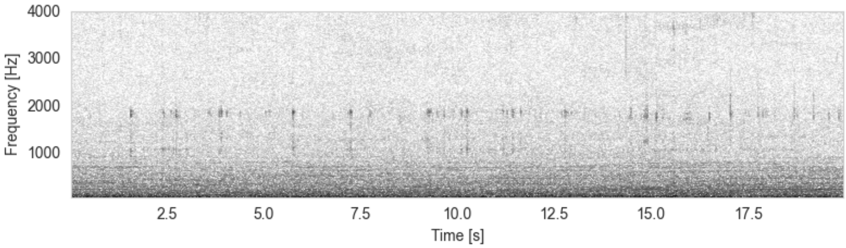
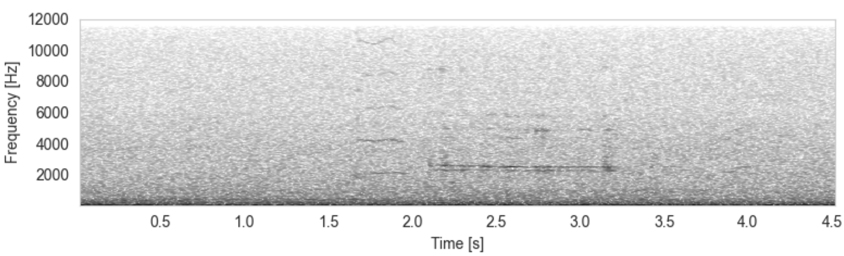


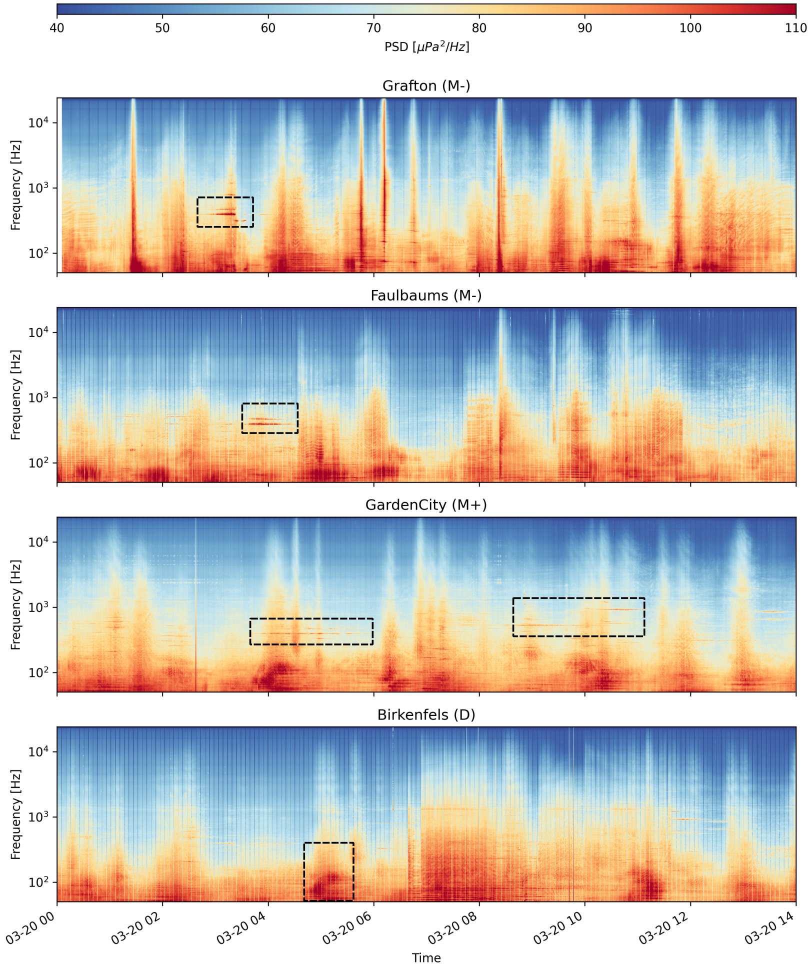
Figure S3.5 Snippet from the GardenCity station the 31st of March of 2021. There is first a whistle and then a Jingle bell (see Chapters 7 and 8 for more information. Similar whistles with harmonics have not been found very often, and they are possibly from a mammal. The Jingle bells are found on all the stations, throughout the year. It is still unknown the source.).



S4 Supplementary Material Chapter 4

S4.1 Figures

Figure S4.1 Examples of several LTSA for the 20th of March of 2021 until 14:00 AM, from stations Grafton, Faulbaums, GardenCity and Birkenfels. This illustrates how LTSAs can be used to select periods of interest for analysis, some examples highlighted with black boxes.



S5 Supplementary Material Chapter 5

S5.1 Tables

Table S5.1 Parameters used for the UMAP dimension reduction and the DBSCAN algorithm, both for the Drifts and the Stationary datasets.

	UMAP		DBSCAN	
	n neighbors	**min. dist.**	**epsilon**	**min. samples**
Drift (full dataset)	10	0	0.5	100
Drift (clean dataset)	50	0	0.9	720
Stationary	5	0	0.47	500

Table S5.2 Parameters used for the grid search during hyper-parameter tuning for the training of the RF.

Parameter	Values
criterion	gini, entropy
min_samples_split	100, 300, 500
max_depth	6, 8, 12
min_samples_leaf	100, 200, 300
max_leaf_nodes	10, 15, 20

Table S5.3 Mean and standard deviations of the values of the one-third octave bands, in time per each category.

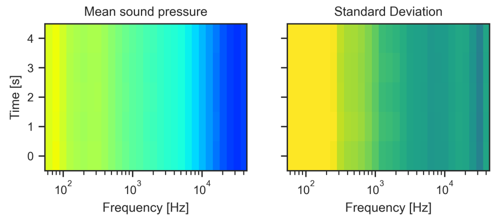
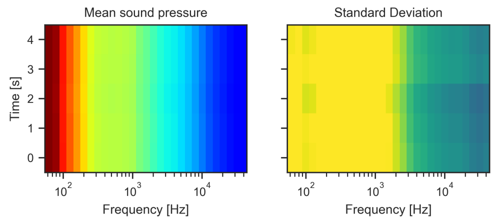
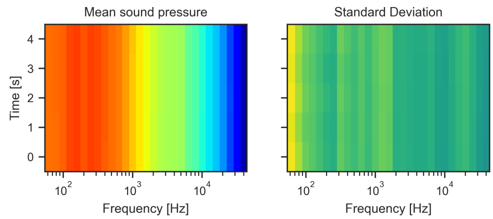
Cluster	**Acoustic Summary**
0	
1	
2	

Table S5.4 *Cont*

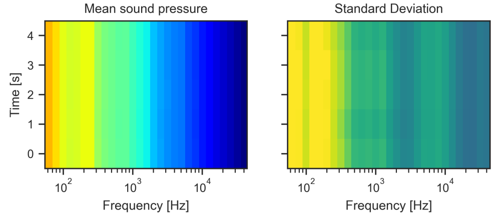
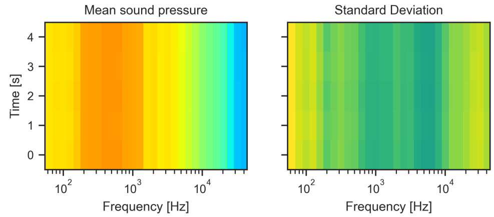
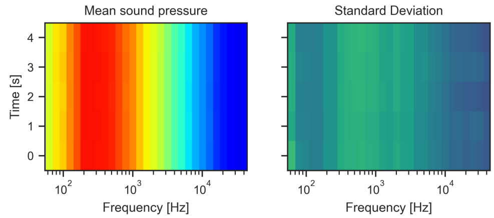
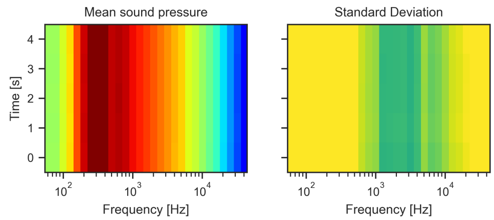
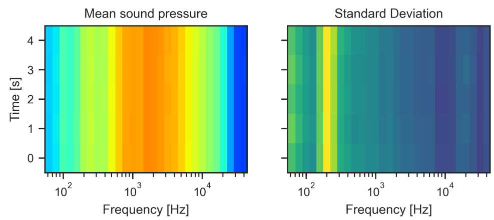
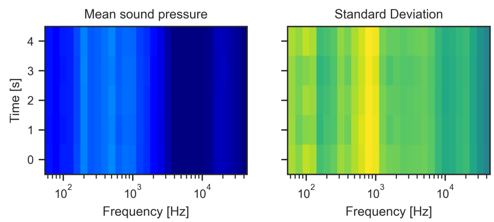
Cluster	Acoustic Summary
3	
4	
5	
6	
7	
8	

Table S5.4 *Cont*

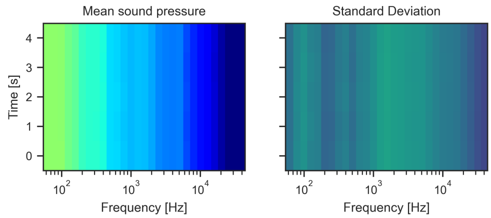
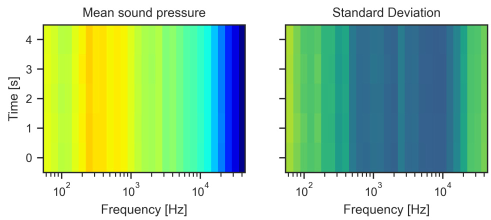
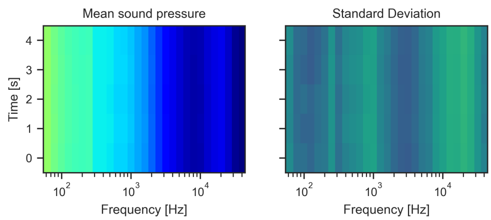
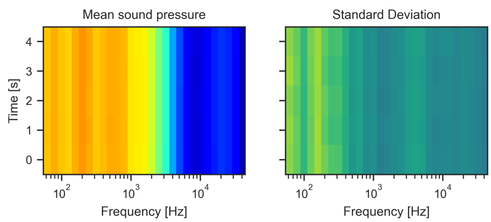
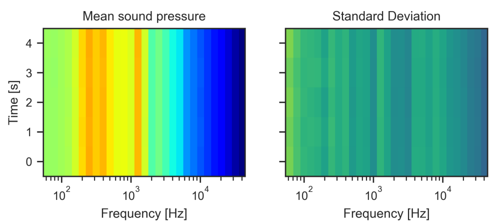
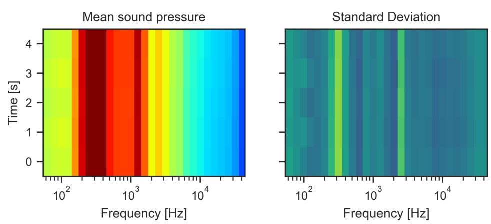
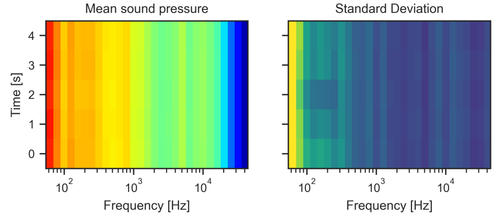
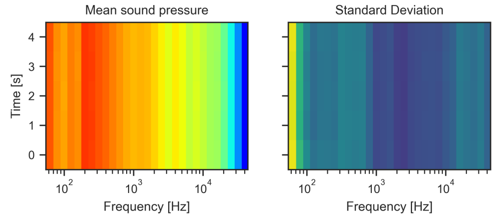
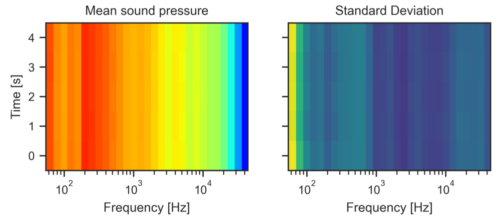
9	
10	
11	
12	
13	
14	

Table S5.4 *Cont*

Cluster	Acoustic Summary
15	
16	
17	

S5.2 Figures

Figure S5.1 *Cont.*

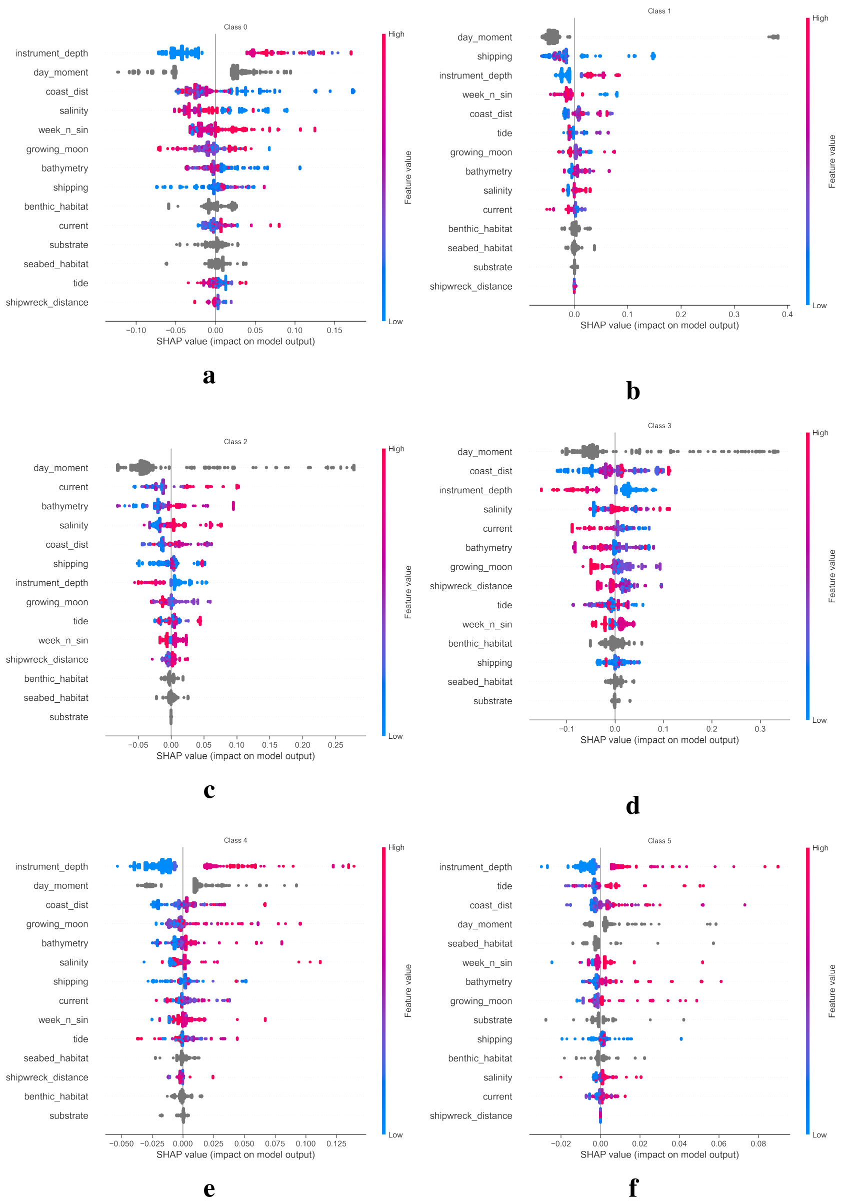


Figure S5.1 *Cont.*

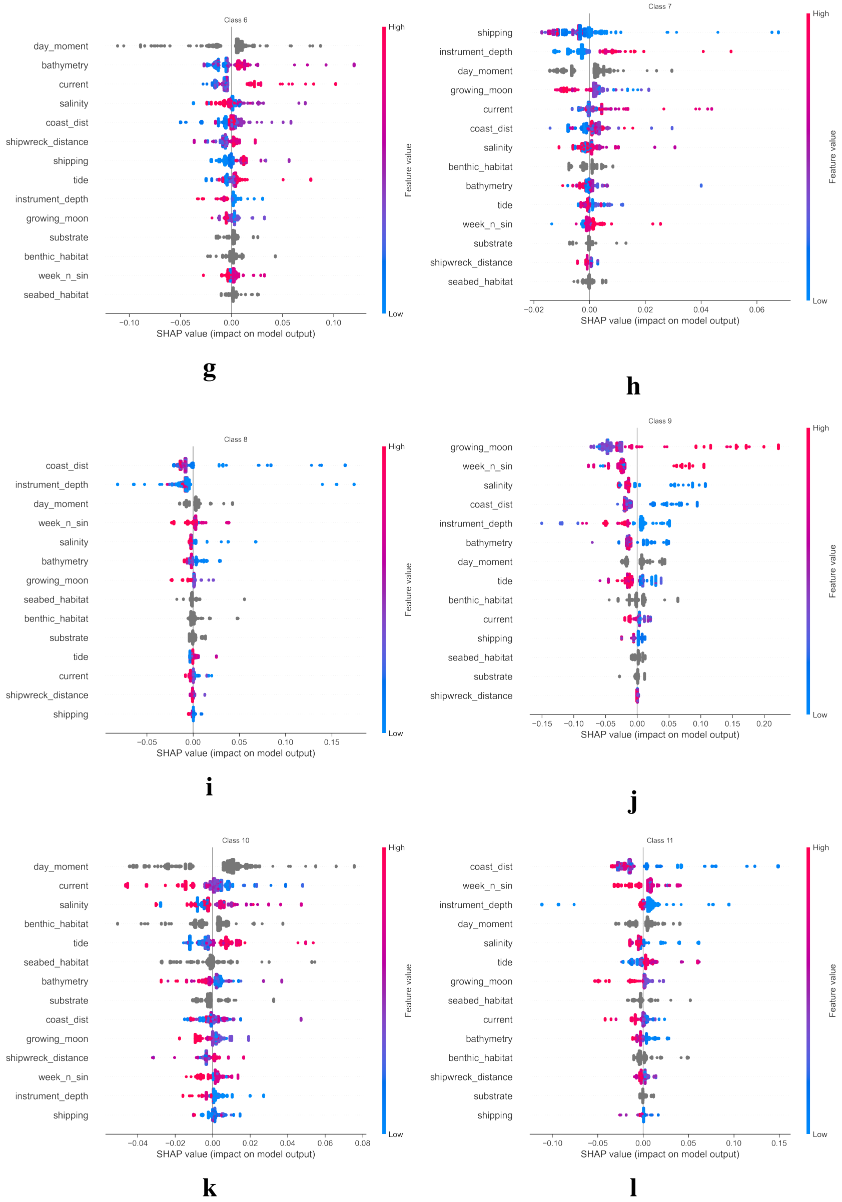


Figure S5.1 Summary of the SHAP values for all the features for class: a) 0 b) 1 c) 2 d) 3 e) 4 f) 5 g) 6 h) 7 i) 8 j) 9 k) 10 l) 11 m) 12 n) 13 o) 14 p) 15 q) 16.

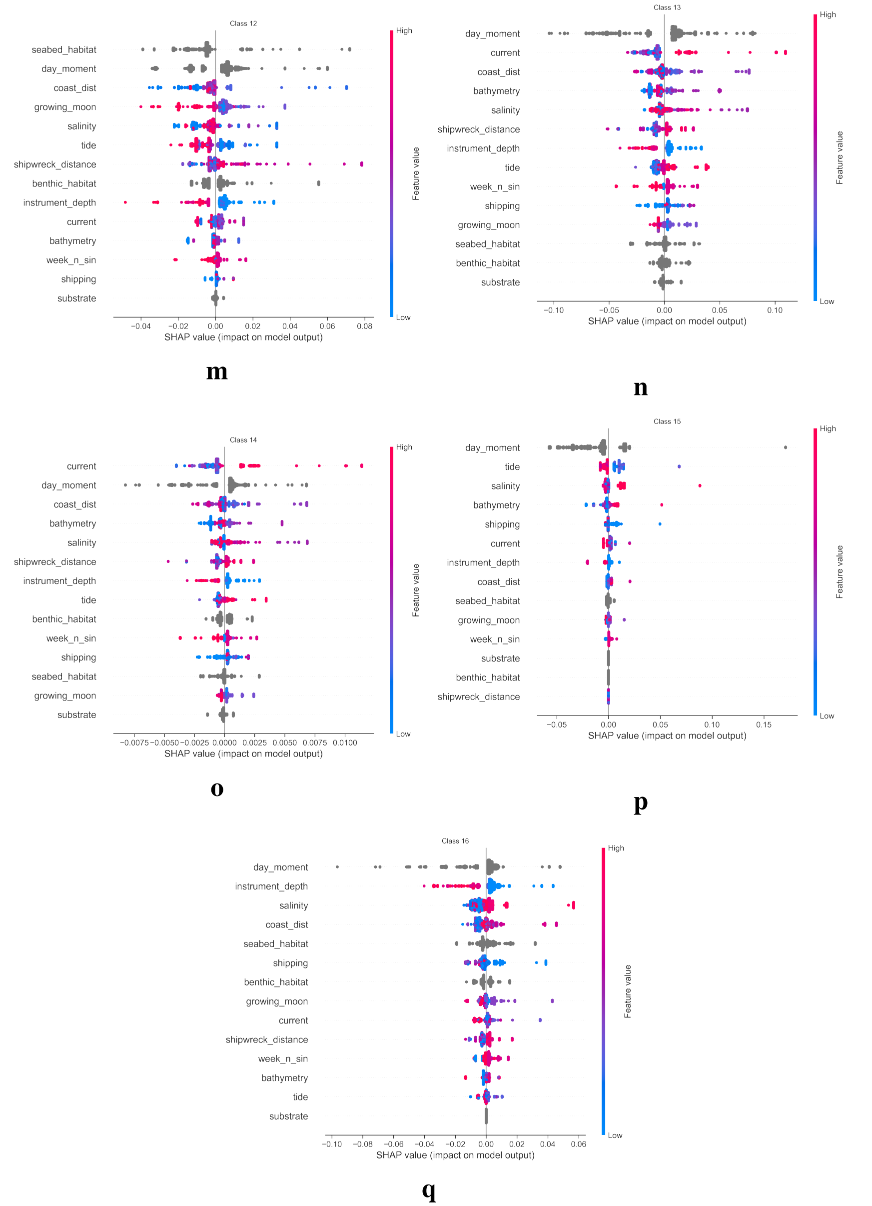
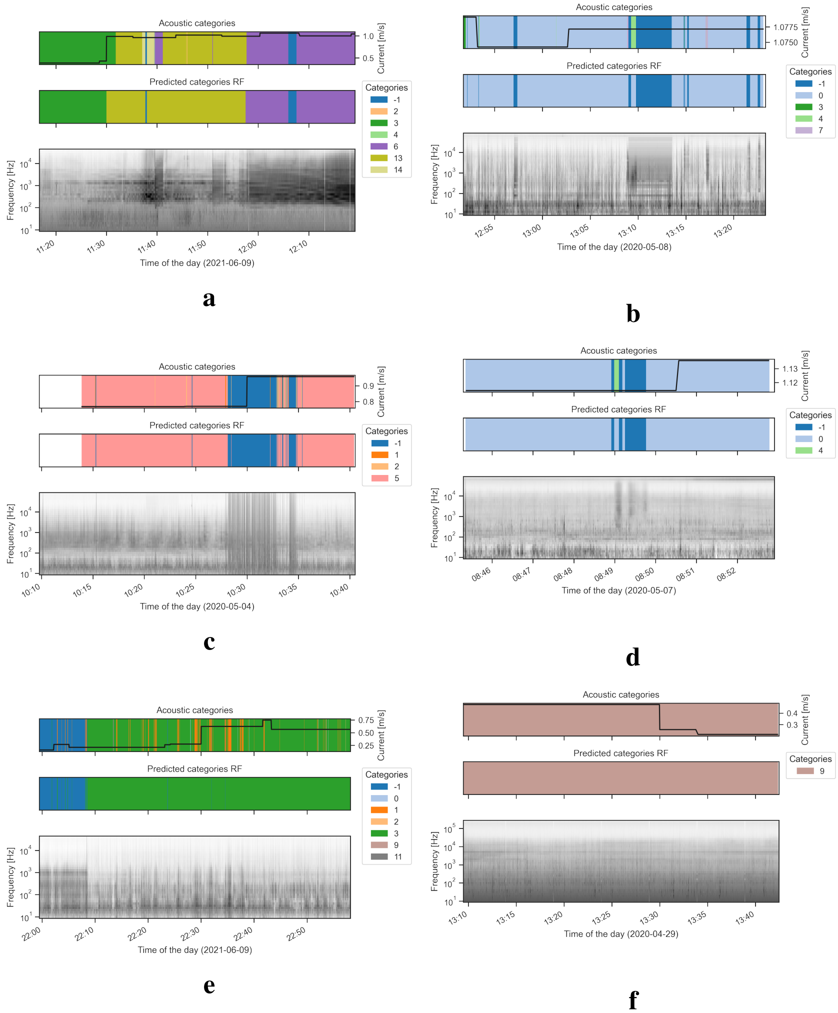


Figure S5.2 Examples of deployments evolution in time, recorded on the a) 9 June 2021 b) 8 May 2020 c) 4 May 2020 d) 7 May 2020 e) 9 June 2021 f) 29 April 2020



S6 Supplementary Material Chapter 6

S6.1 Tables

Table S6.1: Parameters used to run the Yolov8 model for the Miller Dataset. HP stands for hyperparameter

Name	Value	Description	Application
epochs	200	Number of epochs to train for	Train
patience	20	epochs to wait for no observable improvement for early stopping of training	Train
batch	32	number of images per batch	Train
imgsz	640	input images size as int for train and val modes	Train
pretrained	True	whether to use a pretrained model (bool)	Train
optimizer	auto	optimizer to use, choices=[SGD, Adam, Adamax, AdamW, NAdam, RAdam, RMSProp, auto]	Train
deterministic	True	whether to enable deterministic mode	Train
single_cls	False	train multi-class data as single-class	Train
rect	False	rectangular training if mode='train' or rectangular validation if mode='val'	Train
cos_lr	False	use cosine learning rate scheduler	Train
close_mosaic	10	disable mosaic augmentation for final epochs	Train
amp	True	Automatic Mixed Precision (AMP) training, choices=[True, False], True runs AMP check	Train
fraction	1.0	dataset fraction to train on (default is 1.0, all images in train set)	Train
profile	False	profile ONNX and TensorRT speeds during training for loggers	Train
val	True	validate/test during training	Val/Test
conf	0.1	object confidence threshold for detection (default 0.25 predict, 0.001 val)	Val/Test
iou	0.3	intersection over union (IoU) threshold for NMS	Val/Test
max_det	300	maximum number of detections per image	Val/Test
half	False	use half precision (FP16)	Val/Test

Table S6.1: *Parameters used to run the Yolov8 model for the Miller Dataset. HP stands for hyperparameter*

Name	Value	Description	Application
dnn	False	use OpenCV DNN for ONNX inference	Val/Test
lr0	0.01	initial learning rate (i.e. SGD=1E-2, Adam=1E-3)	HP
lrf	0.01	final learning rate (lr0 * lrf)	HP
momentum	0.937	SGD momentum/Adam beta1	HP
weight_decay	0.0005	optimizer weight decay 5e-4	HP
warmup_epochs	3.0	warmup epochs (fractions ok)	HP
warmup_momentum	0.8	warmup initial momentum	HP
warmup_bias_lr	0.1	warmup initial bias lr	HP
box	7.5	box loss gain	HP
cls	0.5	cls loss gain (scale with pixels)	HP
dfl	1.5	dfl loss gain	HP
pose	12.0	pose loss gain	HP
kobj	1.0	keypoint obj loss gain	HP
label_smoothing	0.0	label smoothing (fraction)	HP
nbs	64	nominal batch size	HP
hsv_h	0.0	image HSV-Hue augmentation (fraction)	HP
hsv_s	0.0	image HSV-Saturation augmentation (fraction)	HP
hsv_v	0.0	image HSV-Value augmentation (fraction)	HP
degrees	0.0	image rotation (+/- deg)	HP
translate	0.1	image translation (+/- fraction)	HP
scale	0.5	image scale (+/- gain)	HP
shear	0.0	image shear (+/- deg)	HP
perspective	0.0	image perspective (+/- fraction), range 0-0.001	HP
flipud	0.0	image flip up-down (probability)	HP
fliplr	0.5	image flip left-right (probability)	HP
mosaic	1.0	image mosaic (probability)	HP
mixup	0.0	image mixup (probability)	HP
copy_paste	0.0	segment copy-paste (probability)	HP

S6.2 Others

Algorithm S6.1 Pseudo-algorithm to merge the YOLOv8 detections. j represents a list of the already merged detections. $selected$ are the resulting detections.

```
D is a DataFrame with all the detections
selected ← []
j ← []
for d ∈ D do
    if d ∉ j then
        D_x ← ∀D ∉ j
        compute overlap of each element in D_x with respect to d
        overlaps ← overlaps > 0.5
        if there are overlaps then
            add all overlaps to j
            change d to have the maximum boundaries of overlaps (in freq and time)
        end if
        add d to selected
    end if
end for
```

S7 Supplementary Material Chapter 7

S7.1 Tables

Table S7.1: Summary of all the deployments available from the freshwater dataset. VLR stands for Very Large River, LR for Large River, and MR for medium river. ST stands for SoundTrap. HM for Hydromoth.

Location	River	Water type	Water sub type	Hydrophone	Lat	Lon	Depl. start	Depl. end
Amersfoort	Valleikanaal	MR	Rural	ST S6381	52.12	5.44	3/22/2022	3/31/2022
Amersfoort	Valleikanaal	MR	Bridge	ST S6380	52.15	5.41	3/22/2022	3/31/2022
Amersfoort	Eem	LR	Bridge	ST S6383	52.17	5.37	3/22/2022	3/31/2022
Amersfoort	Eem	LR	Rural	ST S6382	52.18	5.33	3/22/2022	3/31/2022
Barendrecht	Rhine	VLR	Stretch	ST S6380	51.83	4.54	6/2/2022	6/2/2022
Bordeaux	Garonne	VLR	Bridge	ST 6383	44.88	-0.54	4/13/2023	4/15/2023
Caumont	Garonne	VLR		ST 6383	44.45	0.19	4/14/2023	4/17/2023
Damnatz	Elbe	VLR	Stretch	ST S6383	53.14	11.18	6/7/2022	6/8/2022
Duisberg	Rhine	VLR		ST 6382	51.35	6.66	5/29/2023	6/1/2023
Emmerick	Rhine	VLR		ST 6383	51.82	6.27	5/31/2023	6/2/2023
Gorinchem	Rhine	VLR	Stretch	ST S6380	51.83	4.97	6/2/2022	6/2/2022
Hamburg	Elbe	VLR	Estuary	ST S6383	53.60	9.56	6/9/2022	6/10/2022
Leiden 1	Ditch	Ditch	Bridge	HM	52.16	4.46	5/8/2023	5/11/2023
Leiden 1	Ditch	Ditch	no Bridge	HM	52.16	4.47	5/8/2023	5/11/2023
Leiden 2	Ditch	Ditch	Bridge	HM	52.16	4.48	5/14/2023	5/17/2023
Leiden 2	Ditch	Ditch	no Bridge	HM	52.15	4.47	5/14/2023	5/17/2023
Leiden 3	Ditch	Ditch	Bridge	HM	52.15	4.52	5/29/2023	6/1/2023

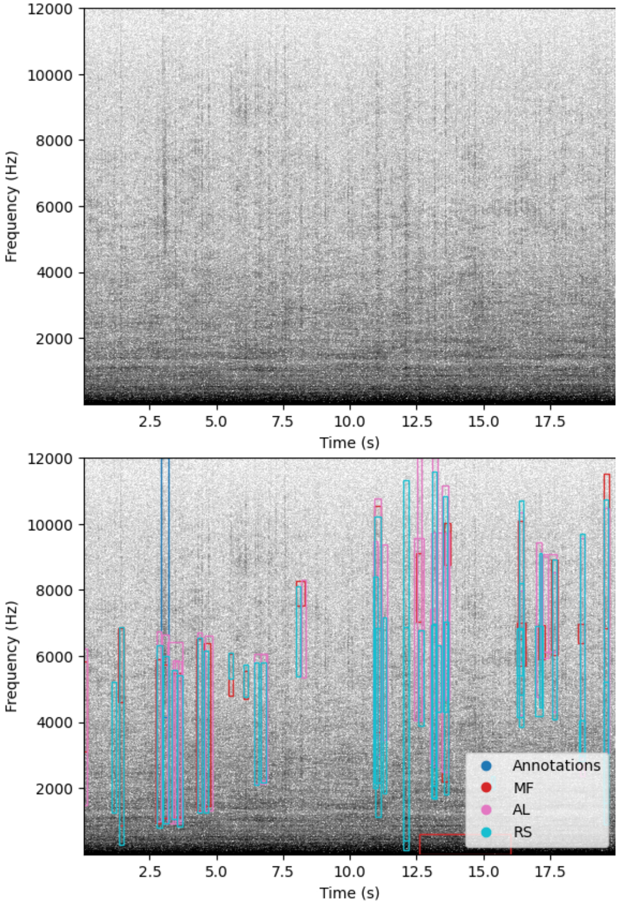
Leiden 3	Ditch	Ditch	no Bridge	HM	52.15	4.51	5/29/2023	6/1/2023
Magdenburg	Elbe	VLR	Spawining grounds	ST S6381	52.11	11.64	6/7/2022	6/8/2022
Meilhan	Garonne	VLR		ST 6381	44.52	0.04	4/16/2023	4/17/2023
Millingen	Rhine	VLR	Distributary	ST S6380	51.87	6.04	5/23/2022	5/23/2022
Niederkassel	Rhine	VLR		ST 6383	50.81	7.03	5/28/2023	5/30/2023
Oegstgeest 5	Ditch	Ditch	Bridge	HM	52.20	4.48	6/4/2023	6/7/2023
Oegstgeest 5	Ditch	Ditch	no Bridge	HM	52.19	4.48	6/4/2023	6/7/2023
Oegstgeest 6	Ditch	Ditch	Bridge	HM	52.19	4.48	6/6/2023	6/9/2023
Oegstgeest 6	Ditch	Ditch	no Bridge	HM	52.19	4.47	6/6/2023	6/9/2023
Oegstgeest 7	Ditch	Ditch	Bridge	HM	52.20	4.47	6/11/2023	6/14/2023
Oegstgeest 7	Ditch	Ditch	no Bridge	HM	52.19	4.48	6/11/2023	6/14/2023
Oegstgeest 8	Ditch	Ditch	Bridge	HM	52.18	4.49	6/13/2023	6/16/2023
Oegstgeest 8	Ditch	Ditch	no Bridge	HM	52.18	4.49	6/13/2023	6/16/2023
Oud-Beijerland	Rhine	VLR	Distributary	ST S6380	51.80	4.62	5/31/2022	5/31/2022
Quinsack	Garonne	VLR		ST 6381	44.74	-0.48	4/12/2023	4/15/2023
Sliedrecht	Rhine	VLR	Stretch	ST S6380	51.82	4.75	5/12/2022	5/12/2022
Veenendaal 1	Valleikanaal	MR	Rural	ST S6381	51.98	5.62	3/8/2022	3/16/2022
Veenendaal 2	Valleikanaal	MR	Bridge	ST S6383	52.02	5.57	3/8/2022	3/16/2022
Veenendaal 3	Valleikanaal	MR	Bridge	ST S6380	52.03	5.54	3/8/2022	3/16/2022
Veenendaal 4	Valleikanaal	MR	Rural	ST S6382	52.04	5.53	3/8/2022	3/16/2022
Vijver Plasmolen 2	Pond	Pond		ST S6380	51.74	5.91	3/23/2023	3/30/2023
Werkendam	Rhine	VLR	Distributary	ST S6380	51.82	4.89	5/17/2022	5/17/2022
Zwijndrecht	Rhine	VLR	Distributary	ST S6380	51.80	4.62	5/26/2022	5/26/2022
le Fleix	Dordogne	VLR		ST 6381	44.87	0.26	4/17/2023	4/20/2023
st Seurin	Dordogne	VLR		ST 6383	44.83	0.06	4/20/2023	4/22/2023

Table S7.2: Description of the resulting clusters and their acoustic parameters. D10 is the 10th percentile of the duration, and D90 the 90th percentile, both in seconds.

Cluster	**Description**	**Possible source**	Mean Low Freq [Hz]	Mean High Freq [Hz]	D10 [s]	D90 [s]
0	Middle harmonic of class 6 between 8 and 10 kHz	Unknown	8439	9351	0.3	0.6
1	Low harmonic of class 7 between 2 and 3 kHz	Unknown	1869	2859	0.3	1.0
2	Repetitive sound, probably from fish. Jackhammer	Biological	94	4823	0.4	1.6
3	Broadband and impulsive wooden scratches/pocs, with multiple repetitions.	Unknown/ Pseudo-noise	985	6248	0.3	0.9
4	Low harmonic of class 6 at 4 kHz.	Unknown	3599	4722	0.3	0.9
5	Medium and high harmonics of class 6.	Unknown	7069	10606	0.4	1.0
6	Squeaking chain sound, with harmonics from 4 to 10 kHz, in general longer than class 7.	Unknown	3223	9952	0.4	1.2
7	Jingle bell-like sound with harmonics from 1.5 to 8 kHz, lasting 1 to 3 seconds, often with an impulsive start	Unknown	1496	9761	0.4	1.2

S7.2 Figures

Figure S7.1 Example of annotations and predictions on a noisy environment. Top image shows a 20 second spectrogram with no annotations. Bottom image represent the manual annotations and the different model predictions. MF stands for Model Final, AL for Active Learning and RS for Random Selection. Annotations refers to manual annotations.



S8 Supplementary Material Chapter 8

S8.1 Tables

Table S8.1 Details of labeled recordings from the LifeWatch Broadband Acoustic Network in the Belgian part of the North Sea

Station	Number annotations	Date	Labeled time
Belwind	11	22-07-2022	00:11:09
Birkenfels	1429	17-03-2021	05:41:42
Birkenfels	328	17-03-2021	00:10:12
Birkenfels	1160	18-03-2021	16:49:48
Buitenratel	8	14-05-2022	00:10:13
Grafton	8	18-03-2021	21:49:04
Grafton	5	20-04-2021	01:21:49.
Grafton	5	12-05-2021	03:44:59
Grafton	47	19-05-2021	17:44:21
Grafton	2	02-08-2021	01:11:35
Grafton	890	07-11-2022	17:41:37

Table S8.2 The preset settings used to process the collected audio data in Raven Pro

Paging	The paging parameters were set with a page size of 10 seconds, employing a 90%-page increment and a 10%-step increment
Spectrogram configuration	Spectrogram settings were preconfigured to display a window duration of 10 seconds, spanning frequencies from 0 to 24 kHz (with an upper limit of 48 kHz)
Color scale	set to 'Grayscale' to facilitate optimal visualization
Additional spectrogram parameters	Additional settings included a lighting level of 4, contrast set at 71, and a view size configured to 1992 points. A Hann window was used with a 75% overlap, and a Discrete Fourier Transform (DFT) size of 2048 samples was utilized

Table S8.3 For the training of the CAE, all parameters chosen were the defaults specified in the code provided by Best et al., (2023) [1]

Learning rate	0.00001
Weight decay	0.0
Batch size	64
Epochs	100
Optimizer	AdamW

Table S8.4 Results of sPCA of AVES- and CAE-extracted datasets, showing the values of alpha embedded in the grid search and resulting number of principal features.

Dataset	Alpha	No. of principal features
AVES-mean	15	30
	18	21
	20	18
AVES-max	35	34
	45	22
	55	13
CAE-crops	32	31
	34	24
	37	17
CAE-original	6	30
	8	20
	10	16

Table S8.5 Number of samples, mean and standard deviation (S.D.) of the 104 grid search results per feature set.

Features set	Number of samples	Scoring metric	Mean	S.D.
AVES-max	12	homogeneity	0.338	0.048
		DBCV	0.347	0.091
AVES-mean	30	homogeneity	0.722	0.053
		DBCV	0.414	0.053
CAE-crops	31	homogeneity	0.265	0.085
		DBCV	0.472	0.118
CAE-original	31	homogeneity	0.551	0.063
		DBCV	0.425	0.126

Table S8.6 Generalized linear model (GLM) fitted with a Gamma distribution showing the significance (** $p < 0.01$, * $p < 0.05$) of association of configured parameters and their interactions to homogeneity and DBCV scores.

glm(homogeneity score ∼ (Number of features + Minimum cluster size + Minimum samples)2, family = Gamma(link = "identity"))		
Coefficient	**p-value**	**significance**
Number of features	0.003	**
Minimum cluster size	0.028	*
Minimum samples	0.239	
Number of features:minimum cluster size	0.001	**
Number of features:minimum samples	0.161	
Minimum cluster size:minimum samples	0.549	
glm(DBCV score ∼ (Number of features + Minimum cluster size + Minimum samples)2, family = Gamma(link = "identity"))		
Coefficient	**p-value**	**significance**
Number of features	0.415	
Minimum cluster size	0.967	
Minimum samples	0.654	
Number of features:minimum cluster size	0.813	
Number of features:minimum samples	0.45	
Minimum cluster size:minimum samples	0.632	

S8.2 Figures

Figure S8.1 Residuals versus fits (left) & quantile-quantile (Q-Q; right) plots of residuals of the generalized linear models (GLM) for homogeneity scores (top) and DBCV scores (bottom). Residuals versus fits plot shows homogeneous variance while the Q-Q plot of deviance residuals shows an approximately good fit.

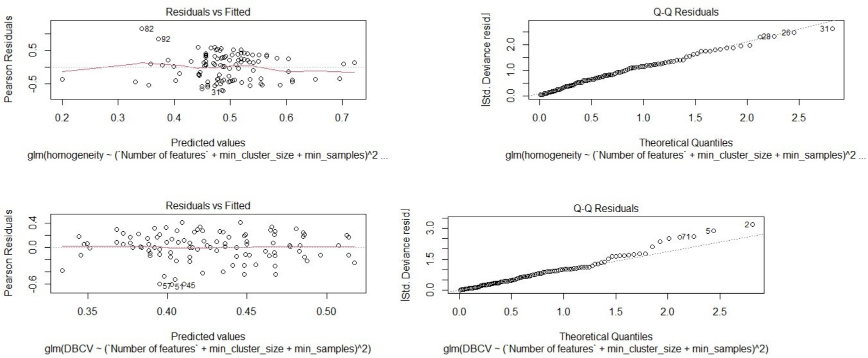
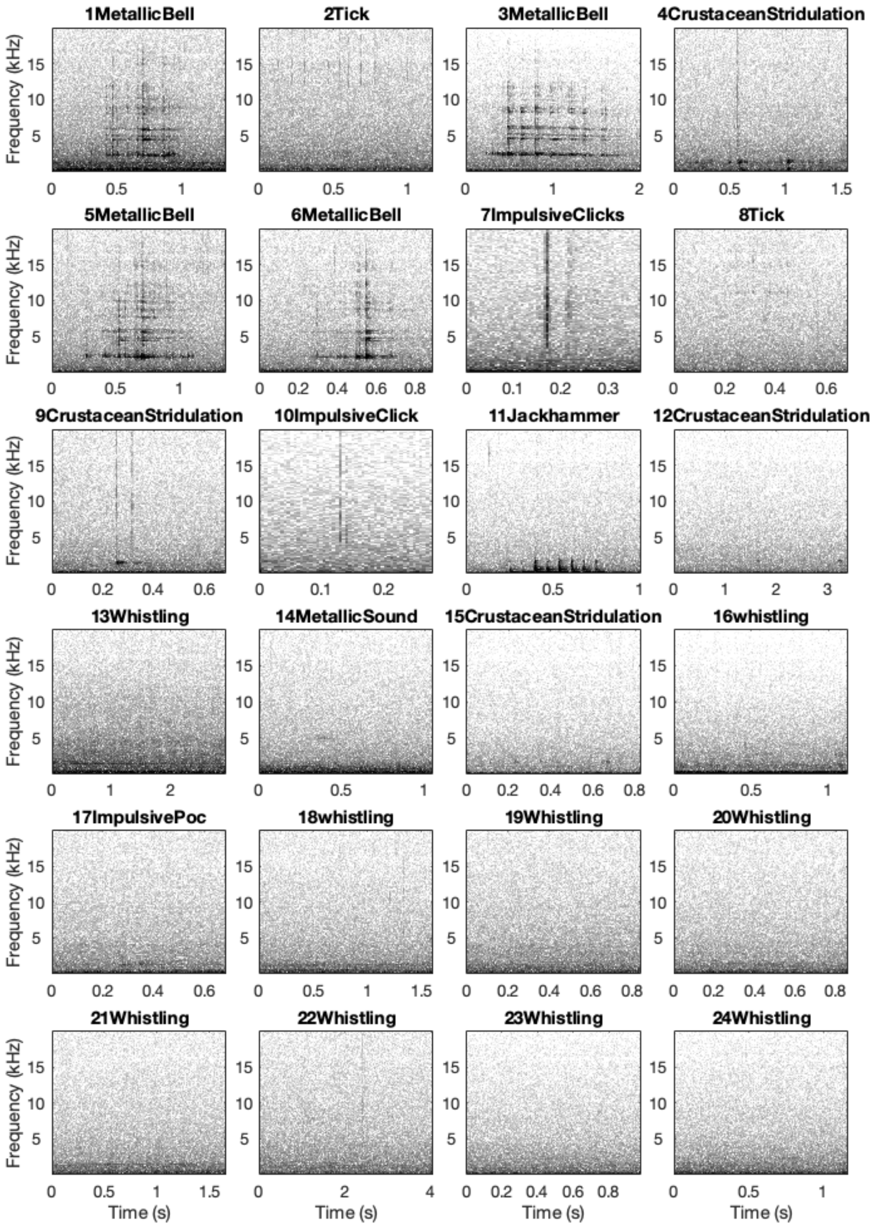


Figure S8.2 Spectrograms of all clusters from the best grid search result which had the highest average of homogeneity and DBCV scores. Each spectrogram is labeled by the cluster # and the representative sound tag which had the highest resemblance to the other sounds from the same cluster.



S9 Supplementary Material Chapter 9

S9.1 Instrument Type

Table S9.1 Description of all instruments used to collect acoustic data. sens. is for sensitivity.

Instrument type	**Treatments**	**Model**	**Serial Number**	**End to end sens. [dB re. 1 μPa/V]**	**Sampling Freq. [kHz]**
SoundTrap	Reef day 3 and 4	ST300 STD	6045	-176	24
SoundTrap	Reef day 1, off-reef day 2	ST300 STD	6046	-176.7	24
ST	REEF_2	ST300 STD	6049	-177.1	24
SoundTrap	OFF_1	ST300 STD	6042	-176.8	24
RTSys	OFF_3, VESSELS_3	RESEA + Colmar GP1190M-LP	EA-SDA14_2003003 GP1190M 134	-176.98	96
RTSys	VESSELS_1	RESEA + Colmar GP1190M-LP	EA-SDA14_2003001 GP1190M 130	-176.98	48
RTSys	VESSELS_2	RESEA + Colmar GP1516M-LP	EA-SDA14_2003003 GP1516M 185	-164.98	48
RTSys	VESSELS_4	RESEA + Colmar GP1516M-LP	EA-SDA14_2003002 GP1516M 191	-164.9	48

S9.2 Sound Treatments

Table S9.2 Different treatments with their corresponding acoustic data collection information. For each day of experiment (Day column in the Table), the reef + vessel combination was done by synthetically mixing the vessel and the reef recordings using Audacity. Mom. stands for Moment. CT stands for Civil Twilight. AT for Astronomical Twilight.

Treatment	Location	Lat	Lon	Depth	Day	Description	Datetime	Mom.
Reef	Nieuweschild	53.07	4.88	X	1	Two times 10 min at different moments of twilight	07/06/2021 02:40	CT
							07/06/2021 05:20	Day
Reef	Kornwerderzand	53.09	5.20	1.35	2	10 min	30/05/2022 20:20	CT
Reef	Nieuweschild	53.07	4.88	X	3,4	20 consecutive min	07/06/2022 19:00	Day
Off reef	Nieuweschild	53.07	4.88	X	1	Sand area close to the reef. 20 consecutive min	13/04/2021 20:39	AT
Off reef	Vlieland	53.22	5.01	1.34	2	Control for artificial reef experiment. 10 min	29/09/2022 06:19	Day
Off reef	Fairplay	51.17	2.62	14.5	3	1 h of continuous off reef sound	06/06/2021 04:19	Day
							06/06/2021 04:42	Day
							06/06/2021 04:54	Day
Off reef	Fairplay	51.17	2.62	14.5	4	30 min continuous, with one distant vessel removed	08/06/2021 15:46	Day
Vessels	Faulbaums	51.33	2.51	15.2	1	3 different vessels: passing close by (11min), short trawling event (4min), loud and long anthropogenic sound (20min)	08/06/2022 14:34	Day
							08/06/2022 14:44	Day
							06/06/2022 17:03	Day
Vessels	Grafton	51.41	2.82	19.14	2	1h20min of continuous boat sounds.	06/05/2022 14:39	Day
							06/05/2022 15:31	Day
							06/05/2022 15:51	Day
Vessels	Fairplay	51.17	2.62	14.5	3	3 different vessels passing far (2, 5, and 5 min)	06/06/2021 03:04	CT
							06/06/2021 03:51	Day
							06/06/2021 04:31	Day
Vessels	Buitenratel	51.24	2.50	6.93	4	Two different boats, one passing by and one close by anchored with constant sound.	14/05/2022 18:55	Day
							10/05/2022 17:13	Day

S9.3 Acoustic Features

Table S9.3 Computed acoustic features and their parameters. Broadband refers to all the frequencies from 0 to the Nyquist frequency (24 kHz). nfft stands for the length of the Fast Fourier Transforms to use.

Metric	Description	Frequency band [Hz]	Parameters	Ref.
SPL	Root mean squared value	broadband		[2]
PSD	Spectrum	broadband	nfft=4096	[2]
Low freq.	Average power spectral density	[0, 1000]	nfft=4096	[2]
Mid freq.	Average power spectral density	[1000, 5000]	nfft=4096	[2]
High freq.	Average power spectral density	[5000, 10000]	nfft=4096	[2]
ACI	Acoustic Complexity Index. Expresses the changes in amplitude in time within a frequency band. Quantifies the acoustic irregularity and variability.	broadband	Hann window nfft=4096 overlap=0.5	[3]
ADI	Acoustic Diversity Index. Quantifies the evenness across frequency bands. A high value would be given if all the frequency bands have the same level, and a low value if one frequency band concentrates all the energy.	broadband	Hann window nfft=4096 overlap=0.5	[4]
AEI	Acoustic Evenness Index. The opposite than ADI. Higher values indicate bigger unevenness in spectral distribution.	[0, 20000]	bin_step=500 dB_th=-50	[4]

$$ACI_{(\Delta fl)} = \sum_{i=1}^{m} \frac{\sum_{k=1}^{n} |I_k - I_{k+1}|}{\sum_{k=1}^{n} I_k} \tag{11.1}$$

$$ACI = \sum_{l=1}^{S} ACI_{(\Delta fl)} \text{ for:} \Delta f = \sum_{l=1}^{S} \Delta f_l \tag{11.2}$$

Where[1],
n is the number of intensity values in a temporal step i,
m is the number of i (temporal steps) in the entire recording, and
S is the number of Δf_l (frequency bands).

$$ADI = -\sum_{j=1}^{S} p_j \ln p_j \tag{11.3}$$

Where[2],
p_j is the fraction of cells (time-frequency) in each jth frequency bands which exceeds a certain threshold (-50 dB as default value),
and S is the number of frequency bands.

$$AEI = \frac{\sum_{i=1}^{S} (2i - S - 1)x_i}{n \sum_{i=1}^{S} x_i} \tag{11.4}$$

Where[3],
i is the rank,
S is the number of frequency bands, and
x_i is the ith proportion of values exceeding a certain threshold (-50 dB by default).

S9.4 Speaker Assignment

The assignment of which speaker and which treatment would be used at each tank per batch was done randomly, resulting in the combinations listed in Table S9.4.

S9.5 Playback Measurements

Prior to the experiment, we conducted recordings of white noise at all the jar positions to assess the differences in sound levels received at each jar. The jars were placed in a way that all of them except jar 3 were at the same distance and position from the speaker. The received PSD at each jar is very similar (see Figure S9.1).

[1] For more information on how to compute these indices, please see Bradfer-Lawrence et al. (2024) [5] and the corresponding web site `https://ecohack.shinyapps.io/Acoustic_Index_Users_Guide`

[2] See footnote 1

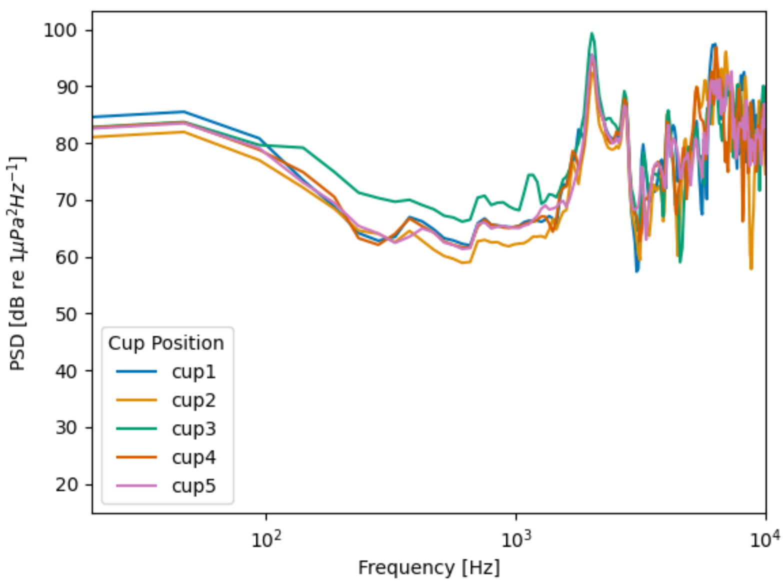
[3] See footnote 1

Table S9.4 Distribution of treatment per tank and speaker. OFF stands for Off Reef, while R+V stands for Reef + Vessel.

Day		Tank1	Tank2	Tank3	Tank4	Tank5
1	Playback	NA	Vessel	OFF	Reef	R+V
	Speaker	NA	3	2	1	4
2	Treatment	Reef	NA	Vessel	R+V	OFF
	Speaker	2	NA	4	1	3
3	Playback	Vessel	OFF	R+V	NA	Reef
	Speaker	1	4	2	NA	3
4	Playback	R+V	Reef	NA	OFF	Vessel
	Speaker	3	4	NA	1	2

Nevertheless, we did not exclude these acoustic differences from having an impact on the settlement and for this reason jar position was investigated using a GLMER model, as described in the statistical analysis section.

Figure S9.1 Power spectrum density received at all the jars when playing a white noise sound.



S9.6 Statistical Analysis

Raw data and statistical analysis are all available on the GitHub repository https://github.com/sschmidlin/larvae-and-sound

Table S9.5 Description and output of statistical models used to determine if random variables had any effect to the base model. Effect stands for the effect assessed. p is the p-value, Pr(>Chisq)

Effect	Model	npar	AIC	BIC	logLik	dev.	Chisq	Df	p
Speaker effect	Base model	10	960.6	1006.5	-470.3	940.6			
	Incl. speaker	13	964.2	1023.9	-469.1	938.25	2.3471	3	0.5
Tank effect	Base model	10	1196.8	1244.9	-588.4	1176.8			
	Incl. tank	14	1203.9	1271.2	-587.9	1175.9	0.9327	4	0.9
Jar position	Base model	9	902.3	942.87	-442.1	884.29			
	Incl. jar position	13	906.1	964.73	-440.1	880.11	4.1733	4	0.4

References

[1] P. Best, S. Paris, H. Glotin, and R. Marxer. *Deep Audio Embeddings for Vocalisation Clustering*. PLOS ONE, 18(7):e0283396, July 2023.

[2] International Organization for Standardization. *ISO 18405:2017*. Technical report, International Organization for Standardization, 2017.

[3] N. Pieretti, A. Farina, and D. Morri. *A New Methodology to Infer the Singing Activity of an Avian Community: The Acoustic Complexity Index (ACI)*. Ecological Indicators, 11(3):868–873, May 2011.

[4] L. J. Villanueva-Rivera, B. C. Pijanowski, J. Doucette, and B. Pekin. *A Primer of Acoustic Analysis for Landscape Ecologists*. Landscape Ecology, 26(9):1233–1246, November 2011.

[5] T. Bradfer-Lawrence, B. Duthie, C. Abrahams, M. Adam, R. J. Barnett, A. Beeston, J. Darby, B. Dell, N. Gardner, A. Gasc, B. Heath, N. Howells, M. Janson, M.-V. Kyoseva, T. Luypaert, O. C. Metcalf, A. E. Nousek-McGregor, F. Poznansky, S. R. P.-J. Ross, S. Sethi, S. Smyth, E. Waddell, and J. S. P. Froidevaux. *The Acoustic Index User's Guide: A Practical Manual for Defining, Generating and Understanding Current and Future Acoustic Indices*. Methods in Ecology and Evolution, n/a(n/a), 2024.