2020

APPLICATION OF AUTONOMOUS UNDERWATER VEHICLES TO THE STUDY OF DEEP-SEA BENTHIC ECOLOGY

Nils Piechaud

Recommended Citation

Piechaud, N. (2020) *APPLICATION OF AUTONOMOUS UNDERWATER VEHICLES TO THE STUDY OF DEEP-SEA BENTHIC ECOLOGY.* Thesis. University of Plymouth. Retrieved from https://pearl.plymouth.ac.uk/bms-theses/427

PHD

APPLICATION OF AUTONOMOUS UNDERWATER VEHICLES TO THE STUDY OF DEEP-SEA BENTHIC ECOLOGY

Piechaud, Nils

Award date:
2020

Awarding institution:
University of Plymouth

All content in PEARL is protected by copyright law.

The author assigns certain rights to the University of Plymouth including the right to make the thesis accessible and discoverable via the British Library's Electronic Thesis Online Service (EThOS) and the University research repository (PEARL), and to undertake activities to migrate, preserve and maintain the medium, format and integrity of the deposited file for future discovery and use.

Copyright and Moral rights arising from original work in this thesis and (where relevant), any accompanying data, rests with the Author unless stated otherwise*.

Re-use of the work is allowed under fair dealing exceptions outlined in the Copyright, Designs and Patents Act 1988 (amended), and the terms of the copyright licence assigned to the thesis by the Author.

In practice, and unless the copyright licence assigned by the author allows for more permissive use, this means,

That any content or accompanying data cannot be extensively quoted, reproduced or changed without the written permission of the author / rights holder

That the work in whole or part may not be sold commercially in any format or medium without the written permission of the author / rights holder

* Any third-party copyright material in this thesis remains the property of the original owner. Such third-party copyright work included in the thesis will be clearly marked and attributed, and the original licence under which it was released will be specified . This material is not covered by the licence or terms assigned to the wider thesis and must be used in accordance with the original licence; or separate permission must be sought from the copyright holder.

Download date: 28. Oct. 2024

UNIVERSITY OF PLYMOUTH

APPLICATION OF AUTONOMOUS UNDERWATER VEHICLES TO THE STUDY OF DEEP-SEA BENTHIC ECOLOGY

By

Nils Piechaud

A thesis submitted to the University of Plymouth
in partial fulfilment for the degree of

DOCTOR OF PHILOSOPHY

School of Biological and Marine Sciences

September 2019

Abstract

APPLICATION OF AUTONOMOUS UNDERWATER VEHICLES TO THE STUDY OF DEEP-SEA BENTHIC ECOLOGY

Nils Piechaud

Rising anthropogenic pressure in the deep sea prompts concerns for its short and long term conservation, however, it remains mostly unexplored. Effective conservation strategies need to be based on a sound understanding of the target ecosystem, or ecosystems, which is not the case in the deep sea, owing largely to the lack of sufficient data. Autonomous Underwater Vehicles (AUV) could help address several long-standing challenges in the study of deep-sea ecology, thanks to their capacity to efficiently sample this remote environment. This thesis aims to investigate how these vehicles can contribute to the study of deep-sea benthic ecology through applying AUV acquired data (presented in **chapter 2**) to address fundamental questions in deep-sea ecology (**Chapters 3 and 4**), as well as asking how the benefits of AUVs, their capacity to quickly gather data in the form of large numbers of seafloor images, can be fully realised (**Chapter 5**). The research conducted in this thesis suggests AUVs are able to quickly and efficiently obtain representative samples, allowing efficient and statistically robust quantification of the density and diversity of benthic epifauna. They can also successfully detect consistent structure in the fine scale distribution of a model benthic epifaunal species (*Syringammina fragilissima*). However, the AUV derived dataset, including high resolution data on environmental variables, failed to clearly establish the environmental parameters driving this distribution. This suggests that although AUVs are capable of gathering large high-resolution datasets, the number of data-points is not the only important criterion for a representative sample. Finally, the application of Computer Vision and Artificial Intelligence methods to the AUV data set demonstrated that useful results can be obtained for some taxa, and the fast development of this technology promises future progress.

Author's declaration

At no time during the registration for the degree of *Doctor of Philosophy* has the author been registered for any other University award without prior agreement of the Doctoral College Quality Sub-Committee. Work submitted for this research degree at the University of Plymouth has not formed part of any other degree either at the University of Plymouth or at another establishment.

Relevant scientific seminars and conferences were regularly attended at which work was often presented; research cruises were undertaken; external institutions were visited for consultation purposes and a paper prepared for publication.

Taught modules attended

- Introduction to Statistical Modelling (Workshop 1) and Flexible Models for Correlated Count or Multi-Category Data (Workshop 2). CREEM (St Andrews) January 2016

The following external institutions were visited for consultation purposes:

- Scottish Association for Marine Science (SAMS). Fieldwork with Freja AUV in Loch Etiv (Visiting John Howe and Karen Bowswarva)

Cruises participation

- RRS James Cook, Wyville Thomson Ridge, Darwin Mounds, Rosemary Bank, Anton Dohrn Seamount, the Rockall Bank, George Bligh Bank, May - June 2016, cruise ref JC136, 6 weeks.
- RRS James Clark Ross, Southern Atlantic, Tristan Da Cuhna EEZ, St Helena. Feb – March 2018, cruise ref JR17004

Publications (or public presentation of creative research outputs):

Presentations at conferences:

- Investigating the effect of depth on species connectivity and distribution in the deep sea. NOC Marine Autonomy & Technology Showcase (Southampton). November 2016
- Identification of benthic epifauna from images using computer vision, Plymsef (Plymouth) February 2018
- Identification of benthic epifauna from images using computer vision, DSBS (Monterey – USA), September 2018

Peer reviewed publication

- Chapter 5 is published as: Piechaud, Nils, Christopher Hunt, Phil F. Culverhouse, Nicola L. Foster, and Kerry L. Howell. (2019) "Automated identification of benthic epifauna with computer vision." *Marine Ecology Progress Series* 615 (2019): 15-30. DOI: https://doi.org/10.3354/meps12925

Word count of main body of thesis: ~42,000

Signed……………………………………………………

Date…03/09/2019..…………………………

Acknowledgments

I would like to particularly thank the crew of NERC ship, R.R.S James Cook for their effort and help in collecting the data used in this thesis making fieldwork an enjoyable experience. I would also like to thank the Autosub support team present on JC136, James Perret, Rachel Marlow, and David Paxton for their hard work to make the vehicle dive and collect data.

This thesis was possible with the help of Kirstin Crombie, Sam Faithful, Chris Hunt, Jamie Cowle, Joe Augier and Phil Culverhouse who I kindly thank for their help in processing and analysing some of the data and comments on some of the manuscripts.

I am also particularly grateful to Daniel Langenkämper, and the Biigle development team for their guidance as well as his eagerness to include feedback within the development of their software.

Also, my grateful thanks go to my deep-sea CRU team-mates, Kirsty McQuaid, Amelia Bridges, Poppy Best, Amber Cobley and Otis Bruner for providing a constructive and enjoyable environment to work in. Special thanks to Rebecca Ross, longest standing member of the CRU for all her advices given over the years. The same goes to my housemates Anaelle and Svenja for leading by example (research-wise). My thanks extend to all past and present members of MBERC particularly my cubicle mates, Mike, Imiy and Poppy.

Big thanks to Chloe Game for reading the many drafts correcting so many typos.

I also thank the many contributors to stackoverflow.com and rbloggers.com who remain uncited but without who this work would have been much harder.

Many (many) thanks to Kerry Howell for the guidance, assistance and patience as well as all the opportunities offered not only during my PhD but throughout the many years of my stay in Plymouth. I will always be grateful.

Many thanks to Nicola Foster for her patience and endless proofing of my half-finished drafts.

Many thanks to Phil Hosegood and John How for their help in the completion of this thesis.

Je remercie également mes parents, Frédérique et Simon pour leur soutient au cours de ce long processus. Cette thèse n’aurait pas pu se faire sans vous.

De temps en temps on tombe sur un farfelu qui croit qu'il a inventé l'eau chaude mais le plus souvent c'est l'adaptation d'un modèle existant.

Venec (Loïc Varraut) Kaamelott, Livre I, 13 : Arthur et la Question, écrit par Alexandre Astier.

Content

Chapter 1: Literature review – Challenges and innovations for deep-sea benthic ecology

Chapter 2: Method - Collection and processing of AUV data

Chapter 3: The effect of sample size on deep benthic soft sediment biodiversity measure.

Chapter 5: Automated identification of benthic epifauna with computer vision

Chapter 6: General discussion and conclusion

List of Tables

List of Figures

Chapter 1: Literature review – Challenges and innovations for deep-sea benthic ecology

1.1 The challenges of deep-sea ecology

The deep sea is commonly defined as the part of the ocean deeper than 200 meters. It is the largest ecosystem on earth covering 65% of the planet's surface (Danovaro et al., 2008, Ramirez-Llodra et al., 2011), up to 95% of the biosphere's volume and could be home to between 0.5 to 10 million species (Higgs and Attrill, 2015).

The study of the deep sea began in the XIXth century first by punctual observations of life at great depth which led to targeted expeditions topped by the voyage of the Challenger (1872–1876), when the existence of diverse life in the deep was irrefutably demonstrated. The following century saw continued exploration of this environment and significant advances in the technology available to the research community to facilitate study (Ramirez-Llodra et al., 2011). With these technological advances came major new discoveries, including the discovery of both gigantic and cryptic species populating the abyss. Given the current rate of species and habitat discovery and description, the deep sea is a virtually infinite source of taxonomic novelty (Snelgrove, 2016a, Costello and Chaudhary, 2017, Costello et al., 2013, Costello et al., 2010).

Although the pace of deep-sea exploration is increasing (Gage and Tyler, 1991, Rex and Etter, 2010, Clark et al., 2016b), this ecosystem is vast. It is unlikely that knowledge of deep-sea ecology will equal that of shallow-water or terrestrial ecology any time soon. The relative "out of sight" state of the deep sea tends to maintain general public interest to a mere distant fascination. However, this hasn't kept commercial activities from expanding into the deep-sea environment (Danovaro et al., 2017a, Van Dover et al., 2017, Van Dover et al., 2014, Van Dover, 2011). Diminishing resources on land will only encourage this movement, subsequently a new chapter in the relationship between man and the greatest wilderness on earth is just beginning.

Deep-sea scientific exploration is ultimately driven by technological development. This technological development has often gone hand-in-hand with commercial exploitation. Although barely explored (Webb et al., 2010, Glover et al., 2010), deep-sea life is facing increasing human pressure and is threatened by a number of anthropogenic activities such as bottom trawling (Clark et al., 2016a), mining (Collins et al., 2013, Vanreusel et al., 2016, Van Dover et al., 2017), plastic pollution (Woodall et al., 2014, Courtene-Jones et al., 2017), oil and Gas exploration (Cordes et al., 2016) and climate change (Mora et al., 2011, Levin and Le Bris, 2015, Balmaseda et al., 2013). These threats to the deep-sea are well identified but their exact effects on the ecosystems are still poorly documented (Van Dover et al., 2014). In this context, a viable and world-wide conservation strategy is urgently required in the world's oceans (Ramirez-Llodra et al., 2011, Barbier et al., 2014, Clark et al., 2016a, Danovaro et al., 2017a, Turner et al., 2017) in order to avoid loss of habitat and species extinctions such as those that have occurred on land (McCauley et al., 2015).

While biologists and managers have now developed strategies for the conservation of specific land and coastal ecosystems (Borja et al., 2016, Pimm et al., 2014), no such strategies are available for the deep-sea ecosystem. Calls for better management of the deep sea, particularly Areas Beyond National Jurisdiction (ABNJ), are multiplying from both the scientific community and civil society (Van Dover et al., 2014, Ban et al., 2013) with some progress already made. Several international initiatives such as resolution 61/105 of the United Nations General Assembly (2003) or the OSPAR Convention (OSPAR., 2008b) encourage nations to identify and map the distribution of vulnerable and / or threatened marine habitats in order to effectively organise their protection. Furthermore, conservation specialists now tend to advocate for a more comprehensive approach to conservation and management, which includes societal preferences and costs (Ban et al., 2013) as well setting priorities after giving relative values to different ecosystems (Marchese, 2014).

The foundations of conservation science that were laid by terrestrial biologists and then exported to the coastal waters and fisheries in the open ocean, are now being adapted to the deep sea (Probert, 2017). It relies on thorough understanding of ecosystem composition and dynamics to assess resistance and resilience, monitor health and predict evolution over time. Practical information provided to policy makers by ecologists comes as interpreted data derived from field observation like vulnerable species occurrence data (Clark et al., 2016b), habitat classifications (Howell et al., 2010b) and representation of extent and distribution of species and communities (Ross and Howell, 2013), often using maps. Interpretation of raw data by ecologists mobilises the fundamental knowledge of the ecosystem, such that the observation can be placed within a wider context and interpreted correctly. In the deep sea, however, the lack of fundamental knowledge renders such endeavours difficult (Van Dover et al., 2014, Danovaro et al., 2014, Anderson et al., 2016, Howell et al., 2016a).

Understanding the fundamental ecology of the deep sea is a major research challenge of the coming century and there is much work left to be done (Higgs and Attrill, 2015, Mora et al., 2011). Illustrations of this state of affairs can be seen in many aspects of deep-sea science. For example, current estimates suggest we have explored between 0.05 - 5% of the deep sea (Danovaro et al., 2017b), which suggests there is much left to discover. Additionally, there are currently more than 243,000 marine species described and estimation of the remaining number of undiscovered species vary, but is probably around 1.5 million (Costello and Chaudhary, 2017). Finally, lacking such fundamental knowledge may hamper conservation effort in particular as it is impossible to establish ecological baselines to which an ecosystem's state can be compared for monitoring environmental health (Crowder and Norse, 2008, Grassle and Maciolek, 1992a, Borja et al., 2016).

The fundamental questions that need to be answered include (but are not limited to): 1) the number of species within a given sampling unit. 2) their spatial dynamic and distribution at fine scale and 3) the relationship with their environment (drivers of distribution. These three elements will be investigated in this thesis for the sessile epibenthic megafauna.

While detailing the other size fractions and components of deep-sea biodiversity will be too much to present here, much of the fundamental knowledge used to formulate the theoretical basis of deep-sea biology are extrapolations of studies based on one functional or taxonomic group, be it sediment infauna, hard substrate epifauna, pelagic mobile megafauna, echinoderms or nematodes. Some of the references used in this thesis are studies focused on some of these other groups and although considered with caution they provide useful insight into ecological phenomenon applicable to benthic epifauna.

1.1.1 The number of species

Although simple in principle, answering the question of how many species live in the deep sea, is a practical challenge (Higgs and Attrill, 2015). The number of species, the alpha diversity, or species richness is the number of species in a specific location but there are other expression of that variable. As diversity is usually measured to be compared between different sites, often over an environmental gradient, the difference in diversity is then referred to as Beta diversity. When multiple sites are aggregated into an ensemble, the resulting regional diversity is referred to as gamma diversity (Begon et al., 2006).

Diversity is the product of the appearance (speciation, or the evolutionary process by which new species appear) and disappearance (extinction) of species over geological time. Speciation takes place when organisms evolve to maximize their fitness to their environment or through genetic drift, when populations no longer exchange genes. Although the mechanisms of speciation are complex, they are commonly split into four main categories: allopatric, peripatric, parapatric and sympatric speciation (Begon et al., 2006). Allopatric speciation is the result of the geographic isolation or separation of two previously connected and interbreeding populations, of the same species. Over time, these separated populations drift away from each other genetically and become two different species. This can be exemplified by the closing of a straight between two oceanic basins, physically isolating populations of the same species living on either side, leading to their eventual divergence into different species. A specific case of allopatric speciation, in which one of the two populations is smaller than the other, is referred to as peripatric speciation. In this circumstance, if the smaller of the populations becomes isolated on an island, thus cut-off from the larger population on the continent, it can undergo speciation. Parapatric speciation, however, occurs when two populations of the same species live in different environments whilst remaining in contact and maintaining a geneflow, albeit reduced. Rather than isolation leading to a gradual drift over time, this speciation mechanism operates through natural selection with two populations evolving to occupy different niches, leading to an eventual decline in interbreeding success. This is exemplified in the marine realm by populations occupying different depth bands, diverging into multiple species, while maintaining contact at their respective upper and lower depth ranges. The final mechanism, sympatric speciation, occurs when small populations becomes reproductively isolated from a main population, while remaining within the same geographic area. In this case, the necessary flow of genes to maintain homogeneity of the species is cut by means such as behaviour or dramatic mutation that affects reproduction.

Environmental changes can also influence diversity by triggering range-shifts in the distribution of each species (Parmesan et al., 1999, Gaston and O'Neill, 2004). The effect of climate change illustrates this phenomenon in multiple instances, where species adapted to a certain climate, or a certain range in temperature, see their populations move to areas where these parameters suit their physiology (Sunday et al., 2012, Sweetman et al., 2017). This can be active, in the case of mobile species. For most species unable to migrate face local extinction in areas where they cannot maintain their reproductive fitness, while their population become sustainable in new areas (Pinsky et al., 2013).

In practice, speciation mechanisms in the present environment are hard to distinguish without access to any historical records. Besides, it is never clear when one species becomes two. Furthermore, recent work indicates that speciation may occur without a complete cut of the gene flow between two populations (Bolnick and Fitzpatrick, 2007). In the marine environment, the exact mechanisms of speciation are likely different from land, as genetic isolation is less likely to occur in this fluid environment (Palumbi, 1994, Ingram, 2011). The greater dispersal ability of marine taxa, enabling connectivity over large spatial scales (Hilário et al., 2015), suggests that evolution to fit different niches and subsequent reproductive isolation could therefore be more common (Puebla, 2009). In the deep-sea, the patterns and mechanisms of speciation are, in general, inferred from comparisons with shallow marine or freshwater ecosystems, due to insufficient data (Miglietta et al., 2011). However, these have failed to explain some of the peculiarities of deep-sea biogeography, more research is therefore needed to describe these patterns more clearly. Unlike initially assumed, the deep-sea is not a uniform landscape with low diversity but is, on the contrary, host to a large number of species (Levin et al., 2001, Rex and Etter, 2010). Many theories have been put forward to explain this diversity (Rex and Etter, 2010, McClain and Schlacher, 2015). Some of these theories contradict each other and research on this subject is slow or even neglected. This is partially due to a lack of converted effort to aggregate data, as well as dedicated effort to evolve the conceptual framing of these theories (McClain and Schlacher, 2015). More knowledge of the exact extent of diversity, as well as its spatial variation in relation to environmental drivers, would greatly contribute to the refinement of its understanding.

The vast volume of the deep Ocean implies that its diversity cannot be subject to a census where the ecosystem is systematically thoroughly sampled and every species encountered is recorded until no unknown remains. It is nonetheless certain that the deep sea is highly diverse (Rex and Etter, 2010).

Many attempts have been made to estimate the number of species in the deep sea. Studies yielded numbers from 0.5 to 10 million species, sparking debate on the methodology of calculation (Mora et al., 2011). Global estimations are difficult as it is uncertain how to extrapolate numbers of species from a very limited number of records (Higgs and Attrill, 2015). Early estimates assumed that the number of species encountered at a small but clearly defined location, of a given surface area or volume, could be multiplied by the surface area / volume of the deep ocean, giving the total number of species. This led to the belief that species richness was higher in the deep sea than in shallower habitats and potentially rivalled tropical forests (Grassle, 1989). This view is now questioned but nevertheless, when it became clear that the deep sea was at least of comparable diversity to the better known shallow coastal waters (McClain and Schlacher, 2015, Hessler and Sanders, 1967), theories attempting to explain that diversity were formulated. The first paradigms, postulated by Sanders (1968), explained diversity by the long-term stability of the deep sea where evolutionary processes had the time to adapt to an increasing number of species to a large number of very specialised niches. However, it was quickly acknowledged that the deep sea is not completely stable, regardless of the scale considered, and that high diversity was rather the result of local disturbances creating patches of different stages of post-disturbance succession, each one a specific niche (Grassle and Sanders, 1973). The reasons why diversity is high in the deep sea are still debated (reviewed in Rex and Etter (2010)), but regardless of the cause, the extent of deep-sea diversity still eludes scientists. Regularly, studies point out how poorly sampled the deep sea remains, even in the 21st century (Costello et al., 2010, Gray, 2002). Whether it is regarding taxonomy (Wiklund et al., 2017), genetic diversity (Taylor and Roterman, 2017) sediment macrofauna (Snelgrove, 1998), meiofauna (Gambi and

Danovaro, 2016), benthic epifauna (Durden et al., 2015), or pelagic fauna (Sherman and Smith, 2009).

At broad (regional and global) scales, estimation of the number of deep-sea species must be extrapolated from a limited number of samples (Grassle and Maciolek, 1992a). Thus, estimates of global diversity need to be based on a good knowledge of local diversity, which is estimated with representative samples in which the number of species is known with certainty. However, ensuring that sampling of one location is complete is a challenge in itself. Additionally, the relationship between surface area and species diversity is not linear, nor is it constant over space and time. There are a wide range of habitats and communities with local variations and differences that are not well defined or understood (Howell et al., 2010a) and the species richness of each of these habitats is also uncertain. Finally, deep-sea diversity is also composed of species with large spatial distributions but are locally rare (low density), and are sampled as singletons which, in turn, tends to exclude them from statistical analysis (Turner et al., 2017, Higgs and Attrill, 2015, Grassle and Maciolek, 1992a). Pooling of multiple datasets could circumvent this problem (McClain and Rex, 2015), but the lack of standardised sampling and analysis procedures makes data sharing difficult (Althaus et al., 2015).

It is clear that the drivers of deep-sea diversity are thus complex, and its contemporary understanding has recently been questioned (McClain and Schlacher, 2015). Concerning benthic epifauna, it is also clear that classic sampling methods are either not sampling a sufficient area, and/or that there is not enough replication of sampling being undertaken. Thus, the main challenge regarding diversity of epibenthic fauna is to determine the appropriate sample size needed in order to provide representative assessment of the number of species present within one station. This is not achievable with classic sampling. It needs better sampling tools and a more standardised sampling method. A thorough assessment of the sampling effort needed to encounter the whole diversity of a station could be translated into a sampling guide, which could then be reused in other surveys, making results comparable. Higher and more standardised sampling effort would yield more representative and thus reliable, results and draw a more accurate image of species diversity in the deep ocean, particularly concerning the representation of rare species.

1.1.2 Fine-scale distribution

The fine scale distribution of species and how it varies in space (the spatial turnover) is an observable consequence of ecological and biological drivers. Therefore, it is a key component in understanding ecosystem functioning (Pringle et al., 2010, Zeppilli et al., 2016, Bowden et al., 2016) and diversity (Grassle and Maciolek, 1992a, Rex and Etter, 2010).

Diversity in the deep sea is, in part, a result of environmental pressures and their variation in space and time (or, possibly, lack thereof). At broad, or global, spatial scales, depth (Howell et al., 2002, Wei et al., 2010b, Carney et al., 2005) and longitudinal gradients best explain observed variations in the diversity of deep-sea fauna (Rex et al., 1993, Watling et al., 2013, Snelgrove, 2016b). However, these parameters, which are easily measured, are possibly proxies for more influential environmental drivers, most notably temperature (Yasuhara and Danovaro, 2016) or energy availability (Woolley et al., 2016, Watling et al., 2013). Other known drivers include oxygen availability, nutrients such as silica or nitrates availability, substrate types and interspecies interactions (Ramirez-Llodra and Billett, 2006, Gage, 2002). Habitat heterogeneity created by these environmental gradients has also been recognised as a driver of large scale species diversity (Levin et al., 2001, Durden et al., 2015, McClain et al., 2010).

At a finer scale however, 0.1m-10km or within basin, the relative influence of the various known drivers changes (Rex and Etter, 2010). Energy and food supply remain major drivers but topography, terrain complexity and substrate types also become more significant (Beazley et al., 2013, Rengstorf et al., 2014, Buhl-Mortensen et al., 2010, Tong et al., 2013). The heterogeneity in these parameters is considered responsible for the diversity of ecosystems observed in the deep ocean. Also at this scale, the influence of biological interactions (competition, parasitism or predation) become predominant (Ramirez-Llodra and Billett, 2006, Henry and Roberts, 2017), albeit to a debated extent (McClain and Schlacher, 2015). Deep-sea species are distributed in patches of various sizes as result of the interactions between these different parameters over space and time. The size of these patches is an important topic of research and long known challenge (Jumars and Eckman, 1983, Jumars, 1976, Jumars, 1975b). Indeed, whether the patches are kilometres or meters wide will have a strong influence on the result of sampling by transects 100 meters apart as illustrated by Brattegard and Fosså (1991). Whether species are evenly spaced within such patches or whether they are aggregated within multiple smaller pockets can also greatly influence that same sampling result.

The causes of this spatial turnover of species has puzzled scientists since the 1960’s and many studies have attempted to quantify and explain it (reviewed in Rex and Etter (2010)). Some have found that at a fine scale, deep-sea diversity can be explained by the patch-mosaic theory (Rex and Etter, 2010, Dayton and Hessler, 1972). According to this theory, species diversity should vary at a small scale because different patches are at alternate successional stages. However other studies in the deep sea have found that species distributions rarely deviate from random (reviewed by Rex and Etter (2010)) which is contrary to the patch mosaic theory (McClain et al., 2011) but is consistent with other ecological research results. These have shown that patterns in distribution can emerge within supposedly constant environments, like arid landscapes (Tarnita et al., 2017, Pringle et al., 2010), that such patterns are strongly influenced by the scale at which observations are made and that random distribution is unusual in nature (Taylor et al., 1978).

All these conflicting theories and evidence illustrate the complexity of this question. Studies of spatial turnover are usually focused on one community, one given scale and limited to one sampling gear, aggravating the biases and making formulation of a coherent theory for the turnover of benthic epifauna difficult. This paired with the difficulty to gather the necessary quantity of data required to investigate it, is probably responsible for the lack of progress in finding an answer (Jumars and Eckman, 1983, Morrisey et al., 1992, Gray, 2002). Regardless, the lack of understanding of how communities are structured is a cause of great uncertainty in deep-sea studies. More effort needs to be dedicated to solving the problem in order to properly describe species diversity and correctly extrapolate it over various scales, as well as design appropriate sampling strategies.

1.1.3 Influence of the environment and drivers of species distribution

While univariate measures of species richness are one aspect of deep-sea diversity, developing effective conservation strategies requires an understanding of the distribution of species and assemblages. Ecological research has demonstrated that the environment largely determines what species/assemblage is found at a specific location (Guisan and Zimmermann, 2000). Theoretically, this translates into: knowing the environment equals knowing the species / assemblage associated with it.

Following this principle, it would also mean increased capacity to predict biological changes through time (Clark et al., 2016b). Furthermore, since some environmental parameters are easier to sample than actual biological information (Brown et al., 2011), a better understanding of the relationship between environmental drivers and biological responses could in turn allow for faster and more efficient exploration and monitoring of marine environment (Brown et al., 2011, Leonardsson et al., 2016, Howell et al., 2014b, Bullimore et al., 2013).

There are other practical implications to encourage better understanding of environmental influences on species distributions. Indeed, one of the most popular ways to summarise biological information and communicate such information to biodiversity managers is through mapping (Ferrier, 2002). While mapping can be achieved by visually surveying terrain and reporting species observations on a map, direct observations are difficult to conduct in the marine environment, particularly in the deep sea. To tackle these limitations, predictive modelling is widely used to estimate extent and distribution of species and assemblages (Villero et al., 2016, Vierod et al., 2014, Yesson et al., 2012, Davies, 2012, Ross et al., 2015, Piechaud et al., 2015, Howell et al., 2011) and to identify drivers of species distribution (Brown et al., 2014). These models rely heavily on environmental parameters like topographic or oceanographic variables, many of which can be remotely sampled and therefore are a more efficient way to produce maps (Brown et al., 2011, Merow et al., 2014, Elith and Leathwick, 2009, Elith et al., 2006). But the knowledge of how these parameters are driving distribution of species (both quantitatively and qualitatively) is still scarce (Brown et al., 2014), particularly at fine scale (Rengstorf et al., 2013), potentially leading to wrong assumptions on the nature of relationships between species presence and a particular environmental setting (Kostylev, 2012). The quality of a model is judged by its capacity to make accurate predictions, validated by new observations which were not included in the model training (Franklin, 2010, Elith and Leathwick, 2009). This comparison requires new independently acquired data (Elith et al., 2006, Anderson et al., 2016).

To effectively describe the environmental drivers of species distributions at fine scale, both environmental and occurrence data are needed, from the same time and location. This encompasses multiple challenges, both logistic (operating multiple sensors at the same time and collecting more data) and scientific (processing and integrating data sets of a different nature). More attention needs to be focused on the understanding of fine scale species distribution drivers and particularly epibenthic fauna, and how that knowledge can be applied to predictive modelling. This challenge involves collecting more data of better quality.

1.1.4 Adapting sampling and analysis to meet the challenges

One way to obtain more data to meet these challenges would be to increase the sampling effort with currently available means of investigation. However, this approach is not only logistically unrealistic, it has other limitations. One of the effects of the difficulty of sampling the deep sea is that the statistical rigor in sampling methodology used by land and coastal marine scientists is rarely met in deep sea studies.

Benthic megafauna can be sampled by multiple means (trawls, towed cameras and various underwater vehicles (Clark et al., 2016b) but currently there is no consensus on the area of the seabed that must be covered in order to provide an adequate sample of the epibenthic megafaunal community, either at the scale of an individual station, study site, or at greater scales. The data collected by a research programme may therefore not be representative of even the sampling station’s local conditions and communities. Furthermore, as not enough of the deep sea has been sampled, rare species can remain undetected, and results are rarely replicated (Zeppilli et al., 2016). For these reasons, data collected by these means is considered insufficient (Glover et al., 2010), biased (Higgs and Attrill, 2015) and limits scientists understanding of deep-sea ecology (Rogers et al., 2015). Pooling data from multiple studies could offer a more comprehensive understanding, however this requires methodological standardisation both between and within nations (Althaus et al., 2015, McClain and Rex, 2015). Collectively these issues hamper the ability of ecologists to understand both the distribution and drivers of deep-sea biodiversity, and by consequence how to model it (Rengstorf et al., 2014).

If we are to meet the challenge of designing effective conservation strategies for the deep sea, our capacity to study this environment needs to be up-scaled by several orders of magnitude. The difficulties of reaching depths of >200 meters, by both physical instruments and underwater remote sensors from ships hundreds of kilometres offshore, raises the cost of a single research vessel to around £40,000 per day (Brandt et al., 2016). This high cost results in limited sampling effort, considering the vastness of the deep ocean (Glover et al., 2010, Levin et al., 2019). These difficulties along with the relatively fewer scientists studying deep sea ecology compared to shallower waters, means that there is more work for less people. Therefore, any increase in the amount of data collected and processed will have to be done through innovative ways rather than increased effort with existing methods.

1.2 Autonomous Underwater Vehicles and data collection

In benthic ecology, an important part of the studies on epifauna is the recording of species presence (or absence) on the seabed and it can be achieved in many ways (Clark et al., 2016b). Of these, image-sampling is gaining much popularity among ecologists (Bicknell et al., 2016, Solan et al., 2003, Durden et al., 2016c) and is the primary tool to sample the deep-sea bed, traditionally via platforms attached to a ship. Seabed imaging is subsequently followed by interpretation, analysis or annotation, so that data can be used in statistical analysis, mapping or modelling (Gómez et al., 2016). The acquisition process is complex in the deep sea as it usually involves lowering equipment to the sea-bed, sometimes kilometres deep, and maintaining a link to the mothership above to guide the imaging system, a daunting logistical, technological and financial challenge (Brandt et al., 2016, Ramirez-Llodra et al., 2010). The mobilisation of so much equipment and personnel makes the process slow and limited in its movements.

The development of autonomous systems is now offering the possibility to gather data without supervision or direct guidance and thus making sampling faster and less costly (Dunbabin and Marques, 2012). This process is used both on land with Unmanned Autonomous Vehicles (UAV), where it is starting to revolutionize spatial ecology (Anderson and Gaston, 2013) and in the marine environment, in which sampling based on Autonomous Underwater vehicles (AUV) shows great promise (Wynn et al., 2014, Wynn et al., 2012, Huvenne et al., 2009, Lucieer and Forrest, 2016) after a long development.

1.2.1 Autonomous Underwater Vehicles

The development and utilisation of AUVs stems from the logistical and technological challenge of controlling vehicles to perform complex tasks in the deep ocean. The use of submersible vehicles piloted by humans (HOV) exist, such as *Alvin* from *Woods Hole Oceanographic Institution* or Nautile from *Institut Français de Recherche pour l'Exploitation de la Mer* (Clark et al., 2016b, Jamieson et al., 2013), however they present a challenge in endurance, safety and ultimately costs, that warrant an alternative (Clark et al., 2016b). Pilots can also be left aboard the ship and control vehicles via a cable, as is the case with Remotely Operated Vehicles (ROV) (Clark et al., 2016b). This is the preferred option for most operations in deep-water, although they require the mobilisation of several highly trained technicians and pilots at a time, while the ship has to be fully dedicated to this activity (Ayma et al., 2016, Przeslawski et al., 2018, Jamieson et al., 2013). Overall, these vehicles can perform almost any sampling task but are very limited in their movement and coverage, whilst expensive to operate (Jamieson et al., 2013, Huvenne et al., 2018). Additionally, ROVs do not always provide adequate images for quantitative studies (Jamieson et al., 2013). The cable of these vehicles remains a necessity as sound-based communication, although possible, is highly restricted in terms of the quantity of information transfer. A wireless underwater remote-control system would be impractical over a long distance, as the vehicle cannot send video feeds and receive real-time instructions from pilots on the ship.

The way around these constraints, was to export the decision-making (or alternatively, store instructions) and control systems to the vehicle, along with the power-source. Thus making the vehicle autonomous and able to carry out its mission and return to the ship or base for recovery upon completion, without human input (Williams et al., 2012a). There are several classes AUVs and the nomenclature is not completely formalised in terms of designs, however they can be split between “hovering” and “cruising” (Przeslawski et al., 2018), which are different trade-offs between manoeuvrability and speed. The Hovering AUVs are designed for precision operation and are therefore highly manoeuvrable. They have several propellers and can move in any direction and dimension, even over rough terrain, but at reduced speed (Przeslawski et al., 2018). The cruising AUVs, are usually torpedo shaped, preferably yellow painted and fitted with a propeller at the aft end. Their speed, up to 2 meters per second, allows them to cover much more ground than the hovering class (Przeslawski et al., 2018). They are the most commonly deployed type of AUV for benthic ecological surveys in the deep-sea and will be the focus of this thesis.

Autonomous vehicles make benthic image sampling extremely efficient compared to those vehicles operating from a ship. They can be programmed with their mission parameters and need no further input after launch. They can achieve large coverage in short periods of time, thanks to a greater speed and freedom of movement in water, than other subsurface vehicles (Perkins et al., 2016). Since they need no instruction during a mission, the ship is free to carry out other tasks, further optimising its time and running costs (Huvenne et al., 2009, Wynn et al., 2012, Brandt et al., 2016). Their size can range from a few centimetres to several meters, and as such they often require little deck space and can be operated by a small team of people. Given that no further attention is required during the mission, a 2 to 4 person team is all that is needed to continuously operate the AUV during a cruise, including overnight deployments. Furthermore, the AUVs can be deployed at a station, programmed to travel distances of possibly tens of kilometres, complete their mission there and return to a convenient rendez-vous point for the ship, thus sparing costly ship movements.

They can even survey areas that are not accessible by other means, such as below icecaps (Nicholls et al., 2006)

AUVs used in benthic ecology navigate primarily by dead-reckoning (estimations based on initial position, speed and direction), since satellite based systems are restricted underwater (Paull et al., 2014). This navigation system is considered accurate, particularly with the most recent technologies, but may be subject to “drift” as it accumulates error throughout the mission, that can in some cases, result in a positioning error that cannot be quantified (Paull et al., 2014, Huvenne et al., 2018). Most AUVs designed to survey the sea-bed have an autonomy ranging from several hours up to a day. This makes their deployment compatible with a standard 24 hours mission that integrates in the general flow of a typical research cruise. Their autonomy depends on the parameter of their mission and can be modulated depending on its objectives; some AUVs designed for endurance can even have autonomies of several weeks or months (Hobson et al., 2012, Wynn et al., 2014, Furlong et al., 2012).

The use of AUVs for benthic ecological surveys and image sampling was pioneered by the French submersible “Epaulard” in the 1980s (Riqaud et al., 2004, Sibuet et al., 1989). Some 15 years later, and with significant advances in robotic technology, AUVs have regained popularity (Wynn et al., 2014). They are now increasingly used in ecology, as the number of publications mentioning AUVs and ecology indicates (Figure 1-1). They have been applied to a variety of ecological sampling tasks such as animal tracking (White et al., 2016), plankton sampling (Reisenbichler et al., 2016) and, benthic faunal survey (Sibuet et al., 1989, Singh et al., 2004, Morris et al., 2014, Simon-Lledó et al., 2018).

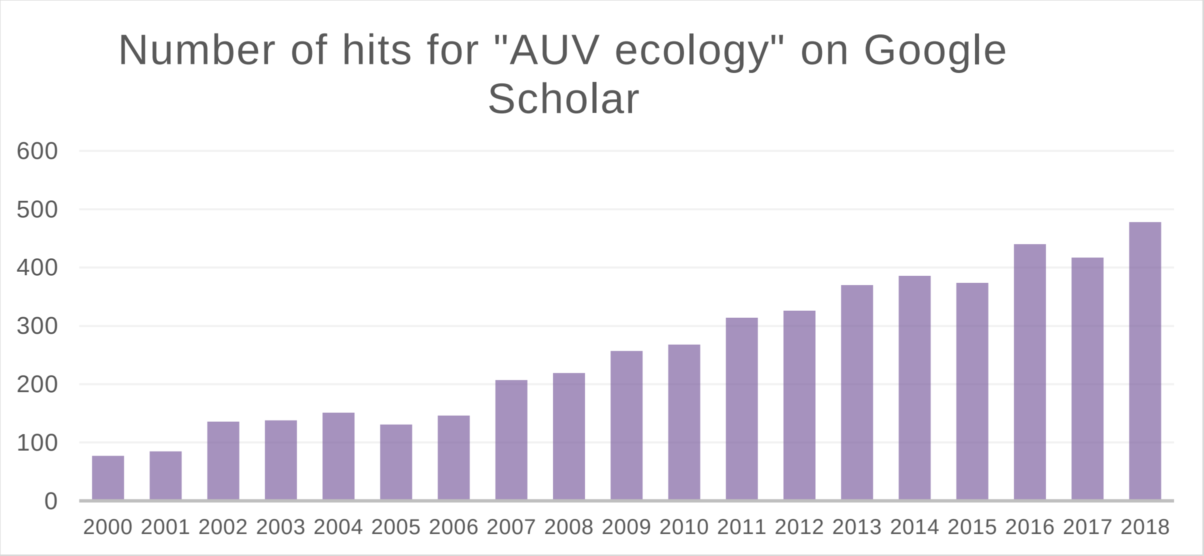


Figure 1-1: Number of results for the search "AUV ecology" on Google Scholar, as an indication of the popularity of the vehicles in the field of ecology since 2000.

One of the particular benefits of these vehicles for application to ecological problems, is their ability to sample a wide range of biological and environmental data, as they can be equipped with various sets of sensors ranging from acoustic, chemical, and optical, depending on the needs of the survey (Lucieer and Forrest, 2016, Morris et al., 2014). Thus, they are commonly used for collecting acoustic data (Grasmueck et al., 2006, Peukert et al., 2018, Williams et al., 2012a, Wölfl et al., 2019) and are extremely useful for mapping as they provide a precise, enduring and stable platform for the instruments they carry (Marzinelli et al., 2015, Morris et al., 2016). They can also take direct chemical measurements with mass spectrometry (Thornton et al., 2015), hydrographic and oceanographic data (Pennington et al., 2016), and even detect hydrothermal vents through their plumes (German et al., 2008) or measure poly-metallic nodules on the abyssal plain (Gazis et al., 2018). The very high resolution of the environmental data they provide can also be used to directly inform ecological surveys (Morris et al., 2016, Huvenne et al., 2016a, Pierdomenico et al., 2015).

The application of this technology (including both class of AUVs) to quantitatively study benthic communities in particular is broad, from tropical and cold water coral reefs (Williams et al., 2012a, Williams et al., 2016, Williams et al., 2012b, Robert et al., 2014, Perkins et al., 2016, Perkins et al., 2019, James et al., 2017), benthic fish communities (Milligan et al., 2016, Tolimieri et al., 2008), deep water kelp forests (Marzinelli et al., 2015) or entire benthic communities (James et al., 2017, Simon-Lledó et al., 2018, Smale et al., 2012, Morris et al., 2014). They have also been instrumental in opportunistic discoveries like unknown cetaceans behaviours (Marsh et al., 2018). The adaptability of AUVs and their capacity to survey both environmental variables, as well as simultaneously collect images, can give a comprehensive view of benthic ecosystems in a single deployment. Besides, the stability of the vehicle also improves the quality of the images compared to other platforms (Foster et al., 2014, Jamieson et al., 2013). The main advantage of an AUV, however, lies in their capacity to quickly cover a large surface-area and collect tens, or even hundreds of thousands of images in one mission (Lucieer and Forrest, 2016); which ultimately leads to cost reductions and more data collected. These large datasets have also enabled complex survey designs to be implemented, granting better statistical robustness to the data (Perkins et al., 2016, Foster et al., 2017, Wölfl et al., 2019). Finally, they have been recognised as very suitable to reproduce surveys in environmental monitoring of the seabed, at an affordable cost, which could be of great importance to facilitate this practice for conservation (Bryson et al., 2013, Sherman and Smith, 2009, De'Ath, 2007, Smale et al., 2012, Williams et al., 2012a, Bayley and Mogg, 2019, Perkins et al., 2016).

1.2.2 Limitations of AUVs

From the literature, the advantages of AUVs for benthic ecology are obvious. AUVs are not a recent technology and they have been used for a long time (Sibuet et al., 1989, Singh et al., 2004). Since their early developments, they have been improved with new payloads and better systems, increased power, autonomy, more accurate navigation, additional sensors or lowered costs (Gafurov and Klochkov, 2015, Huvenne et al., 2018). Yet, although their use in science has increased (Figure 1-1), AUVs studies remain limited and it is still not clear to the benthic ecologists, how to use them, what to expect from them or what exactly could be gained from further investing in them.

Indeed, AUVs and particularly the cruising AUVs on which this thesis is focused, still suffer from a number of limitations that discourage some institutions from acquiring them or choosing to mobilize them over other vehicles, like ROVs and could explain why they are not more widely used in benthic ecology.

To start with, AUVs are not as versatile as ROVs and cannot be adapted to any mission like the tethered vehicles (McPhail et al., 2010).

Unlike (UAVs), AUVs are limited in their field of view by the physics of light movement in water. If images are the only way to identify organisms, this forces the vehicles to fly close to the seabed and they cannot photograph large surfaces at a time. Additionally, although AUVs, the cruising class in particular, can collect very high resolution multibeam data, including in the deepest waters, they cannot match the coverage of a ship-born multibeam due to their speed, power and autonomy limitations (Wölfl et al., 2019). Their poor manoeuvrability also makes them ill-suited to operations on spatially concentrated targets, such as vertical walls of canyons or hydrothermal vents, where fine movements and reactivity are needed (McPhail et al., 2010, Robert et al., 2017). This limits their usefulness to survey some of the most charismatic deep-sea habitats like cold-water coral reefs.

The impossibility for AUVs to make a decision has a direct impact on their capacity to collect physical samples. It is often difficult for scientists using ROVs to quickly decide what animal should be taken during a dive, because the target has to be identified with certainty. The ROV pilots need be able to collect the sample without compromising the safety of the vehicle and causing damage to the surrounding of the target. Automating this process is a challenge that, to my knowledge, no one has realistic expectations to tackle in the near future, for benthic sampling. Physical samples are however necessary to identify animals to the species level as image samples are often unable to provide such taxonomic resolution (Buhl-Mortensen et al., 2015, Williams et al., 2015, Bullimore et al., 2013).

In a more practical consideration, the acquisition of an AUV, and the capacity (skills and facility) to operate it, is expensive and may simply be too great for small research units. AUVs are technologically advanced systems that need to be maintained and require advanced knowledge in engineering; knowledge and experience that is difficult for a biologist to acquire in addition to the rest of their expertise. This is dissuasive for research teams that have already invested in acquiring other sophisticated equipment and complementary skills (Przeslawski et al., 2018). AUVs also do not solve the data analysis bottleneck that exists with other gear, but might in fact, worsen it (Perkins et al., 2016, Schoening et al., 2017). The additional data collected by AUVs is not met with additional data processing capacity, which negates their main advantage for ecological studies. Without the full benefits offered by large size datasets, there is little value for research teams to shift their sampling strategies away from traditional methods, particularly from the much more versatile ROVs. When asking researchers about their perceptions of the advantages of AUVs Przeslawski et al. (2018) reported that although they were aware of their advantages, researchers were concerned by their comparatively reduced reliability (as missions are frequently failed), their unsuitability for high energy environments (high tendency to drift in strong currents), and the relatively small number of vehicles available to researchers.

Several of the problems mentioned may be solved with better technology and to that end, research efforts exist (Gafurov and Klochkov, 2015, Huvenne et al., 2018, Quintana et al., 2018). Besides, some are already addressed by AUVs of specific designs, more adapted to tasks requiring better positioning, or agility (Przeslawski et al., 2018). It is, nonetheless, clear that the use of these new tools doesn’t come without new challenges for ecologists and important work remains to be done on what exactly sets AUVs apart in the study of biological communities. Their efficiency at collecting large datasets has been widely demonstrated and acknowledged and yet they are still treated as an interesting novelty, as illustrated by the large number of papers citing the use of AUVs in their titles as one of the points of note of their study. Will AUVs be consigned to only act as support for the science undertaken with other vehicles or replace them as a cheaper alternative? Or do they truly have the capacity to reveal patterns that were invisible to traditional ways of surveying the seabed? Can biologists fully exploit their advantages despite their limitations?

This issue is not only related to the AUVs themselves, but also to the processing of the data they collect; the two are inherently linked. This naturally leads to the challenge of developing more efficient image analysis methods.

1.3 Image data analysis and Artificial Intelligence

Many of the challenges faced by deep-sea ecologists are linked to the lack of appropriate data in both nature and quantity. While the new generation of AUVs will compensate for the lack of manpower needed to explore half of the planet’s surface the challenge in their use is not just in collecting the raw data (Schoening et al., 2017, Camps-Valls et al., 2017). It lies also in the interpretation of that data and the capacity of humans to process it.

In modern science, the advance of computing methods for data analysis, and subsequent freedom from what is possible by the human mind alone, has allowed for a significant leap in productivity in many areas of science (LeCun et al., 2015). However, the collection and processing (observation and translation) of ecological data is a complex task that it is difficult to hand over to a machine. In the particular case of the benthic ecosystem studied here, the information needed is what animals or groups of animals can be seen on an image. This requires a human mind to interpret the images and transfer that information to semantic form (Gomes-Pereira et al., 2016). In this context, the collection of more images, is only shifting the bottleneck along the processing pipeline. More images of the seabed (of sometimes very high resolution) still simply requires more manual analysis in order to serve ecological and conservation research. Thus, the only way to remove that bottleneck is to pass the image analysis step to any method faster than manual identification and counting of animals on images.

1.3.1 Automated image analysis

The development of image annotation software facilitating the treatment, management and analysis of the images and data produced and subsequently, the sharing of the results, has increased the speed and efficiency of the task in recent years. But there is only so much humans can do, as analysing several thousand images can take months, even to a trained specialist. The resultant data also suffers from many biases and inconsistencies over time (from start to end of the study to changes in focus related to time of the day) as listed in Durden et al. (2016a). Furthermore, manual analysis is poorly reproducible and studies have shown that even annotations produced by experienced personnel in parallel will present high variation in diversity and abundance (Durden et al., 2016a, Beijbom et al., 2015) and even suffers from psychological bias which, for example can lead to observers missing or failing to record large-sized targets (Eckstein et al., 2017). This limits the potential for combining interpretations from several researchers or teams and thus limits the size of datasets that can be studied in a single project to what a small number of people can annotate at the same time.

For all these reasons, the human element within a data processing workflow is regularly identified as the weak link and is effectively limiting the output of scientific research. The natural answer to this problem, following the same reasoning behind the increasing use of Autonomous vehicles and remote sensing (Brown et al., 2011) for environmental sampling, is to automate this process and let computers perform the simple but repetitive and labour intensive tasks, freeing researchers to pursue more novel goals (Schoening et al., 2017).

The need for faster processing of samples to tackle the increase in the amount of data collected is an old problem (MacLeod et al., 2010) and attempts to implement automated species identifications have been made since the early 1980s (Jeffries et al., 1984, Gaston and O'Neill, 2004). Interestingly, Computer Vision (CV, the process of image analysis by computers through application of artificial intelligence) applied to ecology remains a niche research field and has yet to become a common tool despite this long maturation. The trend fits within a wider move towards an increased use of data and computer science in ecology (Hampton et al., 2013, Weinstein, 2018). The applications of CV are many, from recreational and societal applications (https://github.com/AdamMc331/Not-KotDog), agriculture (Lu and He, 2017) and environmental surveys and ecology (Diesing et al., 2016). Identifying various taxonomic groups on images (Gómez et al., 2016, Barré et al., 2017) or by sound (Qian et al., 2017) is one of the most promising applications.

In the marine environment specifically, automated analysis of plankton samples has been pursued for a long time (Benfield et al., 2007, Culverhouse et al., 1996, Rolke and Lenz, 1984), extensively developed and implemented (Schmid et al., 2016). CV has also been employed to quantify environmental parameters like substrate complexity (Lacharité and Metaxas, 2017, Lacharité et al., 2015). Benthic ecologists have also attempted to automate species identification and counting from both fixed platforms (Aguzzi et al., 2009) and mobile vehicles (Beijbom et al., 2012, Edgington et al., 2006, Gobi, 2010, Marcos et al., 2005, Beijbom et al., 2015, Marburg and Bigham, 2016, Schoening et al., 2012). These studies usually reach a relatively high level of accuracy, above 80%, but are in general focused on a small number of species or habitats like shallow water corals. In addition, CV algorithms designed by a team or institution, although reused by the same group (Lacharité and Metaxas, 2017, Beijbom et al., 2015), are rarely exported for wider use, even after publication. Besides, the maintenance of published tools often stops after several years (Lobet, 2017).Thus, despite successes, the skill and material needed to train an algorithm (also referred to as a classifier in some publications) and their lack of flexibility and adaptability, particularly without direct contact with developers, has kept them out of reach for most ecologists. However, this could be about to change.

1.3.2 Computer Vision: Is it time?

This new momentum towards increased automation is following a wider movement, mainly driven by the increased computing capacity of market hardware and the appearance of open access tools. Recently, freely available pre-trained algorithms have allowed non-specialists the ability to train their own classifiers much more easily than before. The time when automated benthic community classifiers become a common tool, seems to have finally come (Williams et al., 2016, Weinstein, 2018).

In 2015, Google released a freely available software named TensorFlow (TF) that is able to build Neural Networks (NN). More importantly, it possessed many inner features such as a Python language Application Programming Interface (API) - whilst its main architecture is in C++, and a built-in function to manage memory allocations, which makes it relatively user friendly, yet fast compared to other frameworks. The release of the software has been enthusiastically welcomed by scientists across various research fields in which it could be applied and is promised to become a widely used tool (Rampasek and Goldenberg, 2016, Marburg and Bigham, 2016, Beijbom et al., 2012, Beijbom et al., 2015, Weinstein, 2018, Favret and Sieracki, 2016).

In a very basic manner, NN and their declinations function like the brain, in the sense that it is composed of many individual units, each able to perform a specific and simple operation (Rampasek and Goldenberg, 2016, LeCun et al., 2015). The combination of these different units and the way information flows through the network, allows for a more complex calculation when they are allowed to interact together. For a specific input, a specific pattern of activation within the network is formed and a new input resulting in the same or similar pattern can be classified as similar to that first input. During the training phase, the network “builds” itself without human assistance other than presenting raw data and thus, requires no parameterisation, as it picks-up automatically on important features for classification. This makes NN somewhat obscure (or “black-boxy”), but also fairly simple to implement. However, they require a tremendous amount of data, up to millions of images, and take a long time to build (Roig Marí, 2016).

The algorithms implemented within TF are already trained to analyse images. This gives TF two advantages: 1) it requires minimal tuning by the end-user and thus, no particular knowledge in machine learning and 2) it requires much fewer training images to be able to classify new images, than if the whole network had to be retrained. This is particularly suitable for ecologists who generally do not have the facility or the skills to train NNs or do not necessarily have abundant material for a given species.

In deep-sea biology, the use of AI and CV could open many fields of investigation, such as full integration with annotation software (Zurowietz et al., 2018), full visual coverage of areas or long continuous transects with images and better replications and consistency of results across projects. Ultimately it should, free some time for researchers to focus on tasks other than bulk analysis work. In the longer term, the potential applications of this technology are extensive and could involve real time animal detection (Seymour et al., 2017), morphometric measurements in situ (Shafait et al., 2017) or hierarchical classification (Bewley et al., 2015).

Therefore, with the increase in quantity of data and images gathered by AUVs, the output of annotation methods has to increase as well. Manual annotation lacks consistency and efficiency to match this challenge, but AI and CV appear to be an alternative and could in part help with this challenge. Despite many examples of successful application and many calls for the community to start using it routinely (Gaston and O'Neill, 2004, MacLeod et al., 2010), papers using AI to annotate benthic images are still rare. However, the potential of AI is undeniable and its use is getting easier and cheaper. It is now efficient enough, requires less specialist knowledge and experience than it used to and can realistically be applied to research projects. Yet, it needs to be tested on practical applications of a scale similar to most marine biological studies. Useful applications should be reproducible by any member of the community while ensuring results can be delivered with the same quality standards and in comparable time frames as manual methods.

Still to be answered is whether AI can work on the scale typical of many deep-sea studies and answer practical benthic ecological questions. What results would it give? And what are its real pros and cons compared to manual analysis?

1.4 Conclusion and aims of the thesis

Technology development could provide new answers to the above-mentioned challenges of deep-sea ecology and help fill the gaps in our knowledge allowing for more efficient management and protection of the deep-sea.

The potential of AUVs and AI combined is enormous. AUVs can collect the necessary quantity of data and CV based analysis methods can process that data. With automatically collected and analysed samples, thorough studies of epifaunal benthic diversity, species distribution and its environmental drivers at fine scales could be achieved.

However, at the dawn of a new age when biologists have to master skills from multiple domains, new challenges are rising and AUVs and AI have limitations that have not yet been addressed and their routine use in research is limited. To be more than a theoretical promise, AI and AUV based methodology has to be implemented in field studies with objectives of not only developing new tools but exploring the results and including them within the much wider frame of ecological research. The capacity of AUVs and AI to unlock new practical research opportunities has to be demonstrated with practical field studies. It is not certain when this technology will be mature and reliable enough to replace traditional methods - if they ever will - but their integration has to start and the community still has to further familiarize itself with the tools.

This thesis will attempt to apply AUVs and AI technologies to case-studies of deep-sea ecology and investigate how these method could contribute to the investigation of benthic biodiversity, fine scale species distribution and the environmental drivers of that distribution.

- **Chapter 2** describes the method used to extract information from the raw AUV data, particularly images, and how it is formatted for the following chapters.
- **Chapter 3** evaluates the impact of sample size on measures of diversity and how AUVs can influence the study of these parameters.
- **Chapter 4** uses a large dataset of abundance of benthic species to study its fine scale distribution and identify its drivers with species distribution modelling.
- **Chapter 5** implements automated identification of benthic organisms with Tensorflow and evaluates its feasibility in practical ecological studies.

- **Chapter 6** summarises, synthetize and concludes this thesis. It discusses the findings of the preceding chapters focusing on the potential of AUVs and AI, and the future of these technologies in deep-sea benthic ecology.

This work aims to provide guidance to other benthic ecologists considering using AUVs in their research and give helpful elements of advice to guide their strategic choices for future projects in order to help the community’s capacity to sample and understand this poorly known environment.

Chapter 2: Method - Collection and processing of AUV data

2.1 Abstract

This chapter details the origin of the samples used in this thesis. It describes the fieldwork during which the data was gathered, the vehicle by which it was collected and the way the images and environmental data were prepared, formatted and analysed to produce the data on which the following chapters are based. It is intended as a practical guide for researchers aiming to reproduce or improve the protocol. It thus contains practical details on the analysis of images and provides the R scripts used to process and format the data to facilitate a wider use of the method. It shows how complex AUV data can be and how much multidisciplinary experience is paramount to take advantages of the large amount of data collected by the AUV.

2.2 Introduction

AUVs have a capacity to bring multiple sensors to the seabed simultaneously, and comprehensively record multiple streams of data from a single location (Brandt et al., 2016, MacPherson et al., 2014, Williams et al., 2016). As such, one AUV mission will usually yield a large quantity of data of various natures and diverse formats (Wynn et al., 2012, Morris et al., 2014) also depending on the type of AUV used (Huvenne et al., 2018, Monk et al., 2018). These can then be combined for an integrated analysis.

All the data used in this thesis were obtained from a single cruise, JC136 (as part of the DeepLinks project), during which a number of AUV dives were conducted. There are three main components within the data, referred to as: 1) image data, 2) geomorphological data (from Multibeam (MB) echo-sounder) and 3) oceanographic and hydrographic data (from CTD and ADCP). The most important component for ecological studies is undoubtedly the image data, on which most of the processing time was spent. The data is therefore organised around the images collected, which act as sampling points or basic sampling units. Each image is associated with corresponding values of oceanographic, hydrographic and geomorphological data, as well as a detailed species composition (list of present organisms and their abundance), following image analysis.

This chapter describes how the data used in this thesis were collected, processed, combined and formatted for use in the following chapters. The objective was to develop a protocol usable by ecologists, where ubiquity and reusability are the primary concerns. A further priority was to enable the comparison of AUV data, to that collected by other vehicles and research teams, as much as possible. The following chapter is aimed at designing the best way to combine different types of data into a functional workflow; from raw data post-fieldwork, to a comprehensive map summarising ecological information, then later to reapply this same workflow to other stations. This provided an opportunity to assess the usability of the cruising AUV Autosub 6000 in a practical situation and to help identify the challenges of applying these vehicles to real-life ecological studies in the deep sea.

2.3 Method

2.3.1 Field work and data collection

The DeepLinks Cruise (JC136) was the fieldwork component of a NERC funded standard grant, aimed at studying connectivity of bathyal benthic populations within the United Kingdom's Exclusive Economic Zone (EEZ) (https://www.bodc.ac.uk/resources/inventories/cruise_inventory/report/16050/) and more specifically, the Rockall Trough. This large basin is located West of Scotland and Ireland, bordered in the North by the Wyville-Thomson Ridge at the end the Faroe Shetland channel, the Porcupine Abyssal Plain in the South, the Irish Margin on the East, and the Rockall Bank and the Rockall-Hatton plateau in the West. The study sites were spread across five seamounts (see Figure 2-1) within the Rockall Trough and an advanced protocol involving sampling across depth gradients, nested within seamounts, was carried out.

Benthic sampling of selected species was undertaken using the Remotely Operated Vehicle (ROV) Isis. In addition, replicated visual benthic surveys, video/still images, were conducted to study community composition and species richness. The initial intention was for the visual survey work to be undertaken by the Autonomous Underwater Vehicle Autosub6000, referred to as Autosub hereafter. However, due to various complications, the Isis ROV was ultimately used for that task. Nevertheless, a number of missions were undertaken by Autosub and the resulting data is used in this thesis.

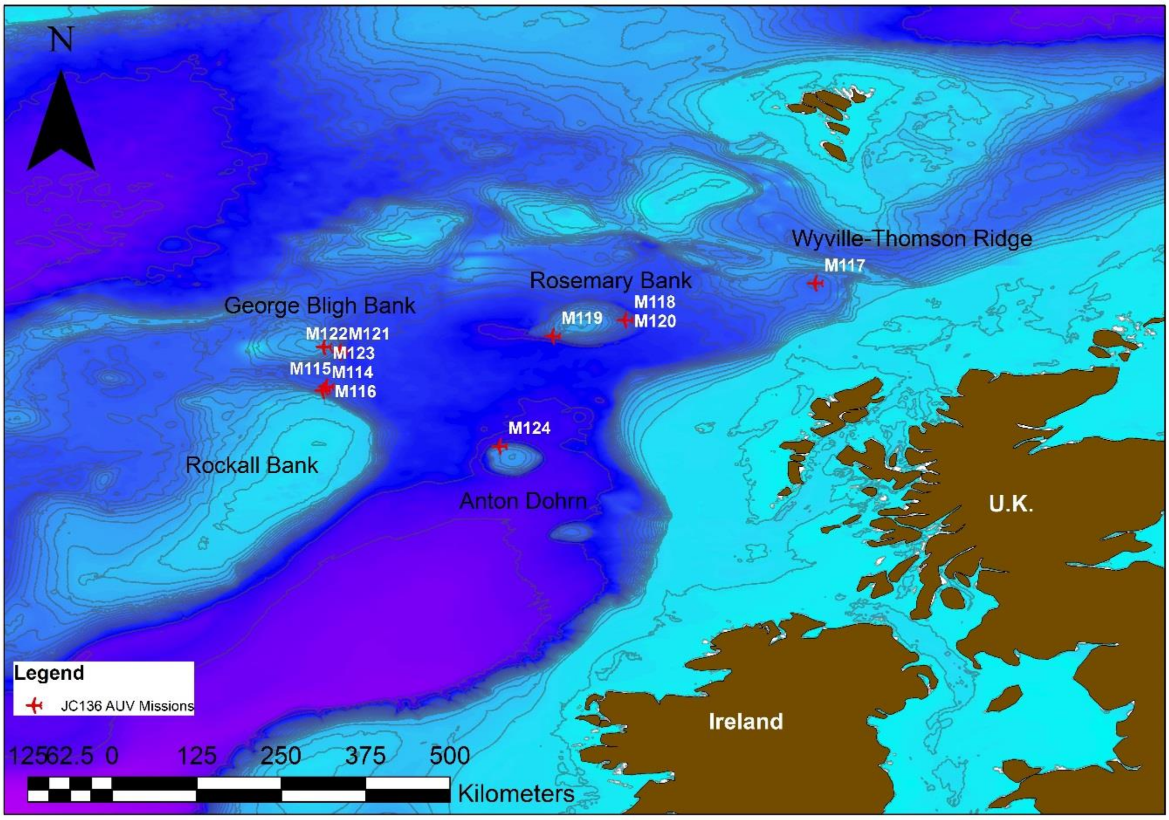


Figure 2-1: Map showing all the stations visited by Autosub6000 during JC136

This section will detail how the raw data, as handed over by the Autosub support team, was processed and formatted to be used in subsequent analysis.

2.3.1.1 Autosub6000

Autosub is a 6000 meter rated autonomous vehicle designed by the NOC Southampton and launched in 2007 (McPhail, 2009). It is a “cruising” AUV, designed to efficiently cover larger amount of ground (Monk et al., 2018). It can travel more than 100 km over a period of 24 hours at a speed of 1.7 m/s and It can take a large payload of various instruments on board (see Table 2-1 for list of equipment and Figure 2-2b for location of each instrument on the vehicle). The initial vehicle design was intended for oceanographic and geomorphological sampling and the photographic mission capacity was added later. Thus, it is better adapted to move within the water column, away from obstacles, than near the seabed. Unlike for the AUVs of the “hovering” class, designed for manoeuvrability and precision sampling (Monk et al., 2018), the high speed at which it is travelling, necessitates that it stays clear of rough terrain where obstacles can block its path or where a collision could happen and cause damage. To avoid obstacles, Autosub has a sonar-based, forward scanning, object detection system. Should something be in its path, it avoids collision by rising in the water column and flying over the object. As its manoeuvrability at full speed is low, this needs to be initiated in advance, sometimes, hundreds of meters before the obstacle (Wynn et al., 2012). For this reason, the sites where Autosub was successfully deployed, on photographic missions during JC136, were those where the seabed was flat and devoid of obstacles, so that the AUV could remain close to the seabed.

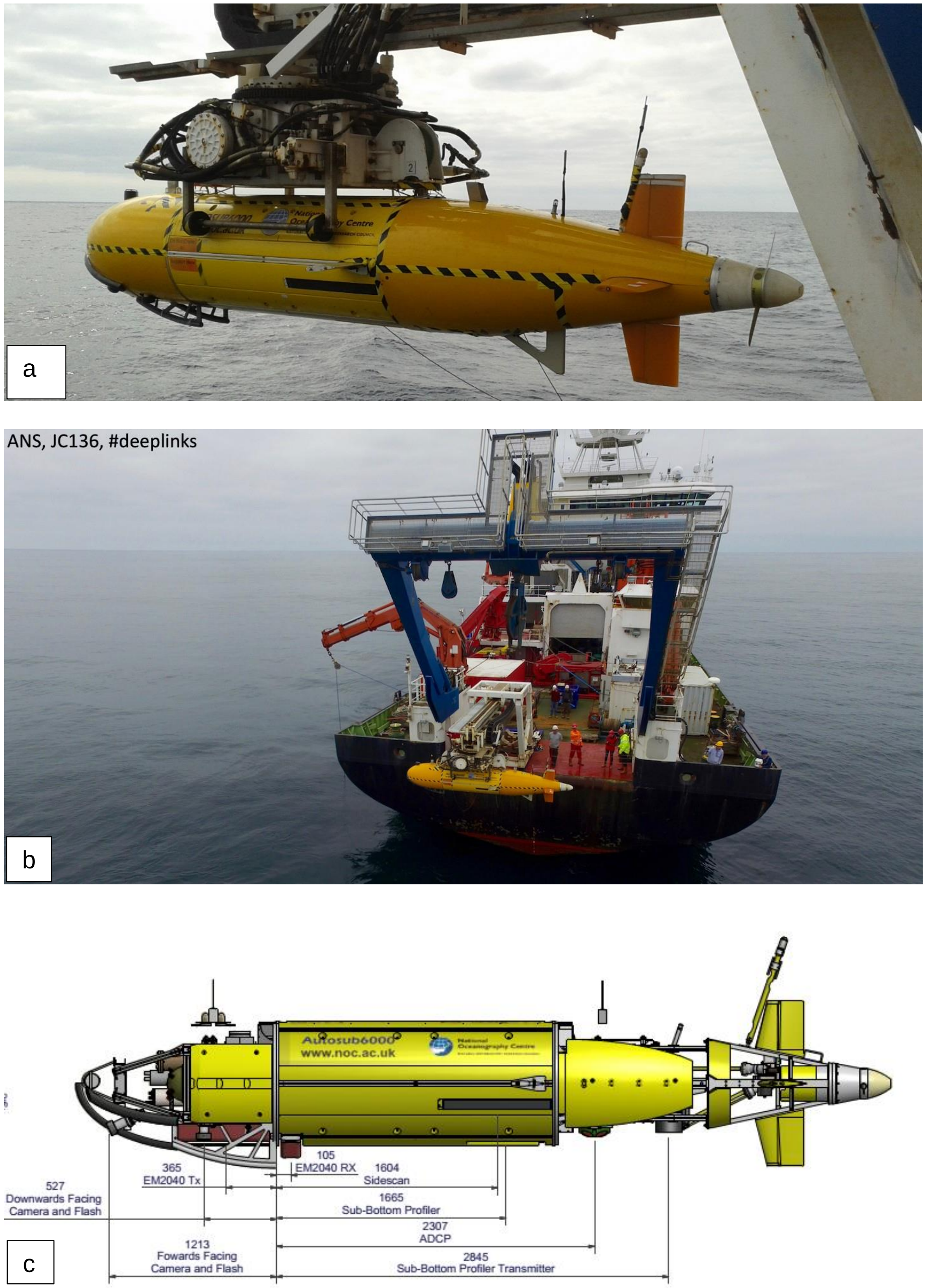


Figure 2-2: The AUV Autosub6000 (from JC136 Cruise report (Howell et al., 2016b). a: Autosub before launch, on the gantry during JC136. b: rear-view of RV James Cook during launch of Autosub. c: Technical sketch of Autosub (Reproduced from (Howell et al., 2016b)).

The Autosub configuration used during JC136 is displayed in Table 2-1.

Table 2-1: Technical specifications of instruments used by the AUV Autosub6000 during the JC136 cruise.

Instrument	Specifications
ADCP	RDI workhorse, 300kHz downwards, set to measure up to 30 4m water bins
Multi-beam	Kongsberg EM2040
Sub-bottom profiler	EdgeTech 2200-M 120-425kHz side scan and 2-16kHz
Cameras	2 x Point Grey Grasshopper, 2 + Flash (1 x downwards, 1 x forward)
CTD	Seabird 911 with 2 x SBE3plus, 2 x SBE4C, 1 x SBE43, Seapoint turbidity sensor, EH sensor, Fluorimeter.

On a typical deployment, the mission parameters are set as a number of waypoints (in 3-dimensional coordinates: latitude, longitude and depth) and the vehicle can be instructed to switch off an instrument during the mission (McPhail, 2009). Autosub is then launched from the gantry (Figure 2-2a and 2-2b), at the rear of the ship. Its position is tracked by the crew until it dives. Once it is underwater, its position cannot be known from the surface. Autosub's navigation is an Inertial Navigation System (IXSEA PHINS), or dead reckoning, and an on board ADCP to measure its speed and distance to the seabed (McPhail, 2009). This system tends to introduce error, or drift, in positioning which grows over time during the mission. This effect is quantified to around 1m per km travelled (Huvenne et al., 2009), but varies from one mission to another.

On photographic mission deployments, in order to take images of the seabed, the AUV needs to fly between 3 and 3.5 meters of altitude and keep a constant speed. While it takes images, it is also recording CTD data, primarily temperature and salinity, and acoustic (Multibeam and sidescan) data directed downward. On board the ship, the AUV is supported by 2 engineers and a technician, who process the data from the different instruments and upload it on hard drives and servers, where it can be accessed by research scientists. Due to the size of photographic data, up 750 GB in one mission, images are provided on a 2TB hard-drive after each mission. The other types of data are provided in a standard template folder, where the data from each on-board instrument is in text files or Microsoft Excel spreadsheets. Most instruments give a reading every 1 or 2 seconds and most of the data can be formatted into a table of point data where each entry has a location, in latitude and longitude, and associated data measured by all sensors.

2.3.1.2 Study site and sampling

The megafauna within the Rockall trough is part of a unique biogeographical province named the Atlantic Deep-Sea Proving and stretching across the ocean to the American Margin and south to the Equator (Davies et al., 2006, Watling et al., 2013). The Wyville-Thomson ridge in the North also act a biogeographical barrier beyond which the fauna belongs to the Arctic province (Watling et al., 2013). At broad scale, Its distribution structured by depth and substrate types (Gage, 1986, Howell, 2010). It is dominated by echinoderms cnidarians, poriferas and Arthropods (Gage and Tyler, 1991, Davies et al., 2006). A number of vulnerable Marine Ecosystems (VME) are commonly found in this area like cold water coral reefs, and coral gardens on hard substrates (Roberts et al., 2006, Howell et al., 2010b) and aggregation of the Hexactinellid sponge *Pheronema carpenteri* (Howell et al., 2016a, Rice et al., 1990) and the xenophyophore *Syringammina fragilissima* (Bett, 2001b, Ross and Howell, 2013) on soft sediments. The existence of these VMEs and the need to better understand their ecological function, connectivity patterns and better predict their extent and distribution is a strong driver of research in this area (Chaniotis et al., 2020, Howell et al., 2010b).

At the start of the JC136 cruise, the intention was to deploy the AUV for a seabed survey on at least three, ideally five, of the seamounts, at 1200 m. The 1200 meters depth band was chosen due to the technical limitations of Autosub, constraining photographic missions to areas of flat seabed which are common at this depth and in this region. However, the limited availability of Autosub in the first days of the cruise, technical failures later, as well as capricious underwater conditions, made comparison between sites impossible. Of the 4 attempted photographic missions, only one (M116, on station 26) was successful and yielded around 120,000 images along with geomorphologic and oceanographic data. Additional acoustic and oceanographic data from another mission on the same site (M115) was also used. Therefore, station 26 became the sole focus of this thesis. Examples of images from successful and unsuccessful photographic missions are shown in Figure 2-3.

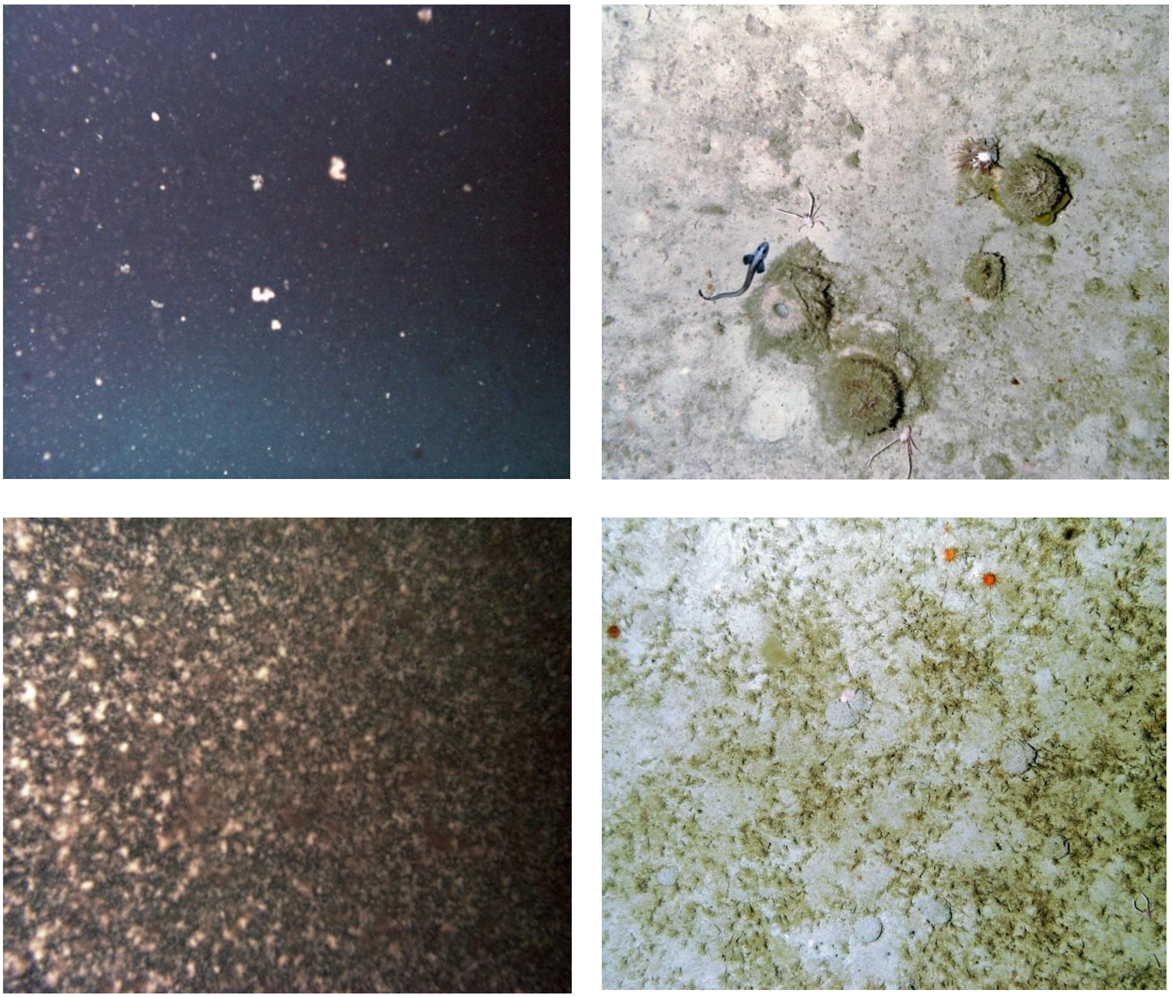


Figure 2-3: Example images illustrating the various visibility conditions encountered by Autosub during its photographic missions.

2.3.2 Images processing and annotation

2.3.2.1 Image and image metadata processing, georeferencing and sorting

The M116 dive occurred near the North Eastern flank of Rockall Bank (Figure 2-4). During a previous mission, Autosub mapped the area and identified potentially challenging terrain in the southeast corner of the survey area, which was avoided in the subsequent mission. The mission plan was a tight mesh pattern, which provided

27 straight transects. This pattern was chosen to approximate full coverage and provide points and transects at various distances from each other. 11 were shortened to avoid the rugged terrain in the southwest and are circa 1.8 km long. The other transects are circa 3.5 km long.

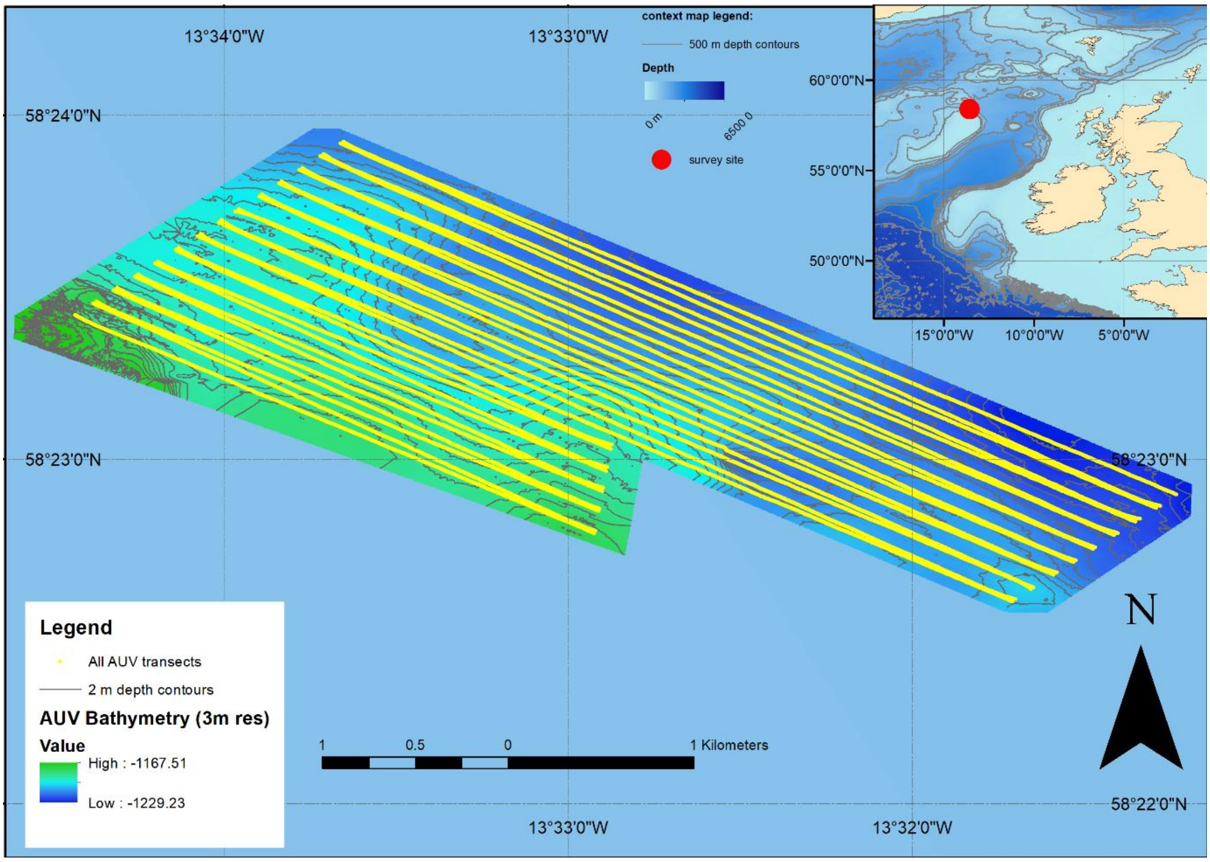


Figure 2-4: Route followed by Autosub while collecting the images during dive M116.

During the dive, the images were collected with 2 Grasshopper 2 - GS2-GE-50S5C cameras (Figure 2-5); one vertical, downward facing and one forward facing, at an angle of 30° from horizontal. The frequency of capture was set to 1 image per second. Image resolution is 2448x2048 pixels.

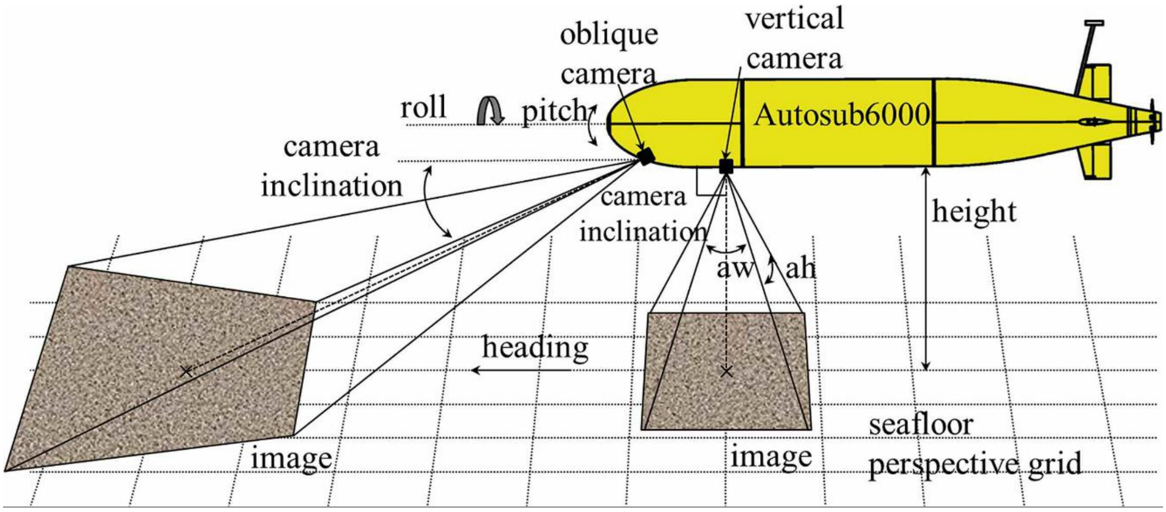


Figure 2-5: Camera field of view on Autosub. Reproduced from Morris et al. (2014).

Due to uneven lighting and the difficulties linked to the uneven size of the sampled surface, images from the forward facing camera were not considered in this study. Of the downward facing images, around 10,000 images taken during the descent and ascent phases, while the AUV was high above the seabed, were also discarded. A further 5000 were taken while the AUV was turning and repositioning; the slowing pace of the vehicle during this phase tended to cause overlapping images and were thus excluded from the transects. They are usable but have to be precisely accounted for to avoid biases in quantification of the organisms present in them. The remaining 55,000 images across the 27 transects constituted 1,600 images in the 9 short transects and 2,600 in the remaining 18. In these transects, image overlap could appear when the submarine reached a certain altitude, which was manually estimated between 3.5 and 3.6m.

Images were provided by the Autosub team in ".raw" format and were then converted to JPEG format with the IrfanView software (Skiljan, 2012). Colours were autocorrected in the process with the default method implemented in the software. An

80% compression was applied to the images for subsequent manual analysis to reduce their file size (in bytes). Comparison with un-compressed images, showed negligible loss of quality. Prior to spatial analysis, each image was geolocated using AUV navigation data based on a common time stamp.

The surface of each image in square meters (m^2) was calculated with the method described in Morris et al. (2014). This work used the characteristics of the camera mounted on Autosub, namely, the vertical and horizontal acceptance angles (*aw* and *ah* in Figure 2-5), the focal length of the camera and the altitude of the vehicle to calculate the size of the rectangle below the AUV. This corresponds to the size of the image if it is perfectly horizontal or at a right angle from the seabed. However due to the pitch and roll of the vehicle that occurs, which is constantly recorded, this has to be corrected. This was performed in a custom Python (https://www.python.org) script, adapted from Morris et al. (2014) original Matlab (https://www.mathworks.com) code.

2.3.2.2 *Manual annotation*

Annotation, the process of extracting biological information from images or translating objects or events in an image to the semantic level (Gomes-Pereira et al., 2016), is the most important step of data analysis in this project. Images were annotated using the open access software Biigle 2.0 (Langenkämper et al., 2017), accessible on https://biigle.de/. Biigle 2.0 is a web browser based software that facilitates the visualisation, analysis and sharing of data from any computer, with a free account and log-in. Images were uploaded on an *Amazon Web Service* (https://aws.amazon.com) server and then remotely accessed via Biigle. Before uploading, JPEG image quality was set to 80% to reduce their file size to around 25% of the original after compression.

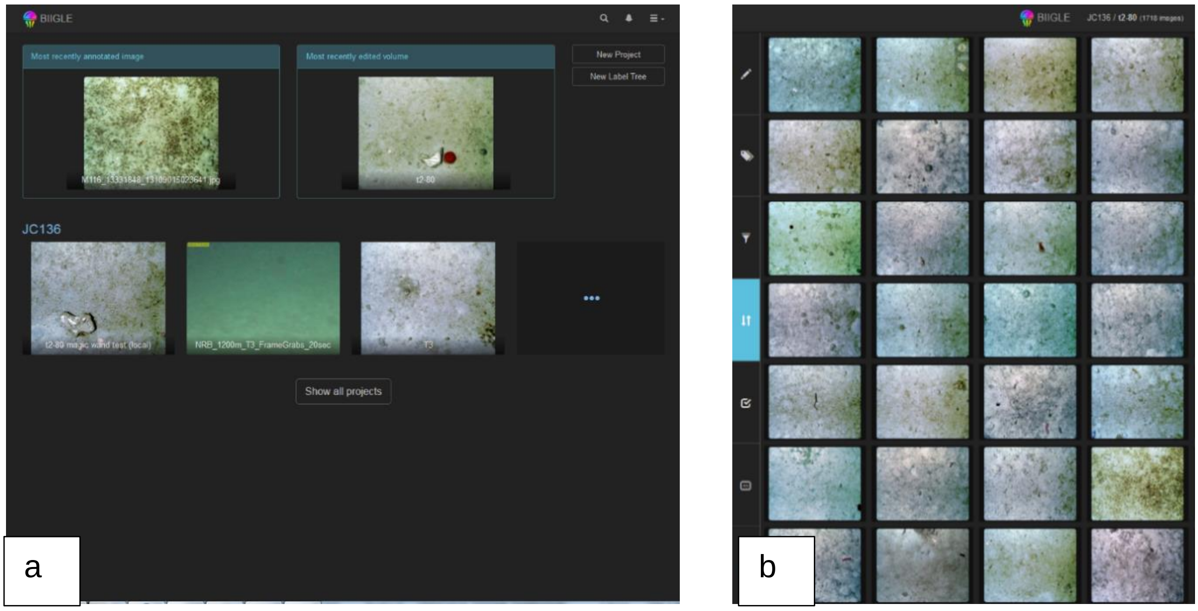


Figure 2-6: a: Biigle main projects and volumes management screen. b: Volume overlook screen showing each pictures miniature.

Organisms present in images were identified as Operational Taxonomical Units (OTUs), following a published OTU catalogue for this region (Howell and Davies, 2016). This catalogue was imported into Biigle 2.0 (Figure 2-6), appearing as a hierarchical 'taxonomic' tree, and subsequently modified to use up to date taxonomic Arborescence from World Register of Marine Species (http://www.marinespecies.org/); the original imported catalogue is available at https://deepseacru.org/2016/12/16/deep-sea-species-image-catalogue/). OTUs correspond to various taxonomic levels and are defined by what can be distinguished on pictures rather than actual taxonomic criteria. Thus, they do not necessarily correspond to a coherent taxa, as it is rarely possible to identify marine animals to the level of species. In general, OTUs for large animals from well-studied groups, have a higher resolution, down to genera and sometimes, even species. The anatomy of some groups also makes them easier to identify, like most chordates or the larger decapod crustaceans. Some groups that are smaller, less well studied or very diverse taxonomically, while homogenous in appearance, are sometimes classified according to very pragmatic parameters like shape and colour. This is the case for encrusting sponges. Some phyla, like bryozoans, are not divided into sublevels.

The images were annotated in a random order, to avoid observer bias also manifesting as spatial bias. The view on the image was zoomed to original resolution (see Figure 2-7b), and the “lawnmower” tool was enabled to restrict the number of animals visible on screen at a time and make the searching and detecting more systematic. The tool sequentially moves the zoomed in window from section to section, methodically going over the whole image (Figure 2-7c). This makes the analysis time longer, but tends to increase detected richness and abundance of epifauna.

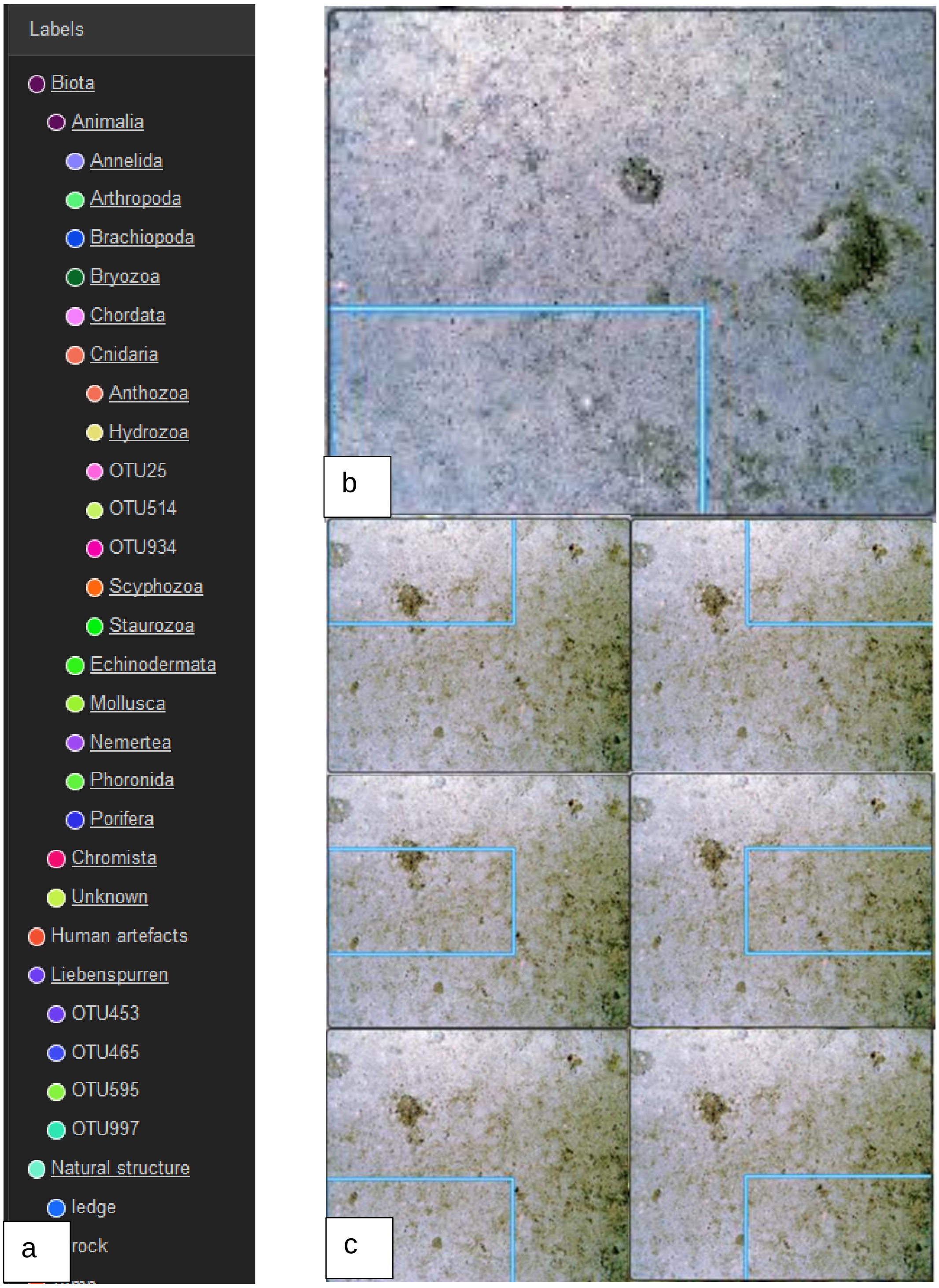


Figure 2-7: Illustration of Biigle's annotation screen. a: Illustration of the phyla and higher level cnidarian tree branches. All underlined labels are parent lables with a various number of sublabels. Each label, including parents can be used in anotation in. b: extent of the section visible when zoomed in. c: window slide with lawnmower tool. The size of the zoomed in window depends on the dimensions of the monitor used for analysis.

During the annotation process itself, illustrated in Figure 2-8, a point or a shape was superimposed over an object or animal. The shape used depends on the average size of the OTU. It is a dot for most small animals as it requires only one click and the size is not quantifiable on these OTUs regardless. It can also be an ellipse, a rectangle, a regular or irregular polygon, a line or a circle; the most commonly used shape in this study. This shape is then associated with an OTU label from the tree; this ensemble forms an annotation.

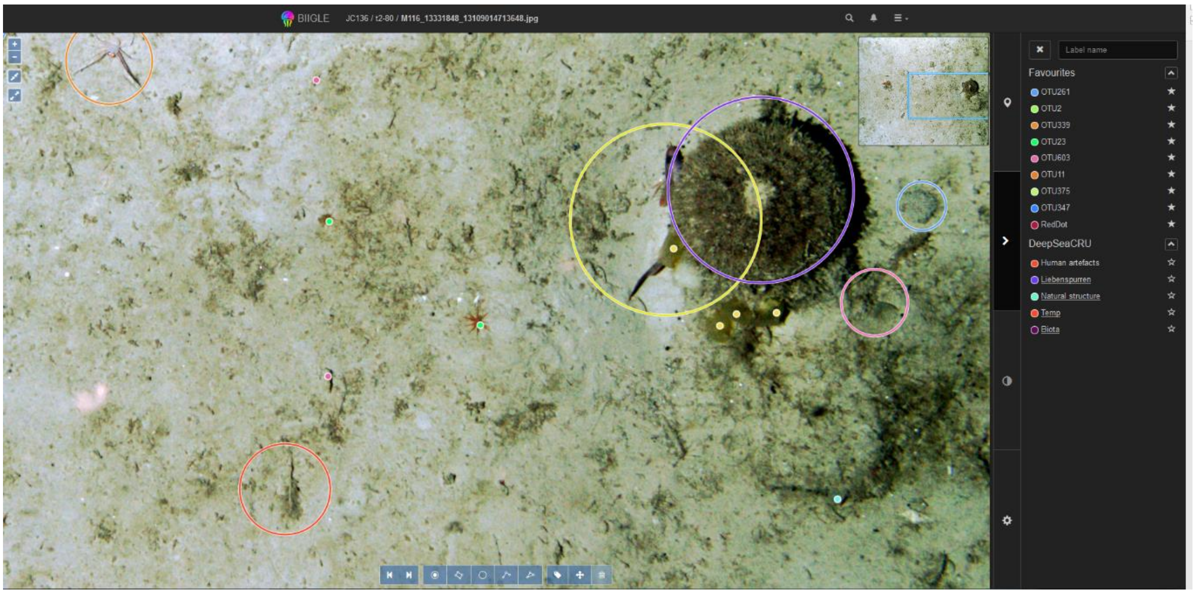


Figure 2-8: image annotation panel in Biigle 2.0. the image is displayed in the centre and the label tree on the right hand side tab.

2.3.2.3 Validation

Every annotation was visually validated. This step made use of the “Largo” (Label Review Grid Overview) evaluation tool included in *Biigle*. This allows each occurrence of a specific label, or OTU, to be cropped from the original image and displayed on the same screen. This enables a very quick visual check and if needed, the correction of errors, as illustrated in Figure 2-9. It can be done by the original annotator or another person with access to the volume. This step permitted the revision or removal of identification mistakes and correction of some inconsistencies. Consequently, some OTUs were also split into 2 or more, whilst others were merged. After a new OTU was created, relevant annotations could also be attached to this new label via the largo tool.

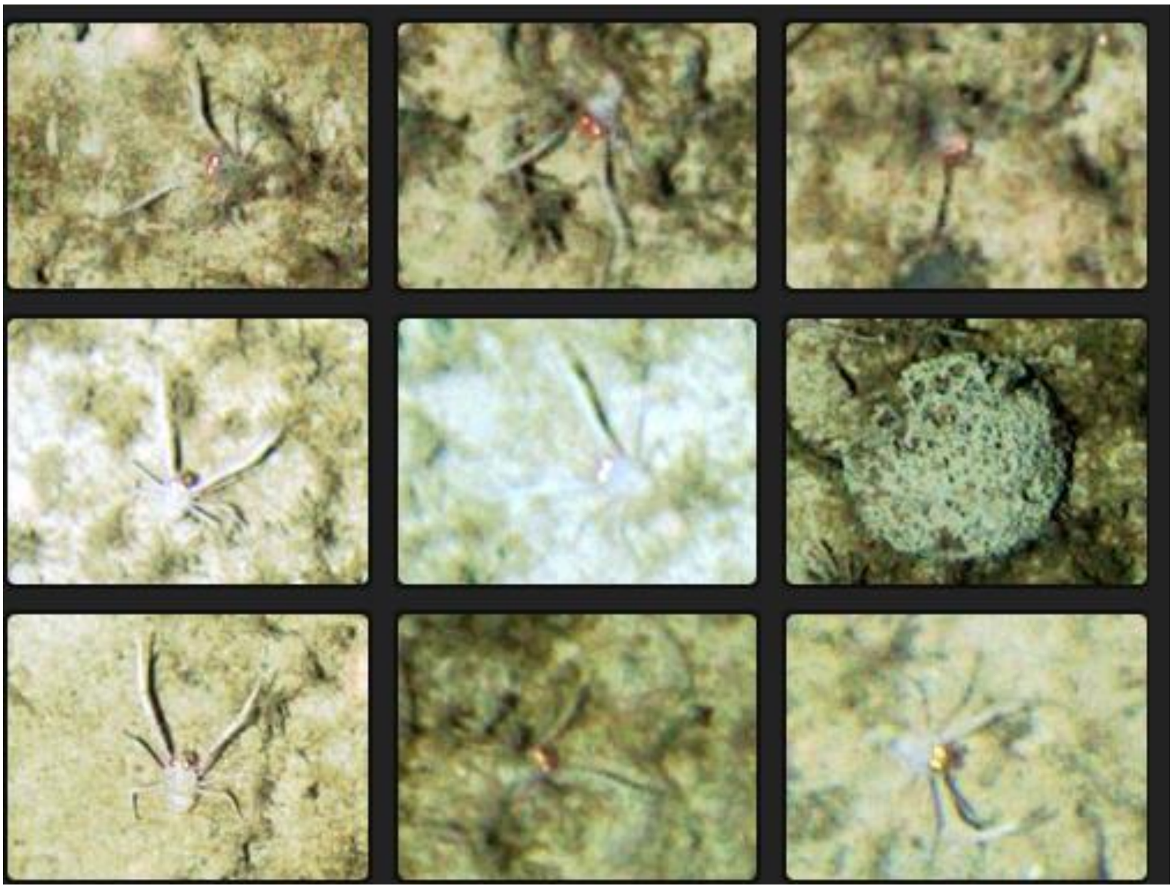


Figure 2-9: Illustration of the use of Largo too to find identification errors: All images of Munida sp. OTU339 - cropped around the animal are shown. Spotting identification errors (like the xenophyophore OTU261 on the right) is easier when all OTUs are displayed side by side.

In total 1,718 images were annotated with this method all from the t2 transect. This represents less than 3% of the images collected by Autosub at station 26. It nonetheless forms a dataset of more than 63,000 annotations of more than 110 OTUs. It took around 5 months to annotate the transect and validate the annotations. More time was later spent on the annotations to correct and revise some miss-identification.

The list of annotations per image was exported to R (R Development Core Team, 2011) for processing and use in the following chapters. A custom R script was used to create tables of information per OTU, information per image, species contingency and presence-absence matrix. The taxonomy imported from World Register of Marine Species (WORMS) (Costello et al., 2013) was also added to this table (also detailed in Appendix A2-1). The R code used to produce these tables has been made available (Appendix A2–2), so it can be used by other researchers to quickly export Biigle’s annotations.

2.3.2.4 Semi-automated counting of OTU261

Although manual annotation produced a large dataset, it represented a very small portion of the images available. Thus, a semi-automated annotation protocol was tested to evaluate the capacity of this method to speed up image annotation and produce additional data.

The number of OTU261, the xenophyophore *Syringammina fragilissima* (Brady, 1883), was measured with the Machine Learning Assisted Annotation (MAIA) expansion of Biigle 2.0. This tool-box implements an experimental procedure to automatically detect objects of interests in images (Zurowietz et al., 2018). The system is based on the Mask R-CNN convolutional neural network (He et al., 2017) and uses transfer learning to "teach" this openly available model to identify targets provided by the user. It needs a certain number of examples of this target species and can then “scan”, searching for similar patterns within unannotated images. The results come in the form of image patches that match the target and are referred to as annotation candidates. These must be manually checked and confirmed as the target or otherwise discarded if they are false positives. In addition, the exact location of these candidates can be modified during this manual verification phase to ensure the whole organism is included and not just the feature picked up by the algorithm.

MAIA was used to detect OTU261 in transects t3, t6 and t14. For each of these, which correspond to a volume in Biigle, some images were manually annotated to gather enough data to train the algorithm. In each volume, at least 200 examples of OTU261 were used to train the CNN, using default parameters and the number of clusters set to 1 or 2. The candidate annotations were all visually checked and those corresponding to *S. fragilissima* were converted to annotations that could subsequently be used for statistical analysis. This process was repeated a second time, with at least 1,000 example annotations detected in the first round. These could then be used as training examples to increase the detection rate and accuracy of the algorithm. Different Biigle sessions of annotations (a process in Biigle that attaches a time stamp to an annotation so they can be filtered by round of analysis) were used to ensure each individual was counted only once in the final dataset. *S. fragilissima* abundance was counted in more than 4800 images with this protocol.

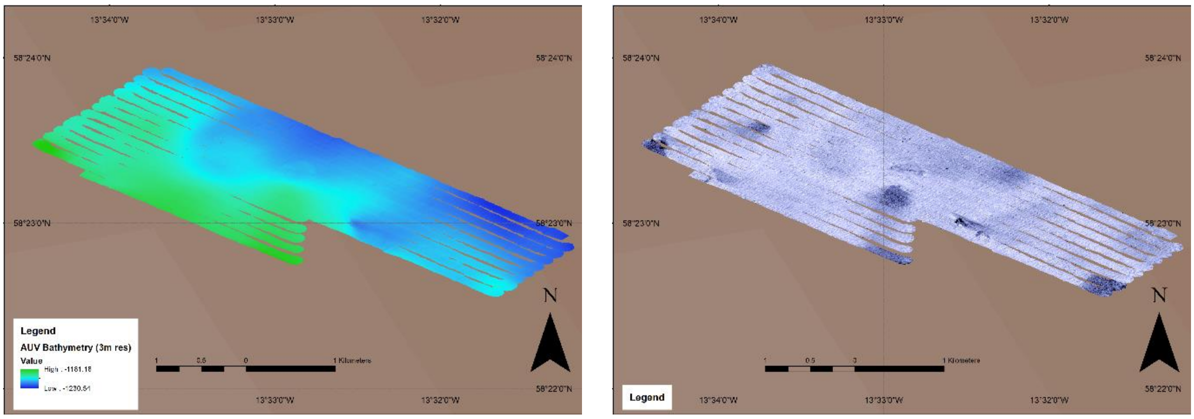
What transpired from this experience is that this object detection system works well, even with only several hundreds of examples of the target species used in training. The algorithm missed an estimated 10 to 15% of the *S. fragilissima* present on the images. The time taken to annotate the data with this method is difficult to measure as it was performed by multiple users (Myself and Jamie Cowle from University of Plymouth) and was partially experimental thus requiring a training stage and extra time for validating the annotation as well as verifying between-users consistency. The time taken to semi-automatically count the number of individuals of *S. fragilissima* in a transect, remains in the order or several days (4 to 7 depending on the size of the transect), if careful checks are performed on the results. This is faster than manual counting, but not by the margin needed to analyse the whole set of images collected by the AUV in a reasonable time, especially if aiming to detect and identify all of the organisms present.

2.3.3 Geomorphological and oceanographic data

Other data components, collected by the AUV, were also explored for later use in the study of the environmental parameters driving the distribution of the benthic organisms seen in the images.

2.3.3.1 Geomorphologic and multibeam data

Autosub is fitted with a Kongsberg EM2040 multibeam sonar system offering up to 400 beams. During dive M115 it was used at 200kHz to collect multibeam data, bathymetry and backscatter, which was then processed by members of British Geological Survey (Sam Faithful and Kirstin Crombie). The bathymetry was later re-processed by a University of Plymouth masters student Joe Augier to produce a cleaner 2.5 meters resolution surface.



a b

Figure 2-10: Acoustic data collected by Autosub during the JC163 cruise. a: Multibeam Bathymetry b: Backscatter

As the AUV travelled at low altitude, gaps were left in the multibeam cover of the terrain between the transects. Some of these gaps were up to 30 meters wide and their presence would limit the surface onto which habitat mapping could be implemented. However, given the low variability of depth in the area, it was considered acceptable to interpolate depth values from the available multibeam to produce a continuous surface. This was done in ARCGIS 10.4 (ESRI, 2014). The values were interpolated using the Inverse Distance Weighting (IDW) tool to produce a surface raster at 2.5 m resolution shown in Figure 2-11. The resulting surface does contain artefacts, which could not be removed without extensive work and could influence later ecological results. No such interpolation was applied to the backscatter as it is more variable. The unpredictable nature makes spatial interpolation inappropriate and may introduce error.

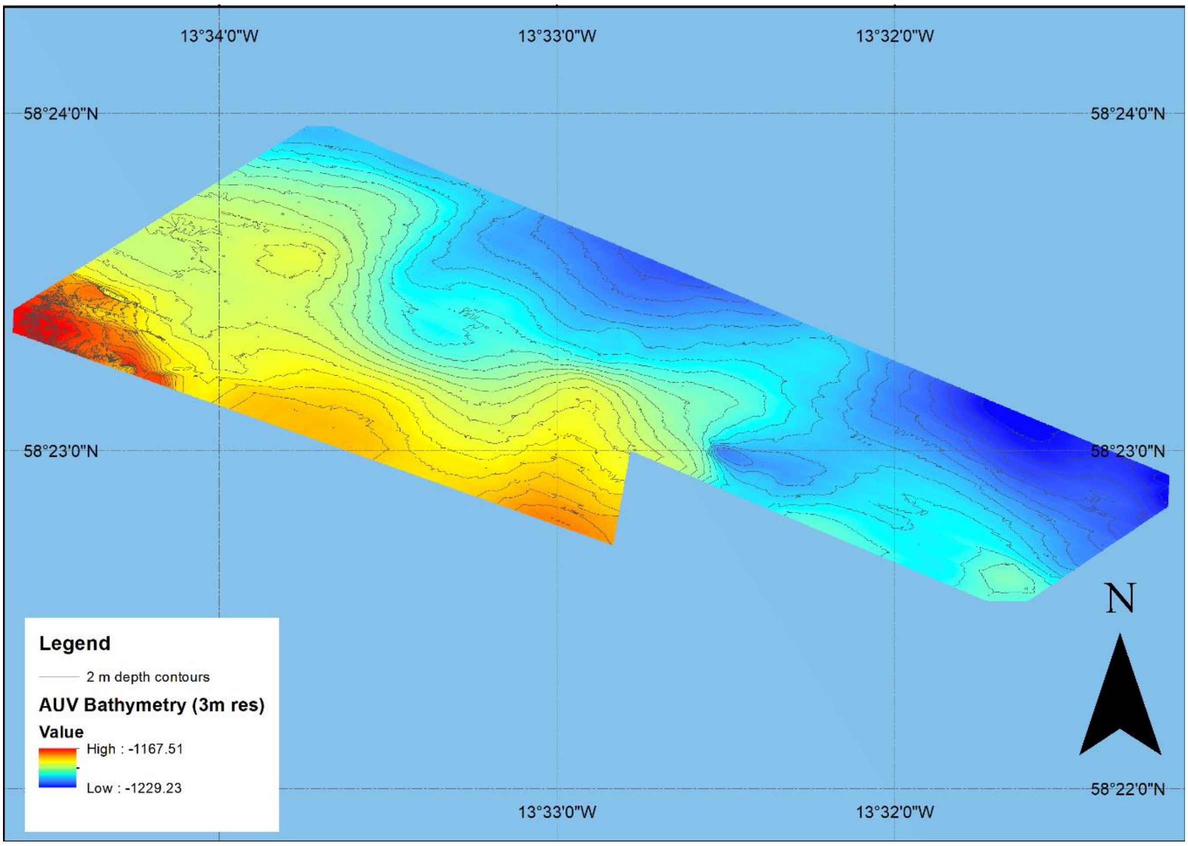


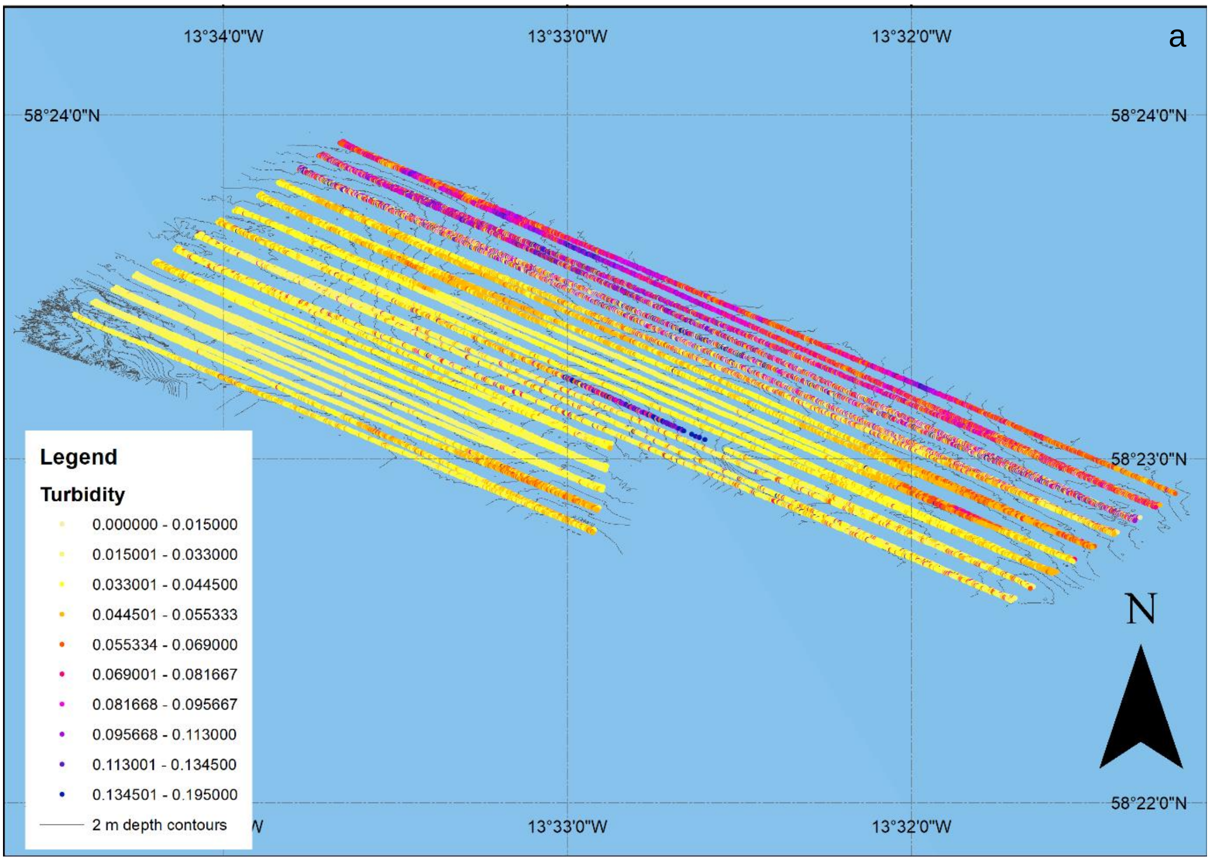
Figure 2-11: Interpolated bathymetry of the survey area at station 26 collected during dive M115

The multibeam data was later further processed to produce other topographic information layers used for species distribution modelling.

2.3.3.2 Oceanographic data

The CTD on Autosub is a Sea Bird Electronics 9 unit, fitted with 2x SBE 3 temperature sensors, 2 SBE4 conductivity sensors and 2 Pumps. A Seapoint turbidity sensor, an SBE 46 oxygen probe and WET Labs ECO-AFL/FL fluorimeter were connected to the CTD.

Variation of dissolved oxygen concentration, salinity, temperature and turbidity were explored by mapping their values along the track followed by the AUV during M116 (examples are shown in Figure 2-12).



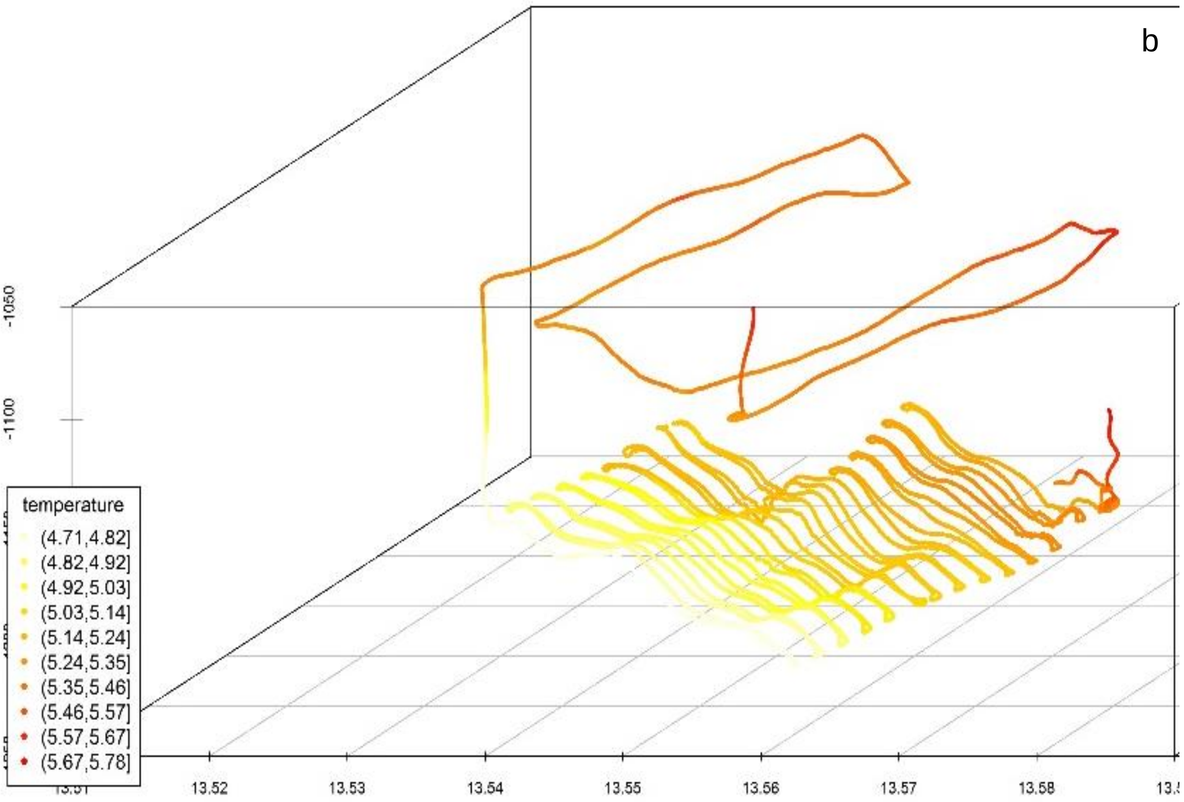


Figure 2-12: Example of oceanographic data plotted in 2D and 3D. a: 2D plot of the turbidity (in FTU) variations along the route of the AUV. b: 3D plot of the temperature (in °C) variations along the route of the AUV. Latitude and longitude are in degrees and depth in meters.

This exploratory approach of the data showed that variations of these oceanographic variables can be observed throughout the dive however it is unclear whether this is spatial or temporal variation. Thus any observed correlation with biological phenomena, such as the abundance of a given organism, was considered with great care in subsequent chapters using these layers.

Variables used in the following chapters, particularly turbidity, were converted from points to a continuous raster at the same resolution as the multibeam bathymetry, with a natural neighbour interpolation.

Owing to time constrains, no further exploration nor processing of the hydrographic data was undertaken.

2.4 Remarks on processing AUV data

This chapter details how the raw data was converted into a format, usable in spatial and statistical analysis. Moreover, it gives practical insights into the use of AUV of the cruising class data and how it involves work on various types of data, as well as careful management and curation of the images, tables and other files. The vehicle's capacity to incorporate data from multiple sensors helps consider a bigger picture in benthic ecology and builds up a multi-disciplinary (biological, geomorphological, oceanographic) approach to ecology. However, this abundant and diverse data is not accessible without relevant skills ('big data' analysis, acoustic data processing, oceanographic data processing, general programming and Geographic Information System), computing power, and relevant software to process the raw data that comes from the vehicle.

Overall, it is an important time investment, not just to acquire but to process the data and learn how to proceed with this task. Hence, some of the data, particularly hydrographic, was not used in subsequent thesis chapters due to the lack of time. Likewise, it took a long time to manually and semi-automatically annotate a very small fraction of the total number of images gathered at one station only. This highlights how much the use of autonomous vehicles increases the complexity of benthic ecological studies and how collecting more data does not easily translate into better understanding of deep-sea ecosystems.

The image dataset detailed in this chapter is used in **chapters 3**, 4 and 5. The environmental data is used in **chapter 4**. The work done in this chapter is far from fully exploiting the vehicle's capacity and it stresses the need for more efficient ways to annotate images which are investigated in **chapter 5**.

Chapter 3: The effect of sample size on deep benthic soft sediment biodiversity measure.

3.1 Abstract

Basic characteristics of deep-sea ecosystems are commonly investigated through simple metrics like diversity or density. Comparable measures of these metrics are important to build a comprehensive understanding of deep-sea ecosystems over multiple sampling sites and environmental gradients, or over repeated surveys in monitored areas so that it can inform the development of a coherent conservation strategy. Obtaining such comparable measures requires standardisation of the sampling protocol, to attribute observed differences to genuine ecological causes rather than biases caused by differences in sampling effort.

Obtaining samples in the deep-sea is now becoming less difficult with the increasing availability of AUVs that are able to sample hundreds of thousands of images or tens of square-kilometres over a single dive. Consequently, investigating the exact sample size needed to observe the entire alpha diversity of a station can become a practical question.

To this end, we used a large dataset collected by the AUV Autosub6000 to study the effect of sample size on the measures of density and diversity. The appropriate amount of sampling effort required to obtain a reliable measure of density was 150 m^2. Depending on the index used diversity could be accurately measured with 25000 to 75000 individuals (1200 m^2 of seabed photographed and a less than 850 meters long

continuous transect by Autosub6000 to 3730 m^2 of seabed or 2600 meters of transect) but could be estimated with replicated or pseudo-replicated samples of 40% that size while also quantifying the variability of these estimates. Finally, we discuss how AUVs, such as Autosub6000, can improve deep-sea benthic surveys.

3.2 Introduction

The deep ocean is the largest ecosystem on Earth but is, in general, poorly explored and understood With rising threats from anthropogenic activities, a better understanding of deep-sea ecology is needed to design efficient conversation strategies (Barbier et al., 2014, Danovaro et al., 2017a, Van Dover et al., 2014, McClain and Schlacher, 2015).

International initiatives, like the United Nations General Assembly Resolution 61/105″ and the Oslo-Paris (OSPAR) Convention for the Protection of the Marine Environment of the North East Atlantic have encouraged the creation of Marine Protect Areas (MPA) to preserve deep-sea ecosystems and biodiversity. Some nation-states, like the United Kingdom, have also integrated the objective of preserving marine diversity into their policies (Defra, 2005, Kroeger and Johnston, 2016, Chaniotis et al., 2020). Efficient conservation strategies must be based on reliable and representative information on the ecosystems it aims to protect, which is currently lacking (Rogers, 2015, Folkersen et al., 2018). This information is needed to place the MPAs in the correct location (Halpern et al., 2010) and ensure they are connected (Hilário et al., 2015) in order to fulfil their role. Another effect of the poor knowledge of the ecology and dynamics of deep-sea ecosystems is the relative lack of understanding of how MPAs help the recovery of protected ecosystems. Carefully designed monitoring strategies are needed to evaluate their efficiency (Bowden et al., 2016, Clark et al., 2016a, Huvenne et al., 2016b). A coherent and future-proof MPA monitoring strategy must rely on repeated and standardized surveys designed to detect changes and must therefore provide comparable measures of descriptive metrics such as diversity and density (Halpern, 2003, Foster et al., 2013, Woodall et al., 2018). This will enable scientists and biodiversity managers to ensure the variations in measures are not the result methodological biases but a genuine biological phenomenon (Kroeger and Johnston, 2016). In other words, baselines need to be established so that deviation from it can be accurately detected and quantified (Rogers, 2015).

Density and, more importantly, diversity are essential descriptors of the state of an ecosystem (Rogers, 2015, SCOR, 1994, Gotelli and Colwell, 2011)). Diversity or species richness is considered high in the deep-sea (Rex and Etter, 2010, Poore et al., 2015), but its exact extent remains unknown (Costello and Chaudhary, 2017). Much remains to be understood on how its different components are affected at various spatial scales, and how it interacts with its environment (Rex and Etter, 2010, McClain and Rex, 2015). Alpha diversity, the number of species present in a specific site, provides the most simple indicator of diversity. This basic quantity, once known, can also enable further comparison with other sites, across environmental gradients (beta diversity) and at larger or regional scales (Gamma diversity) and all comparisons depend on the quality of the initial alpha diversity measure. A better understanding of deep-sea diversity requires a global and coordinated approach but the lack of standardisation of scientific methods, makes comparison and aggregation of data towards a coherent strategy difficult (Levin et al., 2019, Woodall et al., 2018). Ensuring measures of density and diversity are representative and comparable are an important step to tackle this challenge.

Diversity measures have been shown to be strongly influenced by sample sizes (Sanders, 1968, Danovaro et al., 2017a, Jones et al., 2017, Noble-James et al., 2017). Methodological tools to determine appropriate sampling effort to correctly measure local diversity or, at least, accurately estimate it, have existed since Sanders (1968), in the form of species accumulation curves (Gotelli and Colwell, 2011). In practice, sample sizes are a trade-off between ecological necessity and logistical constraints. Samples that are too small can give erroneous measures of ecological parameters. On the other hand, Oversampling and post-hoc resampling or stratification can be a solution to ensure the right amount of data is available, but it can be a waste of precious resources (Noble-James et al., 2017). Given the low density of benthic megafauna and the high proportion of rare deep-sea species within this group (SCOR, 1994), the commonly used sampling gears are unable deliver a sample of the size needed at a reasonable cost (Danovaro et al., 2016, Brandt et al., 2016), but technological innovation could provide new solutions to this challenge in the near-future.

With the increase in access to camera technologies, imagery is now proving a popular way to survey underwater ecosystems (Solan et al., 2003, Romero-Ramirez et al., 2016, Bicknell et al., 2016, Brandt et al., 2016, Durden et al., 2016c). This sampling tool has many advantages over other methods, being less invasive and destructive than physical sampling, preserving the species- habitat relationship, and providing a permanent record of observations given the now virtually unlimited storage capacity available to researchers at low cost. (González-Rivero et al., 2016, Chimienti et al., 2018). Cameras need to be mounted on a mobile platform so that they can cover the required seabed surface and be provided with power, light and other needed support for the system. Since imagery became popular, this platform is either a frame attached to the ship and lowered to the seabed or a remotely operated vehicle (ROV) (Clark et al., 2016b). Dropped or towed frames have the advantage of simplicity and are cost-effective, while ROVs offer better operational control when on the seabed, as well as a wider range of sampling possibilities, including the collection of physical samples (Brandt et al., 2016). Both these methods are limited in their mobility by their compulsory link to the mothership; consequently, the amount of ground they can cover can realistically cover in reasonable time is also limited (Danovaro et al., 2016). Hence, although they have contributed to significant improvement of our knowledge of the deep-ocean, it is commonly accepted that the pace of the exploration they allow is not sufficient to thoroughly explore this environment.

In contrast, Autonomous Underwater Vehicles (AUV) can operate independently from the ship without supervision, allowing the ship to carry out other operations simultaneously (Huvenne et al., 2009, Wynn et al., 2012). The absence of cables also allows AUVs to travel much faster than any gear requiring connection to the ship via cables. This dramatically increases the amount of ground that can be covered in a single mission (Wynn et al., 2014), particularly with the AUVs of the “cruising” class (Huvenne et al., 2018). Thanks to the latest development in battery life, these AUVs can stay underwater from a few hours to a few months, depending on their design and the sampling equipment deployed, and travel hundreds of kilometres, allowing them to gather a tremendous amount of data (Huvenne et al., 2018, Wynn et al., 2014, Wölfl et al., 2019).

Due to this capacity to collect so much data, AUVs can offer the possibility to thoroughly test how sample size can affect density and diversity measures and suggest a minimum sample size that all samples of a benthic community taken in the same area should match, in order to be representative and thus comparable (Soetaert and Heip, 1990). This can hopefully help form a database that can combine data from multiple cruises and studies, enabling a standardized study of the local ecology and its monitoring, in relation to the changes in the environment and implemented conservation measures.

In this study, we evaluate the impact of sample size on density and various measures of diversity in a case-study of the soft sediment epibenthic megafauna at a 1200-meter-deep station off Rockall Bank in the North-East Atlantic. We attempt to provide the necessary amount of data needed to observe and estimate the local density and alpha diversity, evaluate the variability of these measures and discuss the consequences for conservation. We also compare these ecological requirements to the capacity of the AUV used for this survey and discuss how it could impact the study and conservation of benthic ecosystems.

3.3 Method

3.3.1 Study site

In May and June 2016, an area of the sea-floor of the Rockall Trough, located between Rockall Bank and George Bligh Bank, was surveyed using the ISIS ROV and the Autosub6000 AUV. This formed station 26 of the DeepLinks Cruise (JC136) on board the RRS James Cook, within the United Kingdom’s Continental Shelf Limit (see maps in **chapter 2**). The depth of the study area was 1205 m (+- 25 m) and the terrain was relatively smooth. The substrate was fine mud and appeared homogenous.

3.3.2 Data collection

3.3.2.1 Image collection

The AUV Autosub6000 was deployed on two missions over the station: M115 was aimed at mapping the area at high resolution with a multibeam echo-sounder and M116 was aimed at taking images of the seabed (details in **chapter 2**). Autosub conducted an 86.6 km long dive during N116, in which it took more than 80000 2448 x 2048 pixels resolution images of the seabed with a downward facing Grasshopper2 GS2-GE-50S5C camera (Point Grey Research). 60000 of these images were deemed usable for analysis. The AUV was flown at 1.1 ms^{-1} speed, at 3 m ± 0.2 m off bottom and took nearly overlapping images at a 1 second frequency which provided almost full coverage of the seabed. Images taken at higher than 3.6 m off bottom were removed to avoid the risk of overlap and annotating the same area twice.

The approximate dimensions and surface area of each AUV image was calculated using the altitude, pitch and roll of the vehicle and the cameras characteristics, following the method described in (Morris et al., 2014) but since adapted to a Python (www.python.org) script. The surface of an image varies between 1 and 2.5 m^2 and averages at 1.76 m^2. Total time taken to complete the 1900 m long transect used in this study was 29 minutes.

3.3.2.2 Image processing and annotation

The .RAW images taken by the AUV were converted to .JPG format and an 80% compression applied to reduce image size without visible loss in quality with the IrfanView software (Skiljan, 2012). The colours were also adjusted in the same software using the default settings of the colour adjustment tool in the same program.

In total, 1718 raw images of the seabed were manually annotated by a single observer with the Biigle 2.0 software (Langenkämper et al., 2017), using a regional catalogue of Operational Taxonomical Units (OTUs) developed by Howell and Davies (2016), modified with several new OTUs encountered in the present dataset. The order in which the images were annotated was random to limit observer biases becoming locally correlated. Within the Biigle 2.0 software, location (X and Y coordinates in pixels for point annotations, or X, Y and radius for individuals marked using a circle) and identity of individual OTUs annotated within each image were recorded and stored. Individual annotations for all OTUs in both datasets were visually inspected using the “Largo” evaluation tool in Biigle 2.0, to ensure consistency in identification and reduce error. Thus, each annotation was checked at least once by its original annotator.

3.3.3 Data analysis

Prior to community analysis, all members of the superclass Pisces (bony fish and sharks) were removed from the dataset as their high mobility could lead the same individual to be included on several images, introducing bias in their observed abundance. All unidentified individuals (for which no identification could be confirmed at phylum level or lower) were also removed.

3.3.3.1 Quantification of sampling effort

The sampling effort quantifies the investment by the scientists to obtain their sample and it can be expressed in 3 different ways or units, each with its advantages and drawbacks.

The amount of effort is usually decided prior to data collection, at which point, the most practical way to express it is in length of transect or distance travelled by the vehicle in meters. This length was calculated as the cumulative lengths of all images within a transect. This assumes that images are exactly contiguous, which is not the case in practice and can introduce error caused by the image size calculation (estimated at 5-10%). It does not account for the width of the image.

More meaningful to the effective amount of seabed sampled and more comparable to samples from different gear, is the surface photographed during the survey. This can, in theory, by predicted from the specific properties (focal length and opening angle) of the camera used, but is likely to vary due to uncontrollable factors such as water turbidity or terrain roughness at fine and medium scales. Consequently, the exact surface sampled is more likely to be accurately known only after the data has been collected and processed.

Finally, it has been common practice in the deep-sea to express effort as number of individuals since the recommendations of Sanders (1968). This is a consequence of the relatively high variability of deep-sea benthic megafauna, which tends to relate to the quantity of surface sampled. This measure is only realistic in environment where OTUs can be counted as discrete individuals or colonies, as opposed to ecosystems where densities and overlapping continuous cover renders such distinction impossible and requires the use of cover measures to express abundances. At station 26, none of the OTU encountered needed to be quantified in cover. Furthermore, this measure allows for more meaningful standardisation across multiple gear types, when detectability of animals is inequivalently biased by the different camera systems.

3.3.3.2 Estimate of density

Density is the number of individuals per square meters (ind .m^{-2}). The raw abundance of each taxa retained for analysis were divided by the surface of each image to get the density of megafauna within that same image.

3.3.3.3 Estimate of diversity

Diversity was estimated with Hills number of orders 0 (the exact total number of OTU, species richness or Chao number), 1 (the number of typical OTUs or transformed Shannon number) and 2 (the number of dominant OTUs or inverse Simpson number) as described in Chao et al. (2014) and Chao and Chiu (2016). Rarefaction curves (Sanders, 1968) were extrapolated to a sampling effort (in number of individuals sampled) equal to double the number of individuals observed by the AUV. This was decided following recommendations by Colwell et al. (2012) who stated that richness estimates beyond 2 or 3 times the available sample size are unreliable. Asymptotic estimates were provided by the iNext package (Hsieh et al., 2016).

3.3.3.4 Effect of sample size on estimate of density and diversity

Random subsets of 1-20% (in steps of 1%), 30-90% (in steps of 10%) and 99%, of the total number of sampling units (here, images) available were created with replacement. The higher number or small size sampled was decided to increase resolution in the ascending phase of the species accumulation curve. This was repeated 100 times for each sample size. For practical reasons, the resampling was done with images as sampling units (subsample of size 10% is 171 images) but the effort of each subset was also quantified in terms of number of individuals, total surface and transect length.

The density in each of these subsets was calculated as the cumulative sum of all the individuals present in this subset divided by the cumulative surface of all the images in the same subset.

Minimum sampling effort required to observe the estimated asymptotic diversity, of order 0, 1 and 2, was determined with a GAM model of the relationship between effort and observed diversity in the subsample. In some instances, this required extrapolation beyond the existing amount of effort in the available dataset. We also determined the sampling effort needed to observe 99 and 95% of the diversity in the total dataset.

The average diversity estimate across the 100 replicates of each sample size was compared to the diversity of the whole dataset. This would indicate the smallest sample size which can reliably estimate the diversity of the full dataset. This was done for all 3 orders of diversity, as well as with and without inclusion of the singletons for diversity of order 0, to evaluate their impact on our interpretation of species richness.

In order to evaluate the robustness of this method, we also investigated how the average estimate of diversity, within this sample size, was impacted by the number of replicates used to produce it. Average estimates of diversity were calculated for random groups ranging from 2 to 100 replicates (98 groups of increasing sizes). This was repeated 100 times for each replicate size group, resulting in 9800 average estimates of diversity. This allowed visualisation of how estimates of diversity were spread around the mean and if a small number of replicates could give reliable estimate of diversity.

All statistical analyses were performed in R (R Development Core Team, 2011), within the Rstudio (RStudio Team, 2015) environment, with extensive use of the “tidyverse” library for data manipulation and graphical outputs (Wickham, 2017). Rarefaction and sample completeness curves were all computed using the iNext package (Hsieh et al., 2016). Multivariate analysis was performed using the Vegan package (Oksanen et al., 2007) and the geoveg package (Goral and Schellenberg, 2017).

3.4 Results

In total, the AUV encountered 61561 (61555 without the singletons) individuals and 113 OTUs (107 without the singletons). The average diversity was 12 OTU per image (sd = 3.05).

The local epibenthic megafauna was dominated by the xenophyophore *Syringammina fragillissima* (OTU261), a tube worm (OTU375), a small branching sponge (OTU603), squat-lobster, *Munida sp.* (OTU339), unidentified chrysogorgiidae (OTU995), cerianthid anemone (OTU2) and halcampid (OTU23) anemone.

3.4.1 Estimates of density

On average, there were 35.8 individuals per image (sd = 17.027) at an average density of 20.6 individuals per m^2 (sd = 9.69), or 2953 individuals per 100 m travelled.

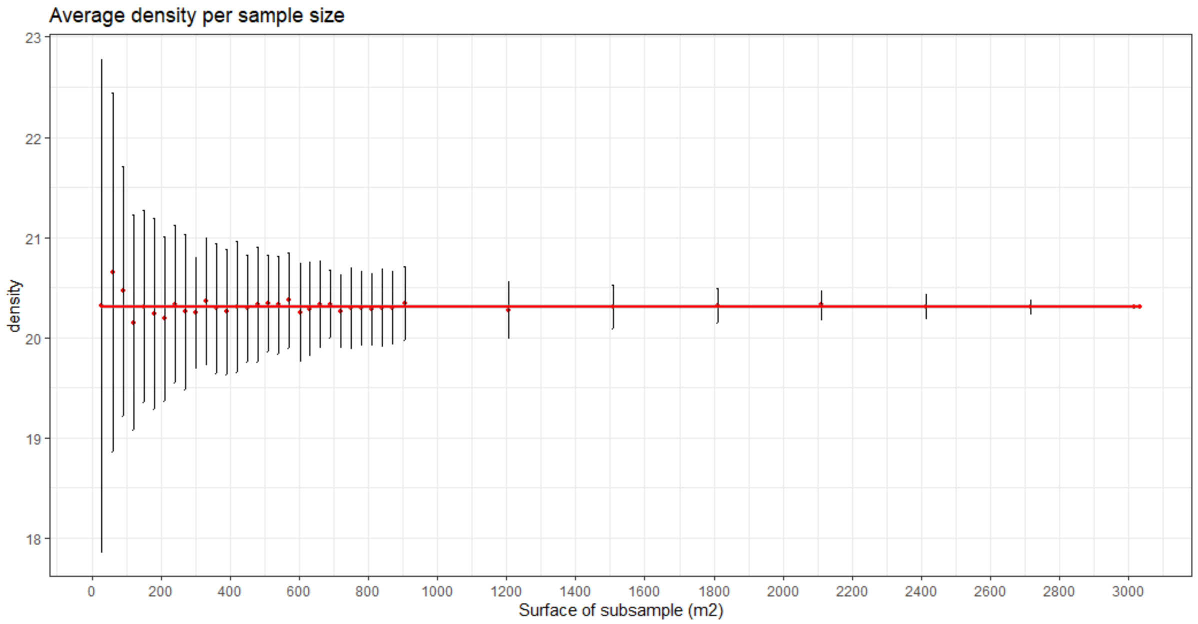


Figure 3-1: Effect of sample size on measure and variability of density at station 26. Red dots indicate the average density observed in subsamples of a given size. Black lines indicate the standard deviation of this measure. The red line is the smoothed curve of average density per subsample.

The effect of sample size, in terms of surface cover (m^2), on the density of megafauna at station 26 is shown in figure 3-1. This shows that although average density is the same across sample sizes, this measure is extremely variable within the smaller samples. The standard deviation of density decreases slowly with increasing sample size, but remains above 0 until the subsets are larger than 90% of the total pool of sample. The standard deviation decreases below 1 with sample of sizes superior to 150 m^2, 3047 individuals and 103 m long continuous transect.

3.4.2 Estimates of diversity

The Chao estimate of total diversity, 117 OTUs, was close to the number of observed OTUs, 113 (values in Table 3-1) and within the error bracket. This indicates that most of the OTUs present at station 26 have been encountered. Excluding the singletons did reduce the estimated species richness down to 107, equal to the observed species richness when singletons are ignored, although below the upper limit of the confidence interval. In higher order diversity, both observed transformed Shannon and Inverse Simpson match the corresponding index estimate, demonstrating all the typical and dominant species were encountered. Removing the singletons had little to no effect on these indices.

Table 3-1: Diversity metrics of the full dataset, with and without the singletons. The table shows observed diversity (Observed, asymptotic estimates (Estimated) extrapolated at 2 times the available effort and upper and lower 95% confidence intervals for diversity of order 0, 1 and 2.

Singletons	Type of OTU	Estimator	Order	Observed	Estimated	95% Lower	95% Upper
with	All OTUs	Species richness (Chao number)	0	113	117	114	133
without			0	107	107	107	110
with	Typical OTUs	Transformed Shannon	1	16.49	16.51	16.49	16.702
without			1	16.48	16.49	16.48	16.694
with	Dominant OTUS	Inverse Simpson	2	8.15	8.15	8.147	8.26
without			2	8.15	8.15	8.145	8.262

The rarefaction curves on the full dataset (Figure 3-2) also showed that the asymptote is not reached with the observed (interpolated) sample, although it possessed a very gradual and gently decreasing slope, that approached a plateau, with a sample smaller than the total 61561. Without the singletons, the curve of the species richness (Chao) reaches a plateau at 107 OTUs and the observed richness is stable.

The curves of higher order indices (Shannon and Simpson), climbed steeply appearing to stabilize with a smaller sample size than species richness, as both seem to approach a plateau with less than 4000 individuals (Figure 3-2).

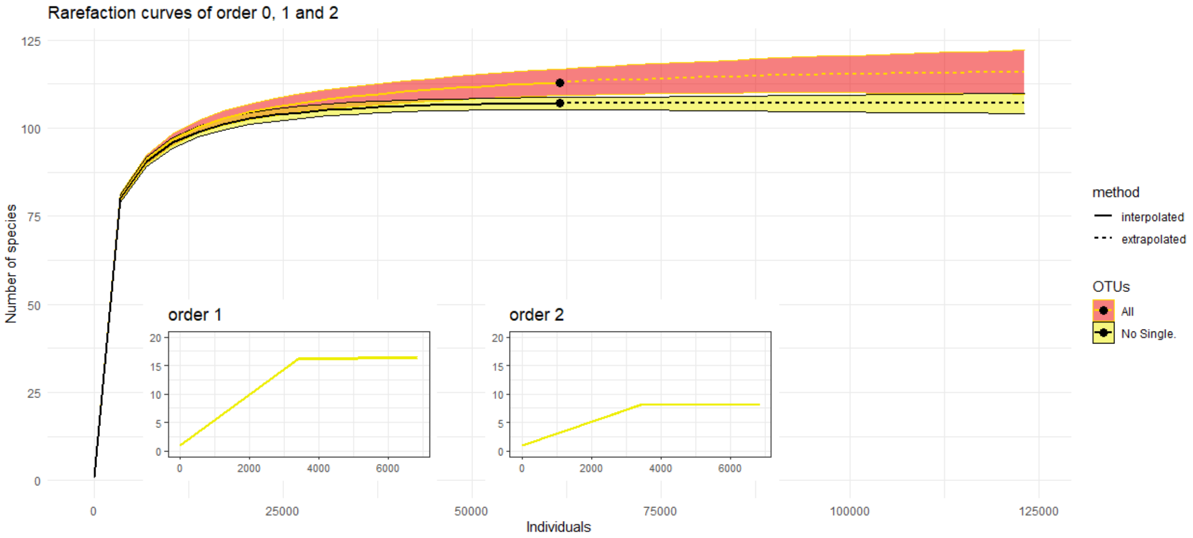


Figure 3-2: Rarefaction curves (continuous line) of the full dataset with (All) and without the singletons (no singles). Curves also show the extrapolated (dashed line) data up to 123122 individuals (order 0), 6000 individuals (order 1 & order 2). 95% bootstrapped confidence intervals are displayed as ribbons around each curve.

3.4.3 Assessment of optimum sample size

Observed and estimated diversity was shown to be affected by resampling the dataset. Table 3-2 shows the estimated amount of sampling effort needed for sub-samples to match the observed diversity (of all 3 orders) in the total dataset. 75772 individuals, 3730 m^2 or 2566 meter long continuous transect would be necessary to encounter all the OTUs present in the entire transect we annotated. 91.5 % of that effort would be enough to encounter 99% of the order 0 diversity (species richness) and 57% of that effort would be enough to encounter 95% of order 0 diversity. The higher order diversity, of typical and dominant OTUs (the transformed Shannon and inverse Simpson indices), can be encountered with less than 33% of the sampling effort and only 5635 individuals, 1208 m^2 and 831 meters long continuous transect are necessary to encounter 95% of the common OTUs. A slightly higher effort was needed to encounter 95% of the dominant OTU (Simpson index) which illustrates the error margin associated with these calculations.

Table 3-2: Estimated amount of sampling effort (in number of individuals, surface and transect length) needed for a sub-sample to encounter 100, 99 and 95% of the diversity (of order 0, 1 and 2) in the total dataset. Values were rounded to the nearest 0 decimal place. Effort is expressed in number of individauls (nb. ind.), surface (Surface (m2) , and length of continous transect (Length (m)).

	Order	Order 0	Order 0	Order 1	Order 2
	Index	Chao	Chao	transf. Shannon	Inverse Simpson
Effort unit	% of total diversity	All OTUs	No Singletons	All OTUs	All OTUs
nb. ind.	100%	75772	47897	24528	24159
nb. ind.	99%	69330	41105	21873	23267
nb. ind.	95%	43699	18354	5635	6849
Surface (m^2)	100%	3730	2358	1208	1189
Surface (m^2)	99%	3413	2024	1077	1146
Surface (m^2)	95%	2152	904	276	341
Length (m)	100%	2566	1622	831	818
Length (m)	99%	2348	1392	740	789
Length (m)	95%	1480	622	190	235

If the singletons are not considered, only 47897 individuals, 2385 m^2 and 1622 meter long continuous transect would be needed to encounter the OTUs remaining in the total dataset; a reduced sampling effort of 36.5% compared to what is needed for all the OTUs. Note that the effort at the asymptote tends to change drastically with small variation in the maximum diversity target. For example, depending on how the target diversity is rounded, it could result in a 15% increase in needed sampling effort to reach it. More importantly, should these indices be rounded to natural numbers to represent a number of OTUs rather than an index, the effort would be greatly changed.

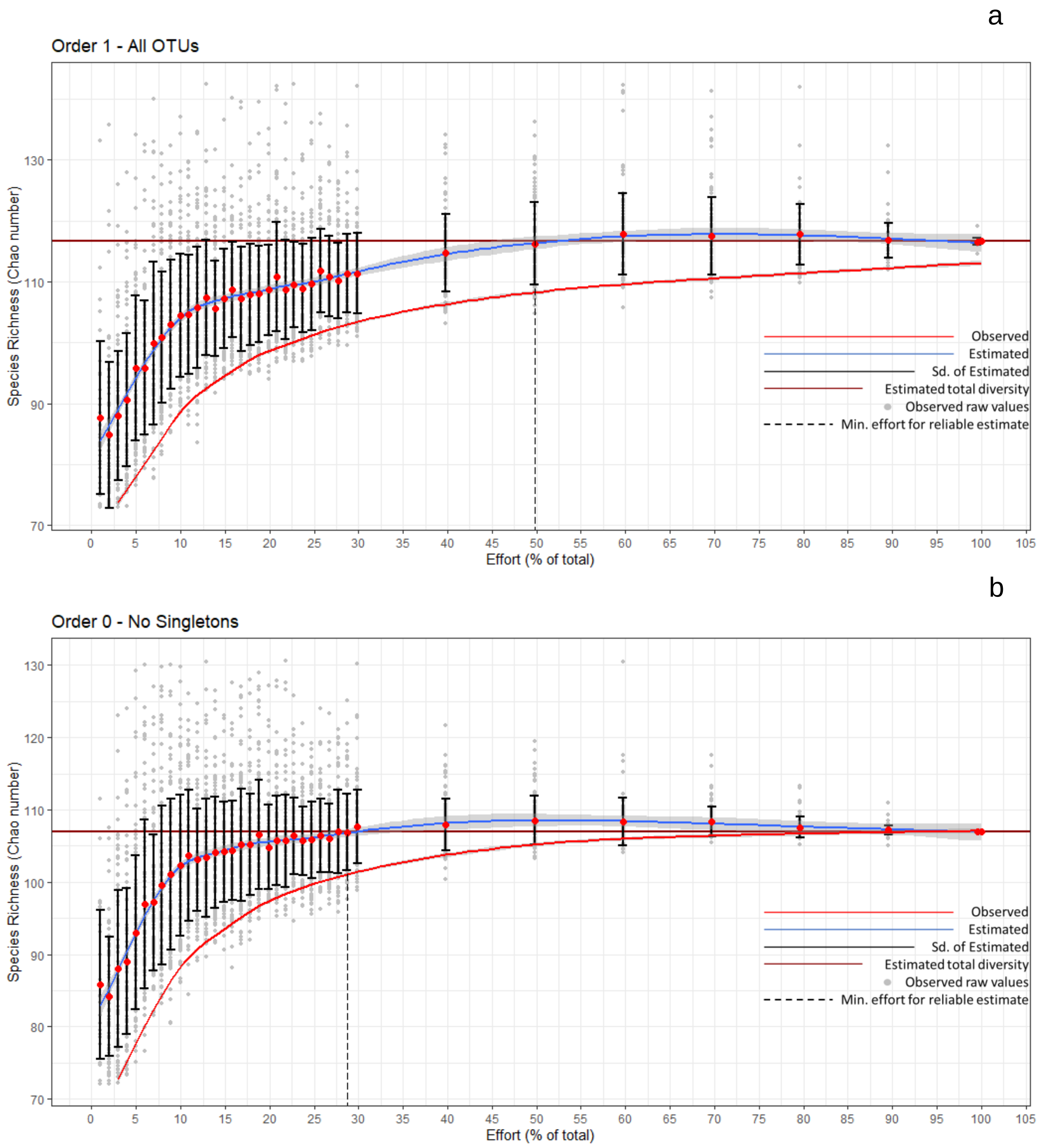


Figure 3-3: Effect of sample size on observed and estimated Chao numbers with (a) and without the singletons (b). Effort, or size of the subsample, is expressed in percentage of the total effort in the whole dataset. The dashed black line indicates the minimum effort or sample size at which the average estimate consistently matches that of the whole dataset. The error bars indicates the standard deviation.

Figure 3-3 shows the relationship between the observed and estimated diversity of smaller subsets. In the whole dataset, no subset ever observed the same number of OTUs as its associated estimate. However, removing the singletons lead the observed Chao number to match the estimate, if the sample is large enough (80% of the total effort). In both cases, the average estimated diversity, over 100 subsets, reaches a plateau with samples smaller than the total dataset. The variability of observed and estimated diversity in individual subsets is higher at smaller sample sizes. This variability decreases with increasing sample size, particularly with singletons removed. Note that, since these subsamples are taken with replacement and the larger ones are far from independent from each other which probably impacts the observed variability.

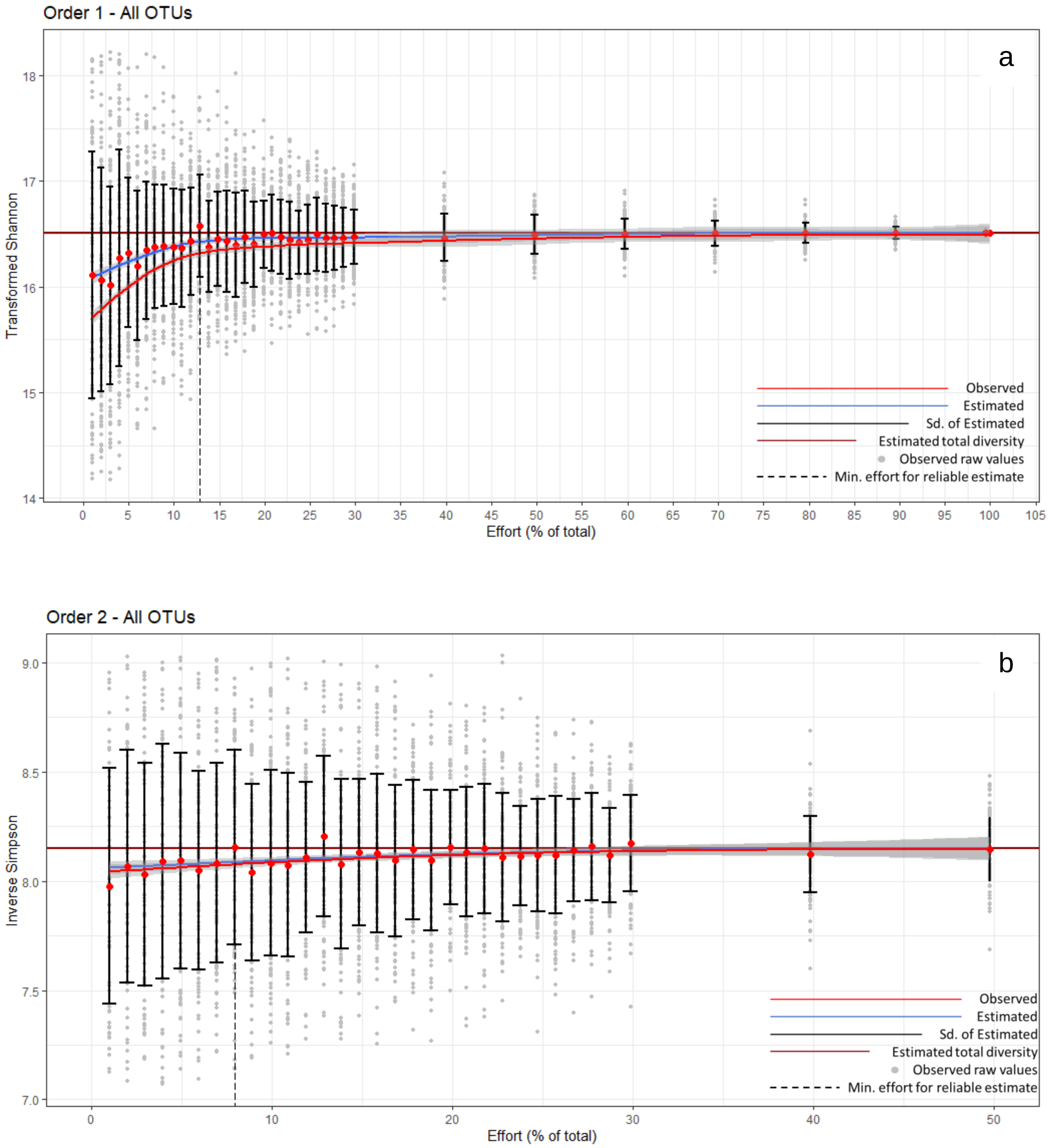


Figure 3-4: Effect of sample size on Observed and Estimated transformed Shannon (a) and Inverse Simpson (b) indices. Frame size in b was adjusted to highlight the behaviour of the curve with samples of sizes below 20000 individuals. The dashed black line indicates the minimum effort or sample size at which the average estimate consistently matches that of the whole dataset. The error bars indicates the standard deviation.

With the transformed Shannon and inverse Simpson indices (Figure 3-4), the minimum sample size that, on average, matched the observed diversity in the whole dataset was difficult to read due to the variability of the measures. Although very small samples can give the correct estimates, the average estimated diversity is not consistently equal or above the target diversity. For both inverse Shannon and inverse Simpson indices, the average estimates of sample of sizes larger than the minimum indicated in Table 3-2 are below that of the whole dataset. Nonetheless, these plots and table do show that a lower effort is needed to obtain the correct estimate of diversity of order 1 and 2, however, the variability of these estimates is relatively lower than those for order 0 diversity.

The smallest subsets size that provided an estimate of order 0 diversity that was equal or larger than the estimated species richness of the whole datasets, with and without the singletons, are indicated in Table 3-3. With all the OTUs, the smallest subsets that, on average, provided that same estimate of total richness (117 OTUs), had 855 images, and represented 50% of the total effort, 30624 individuals, 1508 m^2 or 1037 meters of continuous transect by the AUV. Without the singletons, 493 images were, on average, enough to match the estimated and observed total richness in the dataset (107 OTU). This represents less than 30% of the total effort, 17670 individuals, 870 m^2 or 598 m of meters of continuous transect by the AUV. In both cases, the estimate of individual replicates at that minimum could be above or below the average by up to 15 OTUs.

Table 3-3: Minimum amount of effort in different units for the average estimate over 100 random subsamples to match the estimated diversity in the whole dataset (target richness).

		target richness	nb. Individuals	Surface (m^2)	Length (m)
All	order 0	117	30624	1508	1037
No single.	order 0	107	17670	870	598
All	order 1	16.509	7897	390	268
All	order 2	8.148	4878	240	165

The variability of the estimates decreased with the number of pseudo-replicates used to produce the average estimate. Figure 3-5 shows how the number of replicated sub-samples (randomly sub-sampled with replacement) impacts the standard deviation of the estimated diversity.

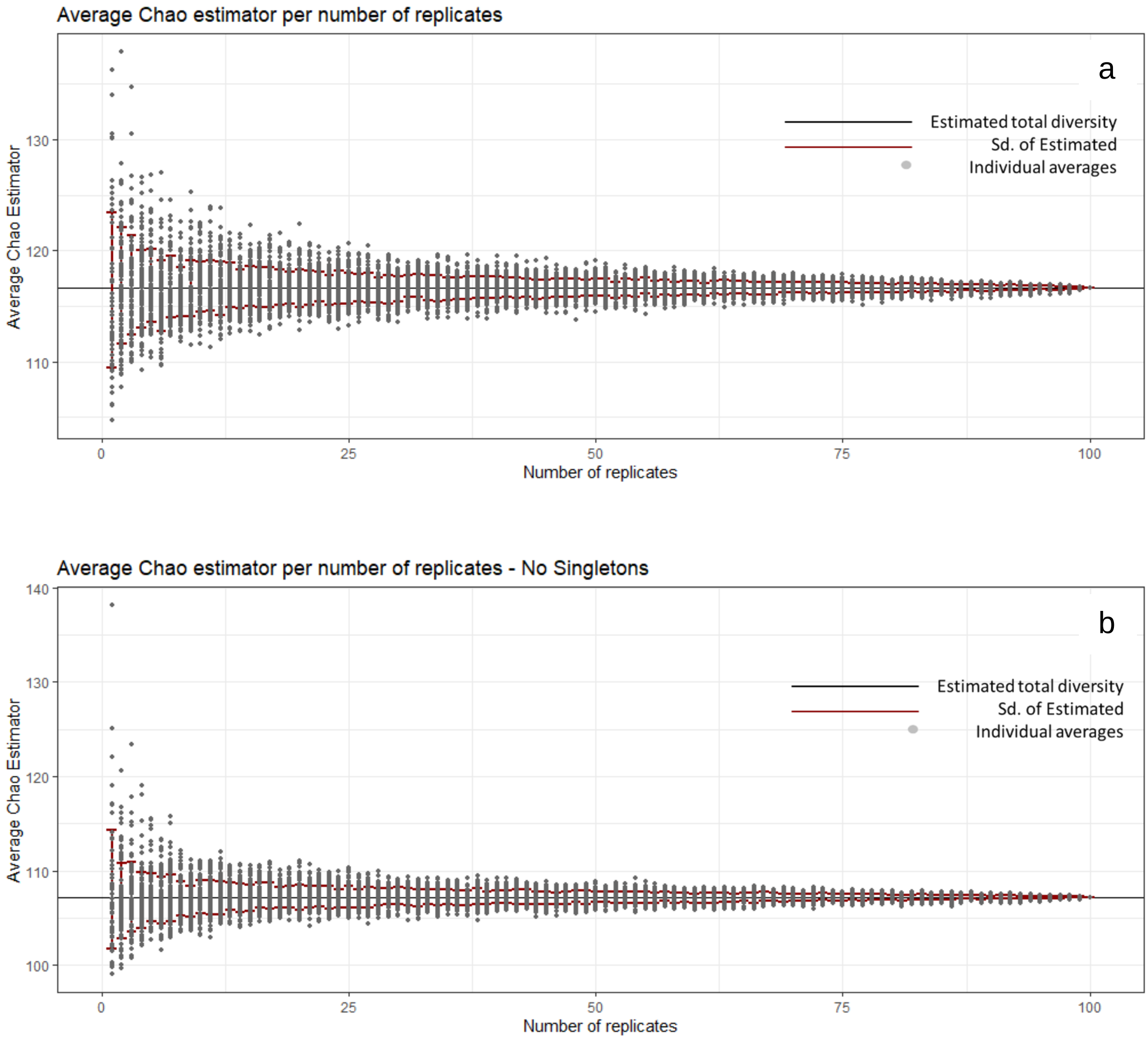


Figure 3-5: Variability of order 0 diversity (Chao) estimates, averaged over increasing number of replicates. The horizontal black line indicates the estimate of the whole dataset with (a) and without (b) the singletons.

Figure 3-5 demonstrates the strong relationship between the variability of the estimated diversity averaged over a number of small random replicates and the number of said replicates. With all the OTUs (Figure 3-5a), averaged estimates of diversity vary widely (+/- 3 OTUs or more) from that of the full dataset when less than

20 replicates were used. For example, the averages of only three replicates could be more than 8 OTUs above or below the diversity of the whole dataset. Removing the singletons did not affect that variability (Figure 3-5b). The same pattern exists with measures of diversity of higher order (plots not shown here).

3.5 Discussion

In this study, we investigated how sampling effort affects measures of density and diversity, in a soft sediment epibenthic community, in the Rockall trough. We also attempted to evaluate the amount of effort needed to fully sample the community.

3.5.1 Impact of sample size on density measures

This study showed that average density (20.6 ind. m^{-2}) can be accurately measured (+/- 0.5 ind.m^{-2}) within a 150 m^2 sample. Although larger samples were shown to give less variable measures of density, this comes at the cost of an increase in effort needed. This illustrates how sampling units, the size of a single image (1.7 m^2), are inappropriate to measure density in this environment, and must be pooled to achieve a sufficient sample size.. Within very large sample sizes, in which sub-samples are less different (as they are sampled with replacement), the level of variability does not decrease to 0. This is likely driven by the patchiness of species distributions, a well-known characteristic of deep-sea ecosystems (Grassle and Maciolek, 1992b, Rex and Etter, 2010, McClain et al., 2011). A more detailed investigation of the heterogeneity in abundance of some of the megafaunal species at station 26, and their possible causes, is conducted in **chapter 4**.

It is important to know if the observed difference in density, between two samples, is within the natural variability of the local megafauna density, or if it could be caused by another factor. In a hypothetical scenario, where a similar study is conducted in the same location 10 years after this one, which measured a difference in density of 1.3 ind.m^{-2} (down to 19.3), this could mean there is a genuine decrease in density, if the sampling units are large enough. However, if these new results were based on measurements over 25 m^2, this would be within the range of variability expected for a sample of that size and, therefore, inconclusive. On the contrary, if the samples represented more than 200 m^2, this observed difference could be a sign of a genuine decrease in species density.

Density can be used to detect broad changes in communities (Ardron et al., 2019). The small sample size needed to accurately measure density and quantify its variability, makes it useful for rapid detection of changes in ecosystems but carries less information than diversity.

3.5.2 Impact of sample size on diversity measures

From our data, it seems possible to observe all OTUs present at station 26 in a sample of 75000 individuals, 3730 m^2 of seabed or 2600 meters of continuous transect by the AUV, using the same settings as this survey. Less than 60% of that effort would be needed to encounter 95% of the total estimated number of OTUs. We also found that 50 % of that effort, less than 31000 individuals, 1500 m^2 and a continuous AUV transect slightly longer than a kilometre, could give a correct estimate, if replicated. This is not to say that this is a large enough sample to provide an exact estimate of total species richness, as the variability between estimates was as high as 15 (+/-) OTUs. We found that only a high number of replicates or pseudo-replicates (more than 20) was enough to ensure that the average richness was close to the true value. The use of true replicates would be preferable but pseudo-replicates can also provide useful insight while keeping sample at logistically realistic size.

Quantifying the variability of diversity measures in the area is valuable information when comparing samples from different surveys. It is important to note here that quantifying variability would be better with independent sub-samples (Hurlbert, 1984, Noble-James et al., 2017). Unfortunately, our dataset was too small to investigate that effect in detail, we therefore had to use pseudo-replicates. Nonetheless, it shows that reporting raw diversity measures, without quantifying variability, may be misleading. Measures of diversity should therefore be replicated whenever possible (Lacharité and Brown, 2019).

We also calculated that less than 25000 individuals, 1200 m^2 and less than 850 meters long transect, would be needed to obtain the same transformed Shannon or inverse Simpson Indices. Less than 30% of the effort needed to estimate diversity of order 0 is needed to measure diversity of order 1 and 2. The variability of the higher order diversity indices is not null, but is relatively low compared to the species richness. Hill’s diversity numbers have previously been described as less dependent on sample size (Soetaert and Heip, 1990) and their response to changes in community composition vary (Nagendra, 2002). They are however, complementary to species richness as they give an indication of the evenness of the diversity (Soetaert and Heip, 1990, Jost, 2006, Simon-Lledó et al., 2019). In the case of station 26, they show how a small number of OTUs dominate the community and that the majority of the OTUs present are rare; as commonly reported in the deep-sea (Poore et al., 2015, Carney, 1997). As these indices account for evenness, they could vary if the relative abundance of the main taxa changed, but not the species richness (Ardron et al., 2019). This information would be a useful application of these indices, particularly if the available sample is not large enough to enable measurement of species richness. It also illustrates how the use of several indices would favour a comprehensive understanding of this ecosystem.

3.5.3 Advice on sampling effort

The methodological advice detailed here aims to increase the rigor of future sampling methodologies according to the goals set by conservation managers in the UK (SCOR, 1994, Kroeger and Johnston, 2016, Noble-James et al., 2017) and globally (Woodall et al., 2018, Levin et al., 2019).

To study the species richness near station 26, or in similar ecosystems, we therefore recommend to aim for a minimum effort between 30000 – 40000 individuals (1500 – 2100 m2 or 1000 to 1400 meter long transects), to allow a safety margin. In addition, replicating this sample at least three times (more if possible) or alternatively use random resampling with replacement, would quantify the variability. This would come to a total of 4500 to 6300 m^2 and 3000 to 4200 meters of continuous transect. This should give a reliable estimate of the total number of OTUs present in the station and these three replicates can be pooled together to produce a species accumulation curve. This will determine if the richness estimate from the replicates can indeed approximate that of a larger sample, within the error margin. A third of that effort would be sufficient to measure density or diversity of a higher order, but all components of diversity should be studied to obtain a representative assessment of an ecosystem (Davies et al., 2017).

The AUV observed 95% of the total species richness with 44000 individuals, but would need nearly twice that amount of effort to encounter the remaining 5%. This single measure of diversity would not quantify the variability of this measure and another transect of the same length and surface is likely to encounter a different number of OTUs. Hence, there is a diminishing return in annotating a large amount of images, as a large part of the effort would be dedicated to the detection of a small number of rare OTUs. The extra effort to observe all OTUs may not be an optimal use of ship, crew and vehicle time. Instead, 40% of that effort, approximately 30500 individuals, if repeated several times could, at least give a reasonable average estimate of the diversity, as well as a measure of the variability; the price being ignorance of the some of the rarest taxa.

3.5.4 Consequences for use of AUVs in deep-sea ecology

The recommendations we make here represent a consequent amount of sampling effort for a single station. It is nonetheless well within the sampling capacity of a cruising AUV such as Autosub6000. In fact, the entire dataset collected during dive M116 along with the data used in this study cover more than 93500 km2 and 80km of continuous transect (probably more than 2.5 million individuals). AUVs are seemingly the perfect tool to cover large amount of ground efficiently and sample images of flat seabed. It is particularly suited for studies where sampling unit size consistency, reproducibility (Borregaard and Hart, 2016), speed and sample size (Perkins et al., 2019, Perkins et al., 2016) are more important than taxonomic resolution and the possibility to collect physical samples, like MPA monitoring or environmental impact assessments (Smale et al., 2012, Wynn et al., 2012, Wynn et al., 2014).

However, the major advantages of AUVs at present only apply to well-known areas with limited risks of collision with large object on the sea-bed (Wölfl et al., 2019). Other disadvantages include the complete absence of physical sampling capability or the impossibility of further investigating objects of interest during the mission. Furthermore, given the lack of efficient satellite positioning over great depth, dead-reckoning navigation used by most AUVs is subject to a certain amount of drift, which can be up to hundreds of meters. AUVs are likely to be increasingly used in the future for surveying and monitoring deep-sea benthos (Williams et al., 2016, Lucieer and Forrest, 2016, Huvenne et al., 2018) and more and more data will be collected by these vehicles (Huvenne et al., 2018, Wynn et al., 2014, Milligan et al., 2016, Morris et al., 2014).

The amount of data recommended here is currently largely impractical for a deep-sea survey due to the bottleneck that exists in in data analysis because images collected still have to be annotated manually (Schoening et al., 2017, Howell et al., 2019, Durden et al., 2016b). Thus, an important change these vehicles can help bring about is a shift of the bottleneck limiting scientists capacity to sample the deep ocean further down the processing pipeline. This is the biggest challenge to the adoption of AUVs as a common sampling tool that currently prevents full use of the tens of hundreds of thousands of images taken by these vehicles (Brandt et al., 2016, Schoening et al., 2017). Indeed, with a hundred or a thousand-fold increase in sampling effort, the data collection phase may not be limiting research in the near future provided it can be analysed. The cost of gathering data can average around 40000 € (36000 Pound sterling) per day (Brandt et al., 2016), analysis time often relies on a small number of individuals and costs a fraction of that. Thus, unlike the data gathering, the data analysis is a more adjustable variable that can be reconsidered a posteriori on a more informed base. This, for example, opens the possibility to continue annotating until targets of diversity are met. This in turn makes harmonisation of effort more realistic as well as more adaptable to local conditions or other parameters known only once the collection phase is complete. Currently at prototype stage but more promising in the future, automated image analysis can also unlock the bottleneck of data analysis (Gaston and O'Neill, 2004, Beijbom et al., 2015, Weinstein, 2018). We strongly advocate in favour of further research on that topic (this will be investigated in **chapter 5**).

3.6 Conclusion

Strictly interpreted, these results indicate that a high amount of effort is needed to obtain a reliable and complete inventory of the taxonomical composition of the benthic megafauna at station 26. Measures of density and diversity of higher order (Shannon et and Simpson indices) need less effort to be reliably measured and these metrics could be considered a useful alternative if the objective of the study can be accomplished with these measures. If a complete list of OTUs is needed, 75000 individuals need to be detected and identified which represents more than 20% more effort compared to what has been performed in this study but replicated samples of 50% of that effort can also provide an accurate estimate while also quantifying the variability of diversity measures.

In the same manner as the arrival of cameras and ROVs complemented existing physical sampling tools like trawls and dredges, AUVs will provide a different view-point of deep epibenthic communities and offer a fresh perspective. AUVs are likely to take an increasingly important role in surveying the deep bathyal and abyssal plains due to their unmatched capacity to collect large amount of data, subsequently enabling robust quantitative studies while optimising the ship and crew time. As methodological rigour requirements for sampling the deep-sea increase (Woodall et al., 2018), these vehicles appear to offer appropriate tool to enable these changes.

Chapter 4: Fine scale distribution of benthic megafauna: a case study in Rockall basin

4.1 Abstract

The fine scale distribution of benthic species and its environmental drivers are an important element of deep-sea ecology, but their study requires a large amount of data that is difficult to acquire. Autonomous underwater vehicles may offer a solution but have yet to be tested in practice. In this study, we tested if the density of the Xenophyophore *Syringammina fragilissima,* measured by an AUV in 6500 images, from four neighbouring transects, could robustly describe the spatial structure of their distribution. We then attempted to extrapolate the pattern observed within the transect, to the rest of the survey area using species distribution modelling techniques. The aim being to determine if the large dataset available was sufficient to produce a reliable modelled map of the distribution of S. *fragilissima*, as well as describe the influence of environmental drivers on its distribution. We thoroughly tested the models with internal and external validation and detected structure in their distribution with densities varying from 0 to 26 ind.m^{-2}, within the study area. The spatial autocorrelation pattern was consistent across all four transects (with minimum distances of 243 to 335 m). The models performed well according to internal validation (Rsquared 0.5 to 0.7). However, external validation with relatively independent test dataset suggested that models trained on smaller subsets were never able to accurately predict the density of S. *fragilissima* (no Rsquared above 0.32, most below 0.05), in other more closely located subsets. This prompts questions on the generalisability of the model trained on the entire dataset. This study suggests that AUVs are suited to study fine scale

distributions of species and their environmental drivers, but big datasets must still come from well-designed surveys.

4.2 Introduction

The deep sea, by convention, the portion of the oceans deeper than 200 metres covers most of the planet and provides many ecosystem services (Borja et al., 2016). It is poorly known (Danovaro et al., 2016, Van Dover et al., 2014) but it is also under increasing anthropogenic pressure (Halpern et al., 2007) and the international scientific community is racing to acquire relevant data to implement effective conservation strategies (Poore et al., 2015, Folkersen et al., 2018, Danovaro et al., 2017a, Levin et al., 2019). The evolution and resilience over time of this complex and vast ecosystem depends on many parameters. This includes abiotic environmental parameters such as topography, temperature, salinity etc… (Brown et al., 2011), as well as biotic parameters such as connectivity (Hilário et al., 2015), diversity (Costello et al., 2010), and species distribution at various temporal and spatial scales (Woodall et al., 2018, Brind'Amour et al., 2009, Glover et al., 2010).

Species distribution patterns at fine scale (finer than a kilometre) are poorly documented in deep-sea benthic ecosystems. This field of study was pioneered in the 1970s (Jumars and Eckman, 1983, Jumars, 1976, Jumars, 1975a, Jumars, 1975b, McClain et al., 2011, Rex and Etter, 2010) but specialists still acknowledge the lack of conclusive progress in the area and advocate for more research on the topic (McClain et al., 2011, McClain and Rex, 2015). Random species distribution is rare in nature (Taylor et al., 1978) and tends to follow specific patterns (Fortin and Dale 2009; Dale and Fortin 2014). Understanding the ecological dynamics and functions of an ecosystem is important to the design of efficient and unbiased sampling that provides appropriate information on which conservation strategies can be based (Brind’Amour et al., 2018, Brind’Amour et al., 2009, McClain and Rex, 2015, Rex and Etter, 2010, Robert et al., 2016, Lo Iacono et al., 2018).

Species distribution studies at fine scale require large amounts of data to represent robustly their variability, but data on deep-sea species is difficult to acquire in quantity (Ramirez-Llodra et al., 2010, Yates et al., 2018). Deep-sea epibenthic megafauna is increasingly studied by using image-based survey (Solan et al., 2003, Brandt et al., 2016). The low density of many deep-sea species necessitates that many images are analysed (or annotated) before sufficient individuals have been encountered and the abundance reliably measured (Perkins et al., 2016) or the full range of variability in abundance has been observed (Rex and Etter, 2010, McClain and Rex, 2015). Robust sampling designs also necessitate replication in the measures made (Chapman and Underwood, 2008), which multiplies the size of datasets by at least a factor of 3, hence raising the challenge of sampling by as much.

Mapping is popular output type of ecological research used to communicate results to policy managers (Brown et al., 2011, Buhl-Mortensen et al., 2015). This approach allows researchers to summarise important information of various nature in an intuitive way. Ecologists have used species distribution modelling (SDM) to extrapolate available knowledge gathered from a limited number of samples to larger areas where data is missing (Elith and Graham, 2009, Guisan et al., 2013). This process is commonly conducted at large scale (Howell et al., 2011, Ross and Howell, 2013) but fewer instances exist at fine scale in the Marine realm. This is due to the rarity of biological and environmental data at appropriate resolution (Rengstorf et al., 2012, Rengstorf et al., 2014, Rowden et al., 2017) and the need for abundant data to improve model performances (Wisz et al., 2008). Validation of SDMs with independent data is strongly advocated by the community of SDM users, as internal validation tends to give over-optimistic performance results (Anderson et al., 2016, Kenchington et al., 2019, Robinson et al., 2017, Elith et al., 2006). However, it is rarely achieved owing to the difficulty of funding studies designed only to confirm previous discoveries. Thus, there is a need to both improve our knowledge of fine-scale species distributions and with that, our ability to predict it through gathering larger datasets over smaller spatial areas. There is also a need for a better understanding of model performances in order to assess their reliability for conservation. This requires careful testing and is easier to perform with large datasets.

Autonomous Underwater Vehicles (AUV) now have the capacity to cover large areas in a short time, compared to other vehicles, and effectively provide more data while cutting the costs of sampling (Huvenne et al., 2018, Wölfl et al., 2019, Jones et al., 2019, Morris et al., 2016, Morris et al., 2014, Wynn et al., 2014). With more large datasets, questions that were previously difficult to address with statistical robustness can now be investigated and there is hope that new insights on benthic species distribution can be gained using these tools (Brandt et al., 2016, Danovaro et al., 2014, McClain and Rex, 2015). AUVs have the capacity to gather not only images, but other types of environmental data like hydrographic, oceanographic and topographic data from the exact same location (Wynn et al., 2014), thus, providing better access to the fine-scale environmental data sought by ecologists.

In this study, we explore how a large dataset of both fine-scale species distribution and environmental data collected by an AUV at a single station can inform our understanding of the distribution of the numerically abundant xenophyophore *Syringammina fragilissima* (Brady, 1883) - referred to as *S. fragilissima* thereafter - in the bathyal part of Rockall basin.

At up to 20 cm in diameter, *S. fragilissima* is possibly the largest single-celled organism on the planet. It lives on areas of soft sediment, deep in the Atlantic Ocean, and can form highly dense aggregations (Bett, 2001b, Hughes and Gooday, 2004); sometimes dominating the benthos as the main habitat-building organism (Howell et al., 2010b). They are usually associated with areas of high surface productivity, which supply abundant organic carbon to the seabed (Tendal, 1972, Levin and Gooday, 1992b). At very broad scales, they are thought to live near geological structures such as banks and margins (Tendal, 1972, Levin and Gooday, 1992b) but can also be found near canyons (Gooday et al., 2011) and seamounts (Davies et al., 2015). At finer scales, they are found near caldera, sediment mounds and walls (Tendal, 1972, Levin and Gooday, 1992b). Their distribution is also known to vary over very short distances, in areas of complex topography and sedimentology, such as in the Darwin Mounds in the North Atlantic (Bett, 2001b).

Little is known of their physiology, but they possibly grow by burst (Gooday et al., 1993). Their agglomerated tests can form a 3-dimensional frame, which is known to house other meiobenthic taxa. Additionally, a higher diversity of endofauna has been in observed in the sediment directly surrounding them (Hughes and Gage, 2004, Levin and Gooday, 1992a, Levin, 1991). The fragility of their structure makes them particularly vulnerable to physical damage and unsuitable for trawl-based studies (Roberts et al., 2000). This has also justified their addition to the list of Vulnerable Marine Ecosystems (VME) indicator taxa, under the United Nations resolution 61/105, and states are encouraged to map their extent and distribution (Ospar, 2008) in order to suitably protect them. Predictive modelling of their distribution has been attempted globally and at broad scales (Ashford et al., 2014) and slightly more locally, at finer scales (Ross and Howell, 2013, Piechaud et al., 2015), however they have not received the same level of attention as reef-forming scleractinians.

S. fragilissima is a suitable model organism for studies of fine-scale distribution for several practical reasons including its ease of detection and identification on images and its density in the study area. Extending knowledge of their ecology and fine scale distribution will directly help inform conservation strategy. We investigate the following questions:

- Is there structure in the fine-scale distribution of *S. fragilissima* within the survey area?
- Is the observed structure adequately explained by fine-scale environmental data collected in-situ? And thus, are we able to produce more accurate SDM models (and maps) at very fine scale?
- What do AUVs, and specifically their ability to gather large high-resolution datasets, offer to future studies of fine scale pattern in deep-sea benthic ecosystems?

4.3 Method

4.3.1 Survey site and image collection

All data were collected at station 26 of the DeepLinks Cruise (details in **chapter 2** of this thesis) on the 29/05/2016, during mission M116 of the Natural Environment and Research Council’s (NERC) AUV Autosub6000. During this mission, the AUV spent 22 hr in the water and approximately 18 hr near the seabed travelling a distance of 82 km. It performed 28 transects ranging from 1.7 to 3.2 km in length, of which 4 were subsequently analysed (Figure 4-1). While proceeding with these transects, the vehicle was programmed to maintain an altitude of 3 m above the seabed, at a speed of 1.1 m.s^{-1}. Taking an image every second, it gave near full coverage of the seabed. Images taken when the vehicle was higher than 3.5 m were excluded from the analysis to avoid the risk of overlapping images. Images taken when the vehicle was lower than 2 m were also removed because it risked biasing the detection rate of the smallest individuals that are less visible from a higher altitude above the seabed. The time spent by the AUV to take the images used in this study was less than 3 hours. Transects were parallel with each other except for t3, which started close (10 m) to t2 and terminated further away from t2 (90 m) but closer to t6. The distance between t2 and t6 is 180 m, the distance between t2 and t14 is 560 m.

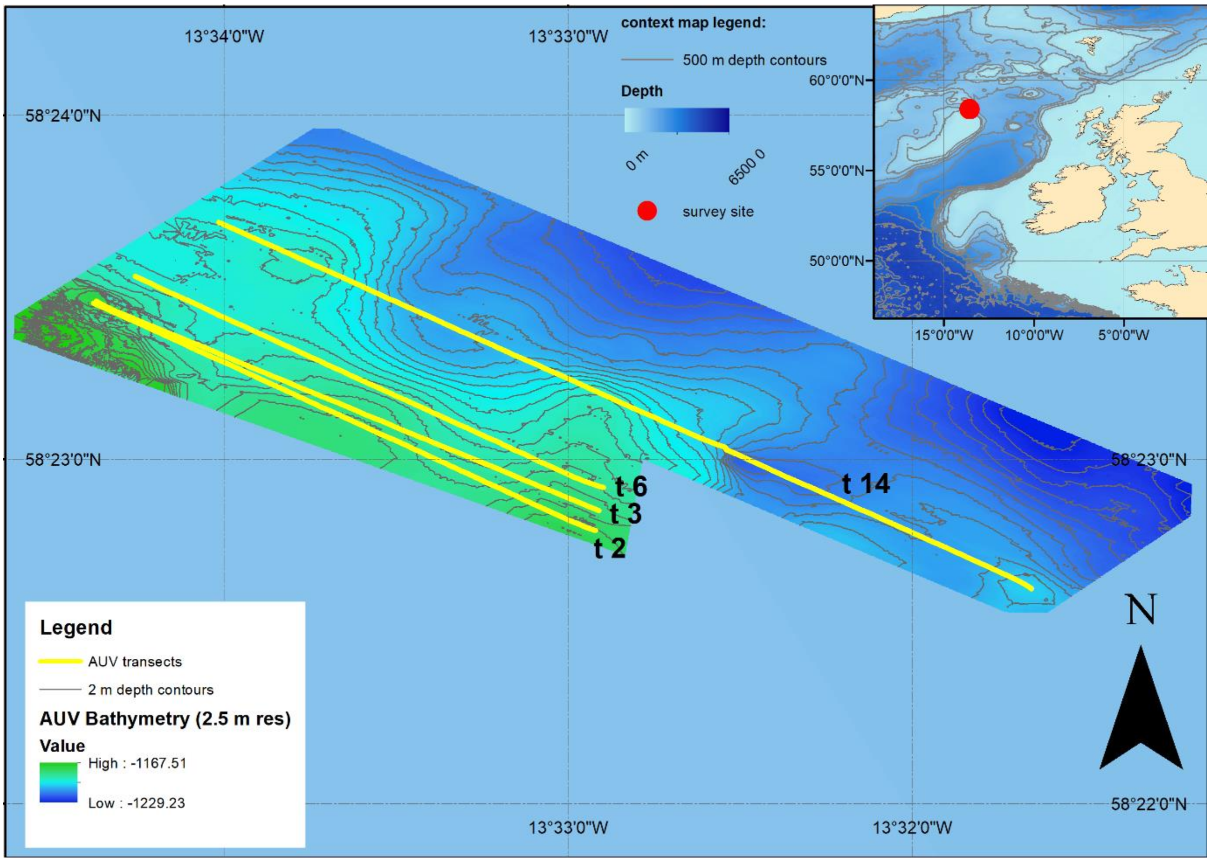


Figure 4-1: Map of the study area in the general context of the Rockall Trough (insert map, top-right corner) indicating the AUV transects annotated as well as the local topography.

Multibeam bathymetry and backscatter data at 2.5 m resolution was collected by Autosub6000 during dive M115 of the same cruise (Figure 4-1).

4.3.2 Image processing and annotation

Raw images were downloaded from the vehicle at the end of the dive and colour correction was applied within the IrfanView software (Skiljan, 2012). The images were annotated by two analysts (Nils Piechaud annotated t2, t3 and t6 and Jamie Cowle annotated t14) within the Biigle (Langenkämper et al., 2017) software using a combination of manual and automated image analysis. Transect t2 was entirely manually annotated. For transects t3, t6 and t14, a small portion of the images were manually annotated to locate at least 1000 individuals. These records were then used as training data in a MAIA (Zurowietz et al., 2018) object of interest detection session with default parameters and the number of clusters set to 1 or 2. The annotation candidates proposed by the algorithm were manually inspected (or refined) so that all individual identification was visually validated. Subsequently, all individual images were visually inspected again to check for individuals that the algorithm might have failed to detect, align the annotation location on the centre of the individual and remove possible duplicates. The two analysts worked together to ensure consistency of identification of *S. fragilissima* in the annotation candidates. The resulting dataset has an identification accuracy above 99% but the true detection rate has not been quantified.

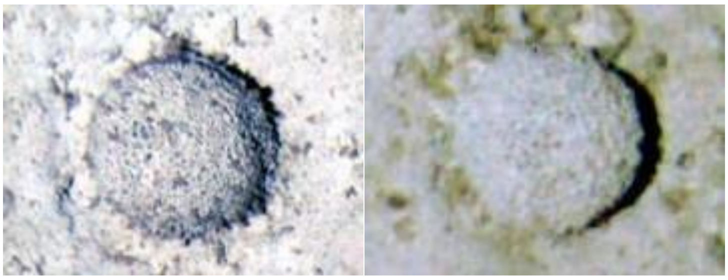


Figure 4-2: Examples of Syringammina fragilissima from AUV images taken during the JC136 – Deeplinks cruise at station 26

The surface area of each image was calculated following the method designed by Morris et al. (2014) and adapted to a Python (*python.org*) script. The abundance of *S. fragilissima* was then converted to density per square metre to compensate for the variation in altitude, pitch and roll of the vehicle and thus standardize the measure.

4.3.3 Environmental variables

We used some of the predictor variables collected by the AUV during the same mission (details in **chapter 2**). Autosub collected acoustic data includeing both backscatter and bathymetry. Bathymetry was not retained as it is correlated to all the other topographical variables. Temperature, salinity and concentration of dissolved oxygen were also excluded as they are unlikely to be drivers of benthic communities given the magnitude of their variability within the study area.

All layers were converted to raster layers at 2.5 m resolution. The bathymetry was converted to several topographic predictors, including slope and Bathymetric Position Index (BPI) (Wright et al., 2005). We used 4 different combinations of inner and outer radii to compute BPIs, 1 to 5 cells, 5 to 25, 10 to 50 and 50 to 250, in order to reflect the variation of topography at different scales. We also used the turbidity measured by the vehicle during the dive, expressed in Formazin Turbidity Units (FTU). This predictor was converted to a continuous layer at the same resolution as the other parameters with a natural neighbour interpolation. Variables and their subsequently used acronyms are listed in Table 4-1.

Table 4-1: Names and description of all the environmental predictors gathered by the AUV at station 26 and used in modelling

Predictor	*Full name*
backscatter	swath backscatter
turbidity	turbidity (FTU)
fbpi1r5	Fine scale bathymetric position index - inner radius 1 , outer radius 5
fbpi5r25	Fine scale bathymetric position index - inner radius 5, out radius 25
bbpi10r50	Broad scale bathymetric position index - inner radius 10, outer radius 50
bbpi50r250	Broad scale bathymetric position index - inner radius 50, out radius 250
slope	Slope

4.3.4 Data analysis

4.3.4.1 Detection and description of spatial structure

To visualize the spatial distribution of *S. fragilissima* and its variation at station 26, its average density within grid-cells of 25 by 25 m was mapped after log transformation and overlaid with local topography.

To test statistically for the presence of spatial structure within the distribution of *S. fragilissima* through quantification of the Spatial Auto Correlation (SAC), we computed Moran's I coefficient on 100 distance classes within each transect. The Moran's I index ranges from 0 when there is no correlation to 1 when the values are perfectly correlated. In addition, we estimated the minimum distance for the Moran's I coefficient to fall to 0 or become non-significant.

4.3.4.2 Fine scale modelling and mapping of S. fragilissima distribution

The relationships between the density of *S. fragilissima* and the individual predictor variables gathered by the AUV were investigated with univariate regression plots.

We used a Multivariate Environmental Similarity Surface (MESS) map (Elith et al., 2010) to remove the areas of station 26 where environmental predictors were outside the range of the training set.

Prior to modelling, *S. fragilissima* density data were reduced (averaged) to one point per cell in the predictor layers.

Different modelling methods can give different outputs (Piechaud et al., 2015). . Ensemble modelling, the combination of the output of several algorithms compensates the relative inconsistency of their individual predictions, mitigates overfitting, and is increasingly used in ecology (Robert et al., 2016, Rowden et al., 2017, Araújo and New, 2007, Berk, 2006). We used three popular modelling algorithms: radial kernel Support Vector Machines (SVMradial), Random Forest (rf) and extreme gradient boosting (xgbtree) to both determine the importance of environmental variables in determining species distribution and predict species distribution across the study area. These models were then aggregated via an Elastic-Net Regularized Generalized Linear Model (glmnet) implemented in the caretEnsemble package (Deane-Mayer and Knowles, 2016). We used the default training parameters supplied by the Caret package (Kuhn, 2008) for each of these methods but also enabled the cross-validation based optimizing algorithm built in Caret. This optimal model selection was based on the R-squared metric. Each of these algorithms have built-in routines to rank the relative contribution of each variable to the predictions. These scores and their combinations were used as a measure of the importance of each predictor in the distribution of *S. fragilissima* within the related training set.

To evaluate the models accuracy, we performed both internal and external validation. For internal validation, we used a 10-fold cross-validation with random subsets of the data used in training to calculate the residual mean squared Error (RMSE) and the R-squared. In this procedure, data points from all transects were included in the training and testing sets. For external validation, we used two slightly different procedures. First we trained a model on 3 transects and tested it on the remaining one. Secondly, we trained a model on one transect and tested it on each of the other transects so that all combinations of training and testing transects were used.

We tested for SAC in the residuals of the models to assess if the predictions were influenced by the spatial structure of the response variable. This was conducted with the same method used to test for SAC in the samples.

We also tested for correlation between the performances of the models and the number of data points used in training.

Finally the model trained on all the data was projected in the entire study area to produce a tentative map of the distribution of *S. fragilissima* at station 26.

All data analysis was conducted in R (R Development Core Team, 2011) and R Studio (RStudio Team, 2015) and with the tidyverse package (Wickham, 2017). Maps were produced with ArcGIS (www.arcgis.com).

4.4 Results

4.4.1 Detection and description of spatial structure

The density of *S. fragilissima* at station 26 varies from 0 to 26.1 individuals per square metre (ind.m^{-2}; Table 4-2). *S. fragilissima* was absent from some images in every transect. The highest densities were observed in t14. Variability in density was high with standard deviation higher than the mean in all transects except t6.

Table 4-2: Specification of each transect surveyed by Autosub6000 at station 26

Transect	***Number of images in transect***	***Transect length (in m)***	***Total Count in transect***	***Mean density in transect (sd)***	***Max. density***	***Minimum autocorrelation distance (m)***
t2	1714	1988	5454	1.787935 (2.08)	15.1	335
t3	1151	1912	3308	1.601815 (1.76)	12.7	302
t6	1314	1844	3811	1.61535 (1.38)	9.19	261
t14	2353	3198	16676	3.907881 (4.00)	26.1	243

The map of *S. fragilissima* density (Figure 4-3) shows that their density can become much higher in some restricted areas. The highest densities seem to appear on local peaks, for example in the southeast sections of transects t2, t3 and t6 or in local troughs, like in the east of t14. Wide but sloped openings like, for example the centre portion of t6, seem less popular.

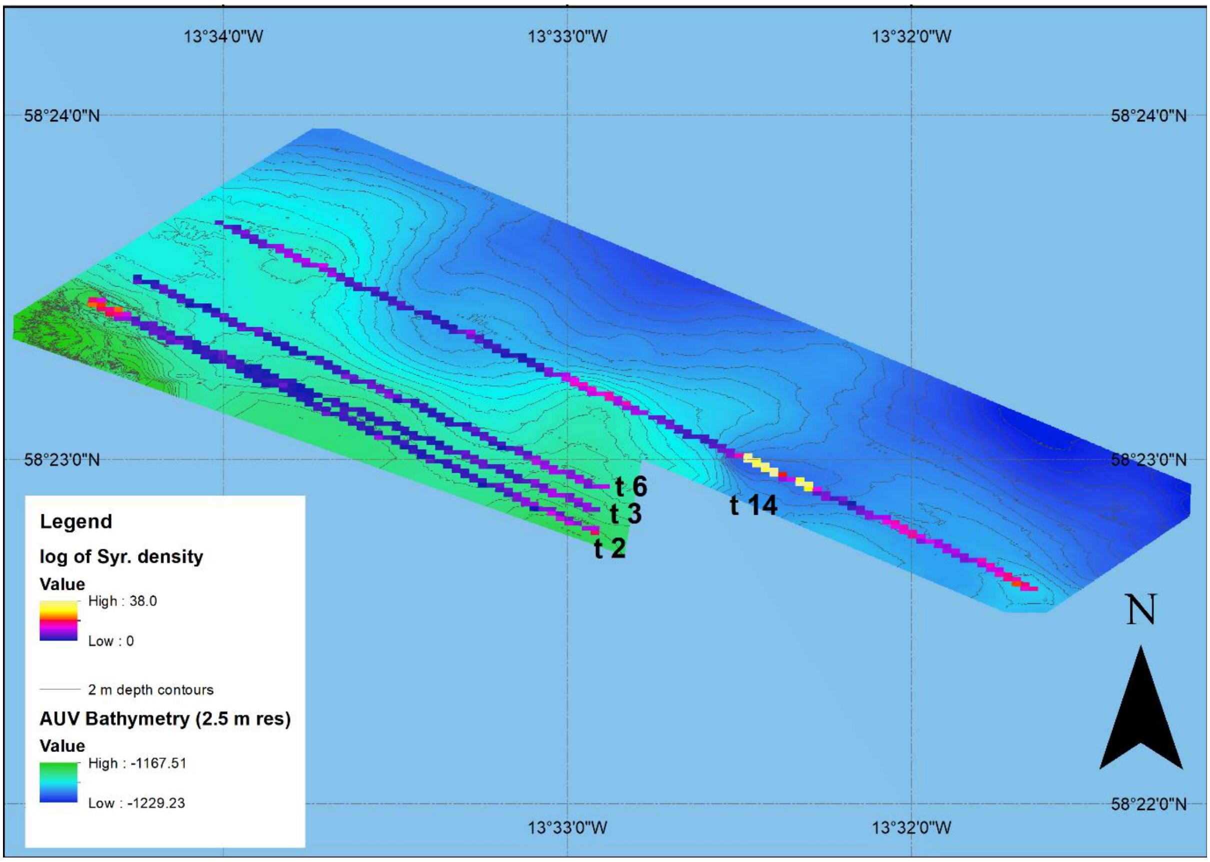


Figure 4-3: Map of the log transformed density of S. fragilissima at station 26. The values were averaged per cells of 25 x 25m for the readability of the map. The yellow areas indicates the highest densities.

Moran's I correlograms show that significant SAC occurs at close distance in each transect (Figure 4-4). This correlation decreases down to 0 or becomes insignificant over a distance ranging from 243 to 335 m (Table 4-2). The influence of SAC is consistent in scale within all 4 transects and does not exceed 335 m in any of them.

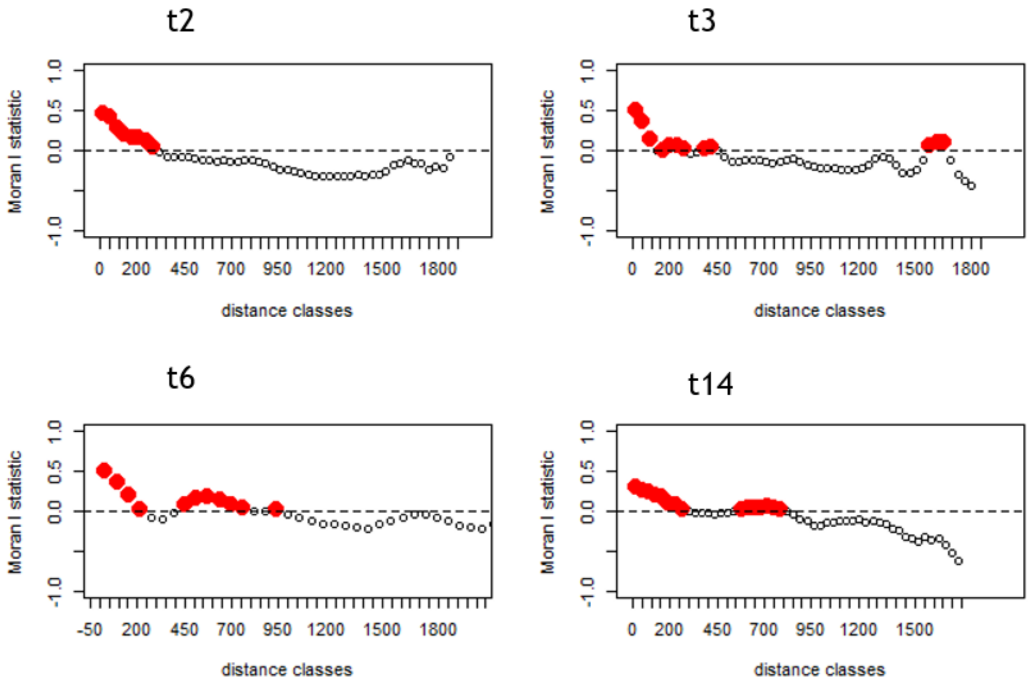


Figure 4-4: Moran's I spatial correlograms of over 100 distance classes. Red dots are distance classes for which the Moran's I was significant.

4.4.2 Modelling and mapping of *S. fragilissima* density

4.4.2.1 Variable importance

The individual relationships between the environmental predictors and the density of *S. fragilissima* does not always offer consistent patterns across all transects (Figure 4-5). For some variables, such as backscatter, a consistent relationship exists, while with others, like the BPI at the finest scales, the relationship is completely different for each of the transects.

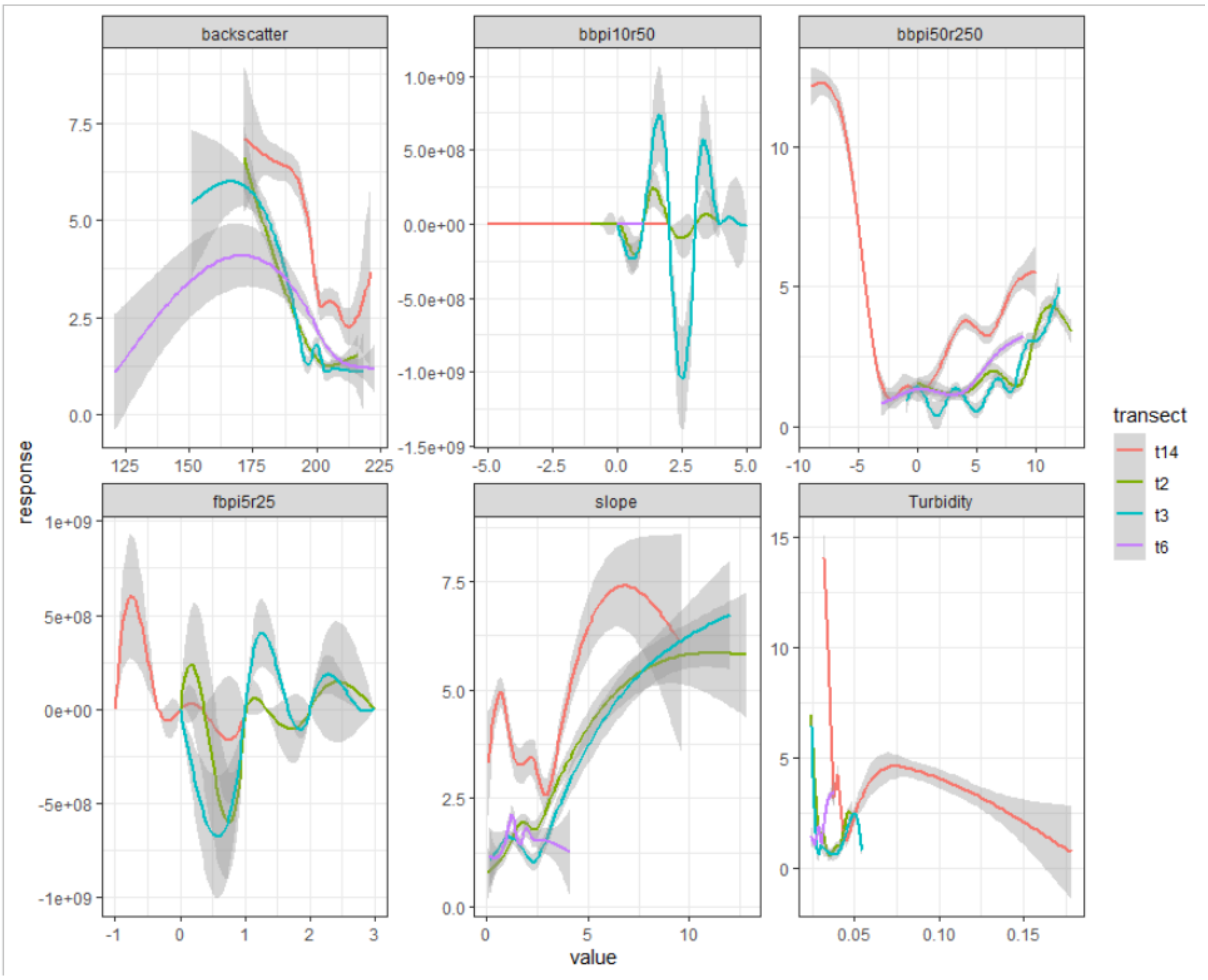


Figure 4-5: Regression plots of relationships between density of S. fragilissima (response) and the environmental predictors (value – different in each plot). The smoothed lines were calculated with gam models. Lines are coloured by transects.

The variable importance plot (Figure 4-6) shows that turbidity, backscatter and the broadest scale BPI (bbpi50r250) were the main drivers on average but important differences exist between the different algorithms in the ensemble. The variable importance values in ensemble models trained on each individual transect varied extensively from one transect to another (Appendix A4-1 and A4-2). However, the BPIs at finest scales (fbpi1r5 and fbpi5r25) are never particularly important while turbidity, backscatter and the broadest scale BPI (bbpi10r50) are always rather important.

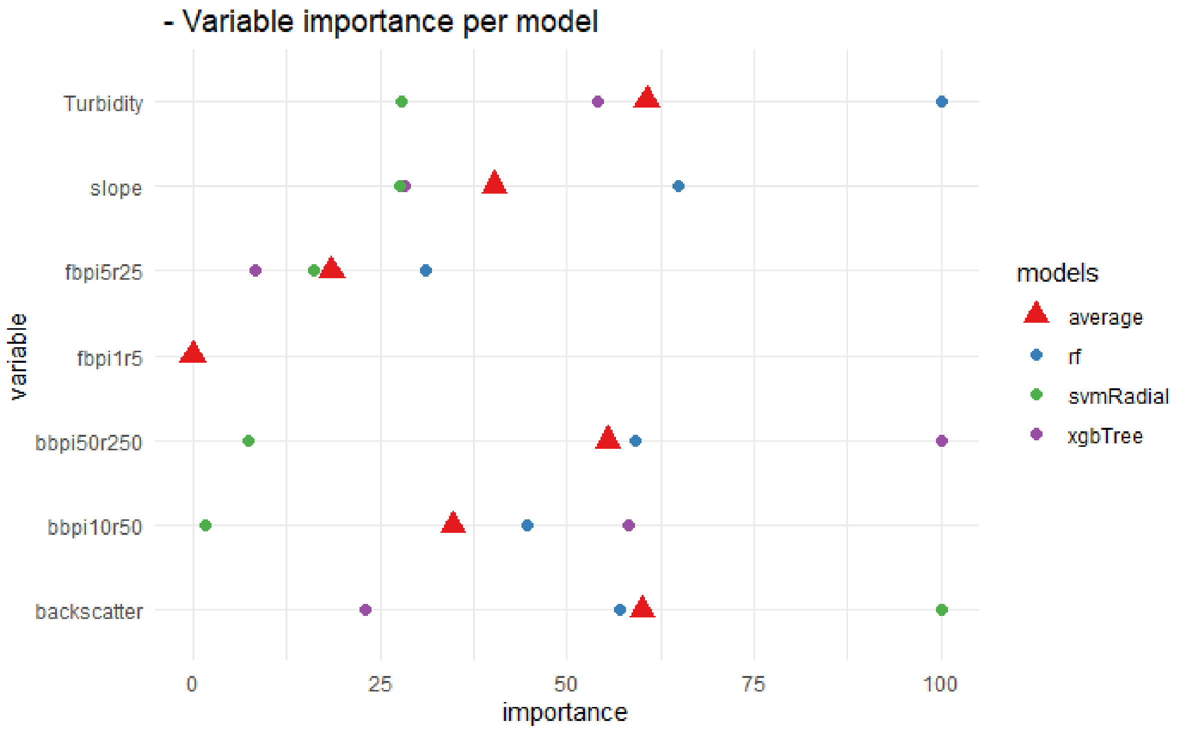


Figure 4-6: Relative variable importance plots of the ensemble model trained on all the data. Points are coloured by algorithms (rf: Random forest, svmRadial: Support Vector Machine, xgbTree: Extreme Gradient Boosting) and their average importance is displayed as red triangles. The higher the score the more important the variable is.

We evaluated the SAC patterns within the residuals of the models (Appendix A4-3) for the model trained on all transects, all the models trained on 3 transects and the models trained on individual transects. Although in some instances, a significant correlation existed in the residuals of the models, the values were always very low and inferior to the SAC and the response variable by an order of magnitude. We thus concluded that the effect of SAC on the model’s error was negligible and that the structure observed at station 26 was driven by the environmental variables measured by the AUV. Figure 4-7 illustrates how in the full dataset, the SAC in the residuals, although present, is negligible compared to the SAC in the actual response variable in the training set.

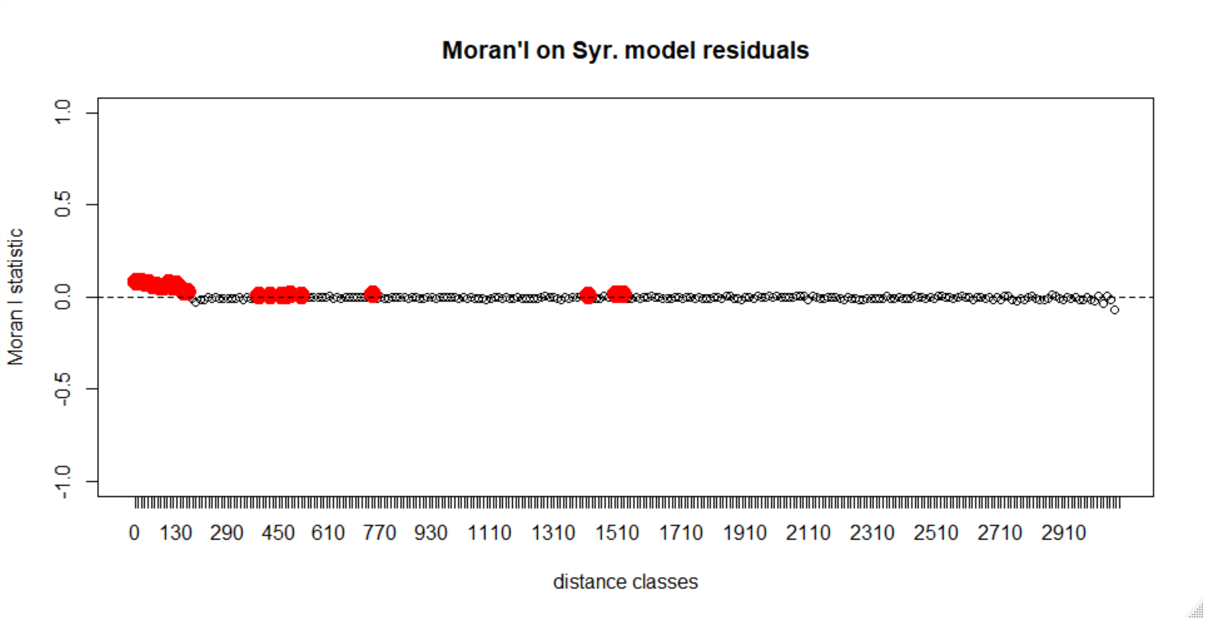


Figure 4-7: Moran's I correlograms on the residuals of the model trained on all the data available. Red dots indicate the distance classes at which there is significant correlation.

4.4.2.2 Mess maps

The mess map of the complete dataset (Figure 4-8) showed that most of the survey area is within the environmental data range (> 0) and only small fraction of the cells were discarded. Maps of all different combinations of testing and training sets are given in appendix A4-4 and A4-5. Some testing sets indeed fell outside the area of suitability of corresponding training sets (like t14 with t2 or t3 as training sets) while others did not (like t2 with t3 as training set).

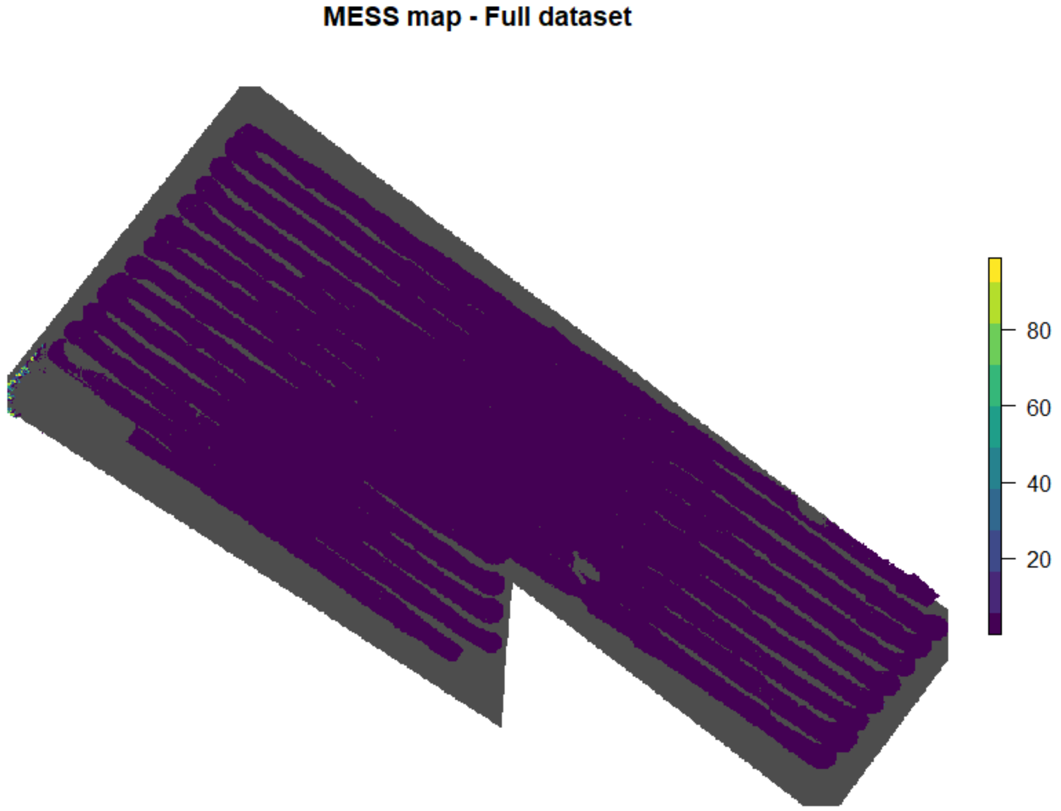


Figure 4-8: MESS map of the full dataset thresholded to only display positive values (where values for predictor variables are within the range of those used in the training dataset). The grey area outlines the extent of the surveyed area seen in Figure 3-1. Most of the cells have positive values retained but a small fraction, particularly near the centre of t14 are removed.

4.4.2.3 Internal Validation

The internal validation of the full model indicated a good correlation between the predicted and observed values (R-squared = 0.69) but the RMSE (1.52) indicated an error of the same order of magnitude as the average density, potentially reflecting the noise in the data. The relatively high RMSE compared to the average density within the full dataset shows that the predictions are potentially far from the observed values in some instances. However, the model is considered valid as it would still identify the areas that are relatively more suitable than their surroundings (Table 4-3). For all models internal validation measures often showed good or very good model performances. On individual transects, the R-squared was never inferior to 0.35 and was up to 0.76. Models trained with the t14 transect included in training tended to show better performances regardless of whether other transects were included or not.

In the results of the internal validation, there is a strong positive correlation (0.52) between Rsquared and the number of datapoints in training indicating that the size of the dataset has a positive influence on model results.

4.4.2.4 External validation

Conversely, external validation showed very poor performances (Table 4-3). No single model was able to predict accurately the density of *S. fragilissima* in transects on which it had not been trained. No model achieved R-squared performances higher than 0.32, regardless of the training set. Error measured by both RMSE and R-squared were always higher in external validation than internal. In 9 cases out of 16, the R-squared in external validation was inferior to 0.05. The only cases where external validation reached an R-squared higher than 0.3 was when training was on t2 and testing was on t3 or vice-versa. Training on all transects except t3, and testing the model on that transect, had an R-squared of 0.29. Unintuitively, training on all transects except t2 and testing on that transect gave an R-squared of 0.06. Interestingly, the models that produced the highest external validation were not necessarily the models that have the highest internal validation.

Based on external validation, there was a weak but positive correlation (0.09) between Rsquared and the number of data-points in training. It is also worth noting that in some instances models trained on more than 2900 data-points performed just as poorly as models trained on 723 data points when predicting on some transects, like t6. Finally, there was no apparent link between the performances and whether or not the majority of points in the testing set are in areas considered similar by the mess map (See Appendix 3-5 for individual training sets MESS maps).

Table 4-3: Internal and external validation performances of models trained on individual transects and tested on another individual transect or trained on all transects but one which is used for testing. The number of datapoints in each training set (after reduction to one point per cell) is also indicated and was used to calculate the correlation between performances and size of training set given at the bottom of the table

Training transects	***Number of datapoints in training***	***Internal***		***Testing transect***	***External***	
		RMSE	**Rsquared**		**RMSE**	**Rsquared**
t3	**723**	1.08	0.57	**t2**	1.590956	0.32
t6	**768**	0.97	0.36	**t2**	2.399424	0.00
t14	**1289**	1.83	0.76	**t2**	3.516431	0.04
t2	**885**	1.28	0.54	**t3**	1.437002	0.31
t6	**768**	0.96	0.37	**t3**	2.035081	0.07
t14	**1289**	1.84	0.76	**t3**	2.939172	0.15
t2	**885**	1.27	0.56	**t6**	2.269925	0.10
t3	**723**	1.06	0.53	**t6**	1.40432	0.01
t14	**1289**	1.83	0.76	**t6**	3.067139	0.01
t2	**885**	1.27	0.56	**t14**	4.220978	0.01
t3	**723**	1.07	0.56	**t14**	4.150438	0.02
t6	**768**	0.97	0.35	**t14**	4.070819	0.03
t3, t6, t14	**2780**	1.49	0.74	**t2**	2.782134	0.06
t2, t6, t14	**2942**	1.59	0.72	**t3**	1.601562	0.3
t2, t3, t14	**2897**	1.63	0.71	**t6**	1.916194	0.04
t2, t3, t6	**2376**	1.18	0.47	**t14**	4.215869	0.05
Correlation between Rsqured and amount of training data			0.52			0.09

Finally projection of the model trained on all data over the entire study area (Figure 4-9) suggests that *S. fragilissima* is distributed in patches of high densities with two aggregations of large size near the North-eastern border of the study area.

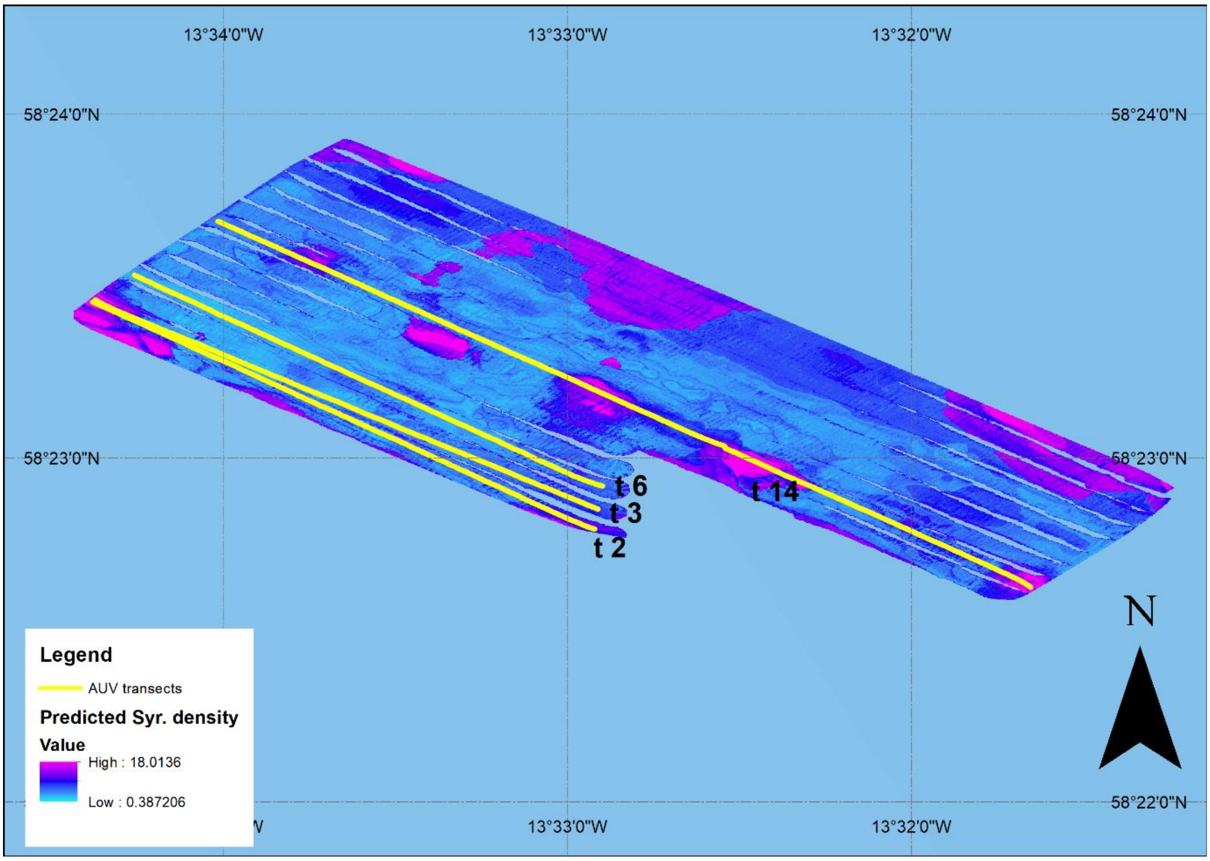


Figure 4-9: Predicted map of S. fragilissima distribution by the ensemble model trained on the full dataset

4.5 Discussion

In this study, we attempted to determine if a large dataset of biological and environmental data, collected by an AUV in a single dive, could provide a useful insight into the fine-scale distribution of deep-sea species, using the listed VME indicator taxon *Syringammina fragilissima* as a model species. We tested if measured density of *S. fragilissima* in four transects within a relatively short distance from each other could robustly describe the spatial structure of their distribution. We then attempted to extrapolate the pattern observed within the transects to the rest of the survey area with species distribution modelling and determine if the large dataset available was sufficient to make a reliable map of the distribution of *S. fragilissima*, as well as study how much influence the environmental drivers have on its distribution at the study site.

The large amount of data available allowed us to thoroughly test the models with internal and external validation and evaluate how various subsets could explain the local distribution of *S. fragilissima*.

4.5.1 Spatial structure of *S. fragilissima* aggregations

We were able to detect a consistent structure in the spatial distribution of *S. fragilissima* Indeed their density was far from homogenous and varied widely in all transects. For example, densities of up to 26 ind.m^{-2} were observed within a relatively short distance (less than 200 m) from areas where their density was close to 0. The SAC patterns indicated that *S. fragilissima* density at a given point is correlated to that of its surroundings within a 240 to 330 m radius and that correlation decreases with distance. This does not seem random as this SAC pattern is relatively consistent across the different transects but it disappears in the residuals of the models, which would indicate that it is mostly the result of variation in environmental predictors.

Although the drivers of xenophyophore distributions are not known with certainty, they seem to vary with the scale like the rest of deep-sea megafauna (Rex and Etter, 2010, Rogers, 2015). At broad scales, Ashford et al. (2014) have identified depth, oxygen availability, nitrate concentration, amount of organic carbon, and temperature to be the most important parameters driving the distribution of xenophyophores. The available parameters in this list, depth and temperature, do not vary enough within the samples collected at station 26 to explain the local distribution of xenophyophores in our model.

At fine scale, our observations are coherent with other observations and predictions of the distribution of *S. fragilissima* aggregations. At very fine scale, they have been observed on flat muddy terrain, such as near the Darwin Mounds (Huvenne et al., 2016a, Bett, 2001a), as well as steep slopes in Canyons (Gooday et al., 2011). At station 26, the shape of the terrain appears to be an important driver of their distribution, as observed in other areas where their close proximity with raised features has been reported at fine and broad scale (Huvenne et al., 2016a, Davies et al., 2015, Ross and Howell, 2013), although not systematically. It has long been hypothesized that the dense yet patchily distributed aggregations of xenophyophores observed in specific locations, like the Western Darwin Mounds, were associated with the scoured tails created by currents behind raised structures (Bett, 2001a). A counter example exists where this association was not observed and the xenophyophores were very widespread (Howell et al., 2014a). At station 26, high densities were found both on local peaks and troughs and it is possible the ecological niche of *S. fragilissima* is defined by exclusion from other areas by other more competitive species, which would explain this seemingly incoherent preference for both troughs and crests.

Backscatter, broadest scale BPI and turbidity were the main drivers of *S. fragilissima* density in most training sets; however, the single most important driver is unclear. This relative inconsistency could be interpreted as a sign that these variables only partially correlate with the actual drivers of *S. fragilissima* distribution and the strength of that correlation may vary locally, resulting in this mixed pattern. More research is needed to confirm or inform this hypothesis, possibly by investigating in details the statistical links between the response variable and each individual predictor.

The predominance of these drivers is also consistent with previously published results. Ross and Howell (2013) and Piechaud et al. (2015), who worked on *Syringammina fragilissima* in the North-East Atlantic (but at a broader scale than this study), also identified depth as the main driver of xenophyophore distribution but, like in our results, they also highlighted the importance of slope and fine scale BPI. Both these studies predicted the presence of *S. fragilissima* in the vicinity of the location of station 26. The density of *S. fragilissima* has already been linked with visible variation in the backscatter on the Darwin Mounds (Huvenne et al., 2016a, Huvenne, 2011). We hypothesized that the higher turbidity could result from localised resuspension of organic matter by the currents or the accumulation of marine snow and, thus, acts as a proxy for the amount of nutrients available to *S. fragilissima.* This parameter and its effect on the local community is likely to vary over time (Davison et al., 2019) and we could not estimate this variability in this study as the AUV did not repeat measurement over time at the same locations. Besides, it may be subject to measurement errors, if, for example, the AUV flies too close or too far from the seabed. Nonetheless, several authors have given credit to the hypothesis that, at both fine and broad scales, *S. fragilissima* is associated with areas of high influx of organic matter, which was itself influenced by topography (Levin and Gooday, 1992b, Gooday et al., 2011). Although we did not quantify the amount of organic matter visible on the seabed in this study, Morris et al. (2016), who used AUV images to perform these measures, observed a high spatial heterogeneity in the distribution of organic matter and observed a large influence of topography on this driver. This could be the reason why different predictors appear as most important in different locations as they all individually or collectively correlate with or even influence the local concentration of organic matter. All these predictors might thus simply be imperfect proxies for the best feeding grounds.

At station 26, *S. fragilissima* abundances above 25 ind.m^{-2} were observed, which are higher than the previously published values for this species in this area. Roberts et al. (2000) and Bett (2001a) reported average densities of 7 to 10 ind.m^{-2}, while we observed more than double that density although with a different imaging platform which is a source of discrepancies (See **chapter 3**). The environmental setting of station 26 could be closer to optimum for *S. fragilissima* than previously studied locations, thus, presenting a clearer image of their ecological niche and helping to further understand their ecology.

Overall, there are less data available for *S. fragilissima* than for other VME indicator taxa, and their exact preferred niche may be very specific to local conditions and driven by predictors that have yet to be clearly identified and could themselves vary between locations. We, however, have observed how heterogeneous the distribution of *S. fragilissima* can be within the Rockall Trough. The large size of our dataset and the number of replications enabled robust measures of the scale of that heterogeneity and enabled us to link it to fine-scale variation in topography and oceanographic predictors. This shows that AUVs can quickly provide valuable data to describe poorly known phenomena and demonstrates the efficiency of these vehicles. It highlights how large datasets can give useful insights to understand ecology and distribution of a VME and inform its conservation.

4.5.2 Mapping and modelling: possible reasons for poor performances

Accurate mapping of the extent and distribution of a VME, such as *S. fragilissima* aggregations, requires more than simply describing variation in their density within a small restricted area. Extrapolating knowledge from localised observations to unexplored areas is essential for their management (Ross and Howell, 2013) but our dataset failed to produce models that could make accurate predictions within even the study area itself when externally validated, despite being considered good models when internally validated.

The various SDM models we trained established that the available environmental variables explained an important part of the variation in *S. fragilissima* density with Rsquared values as high as 0.76. The models performed well according to standard internal validation procedures and these performances seemed to benefit from the large amount of data used in training. However, external validation with relatively independent test data showed that these models were never able to accurately predict the density of *S. fragilissima*, even in transects located very close and sometimes at shorter distance than the minimum SAC distance. In all cases, models trained on three transects failed to accurately predict the density of *S. fragilissima* in the fourth transect. We therefore cannot be confident that the predictions of the general model trained on all transects (shown in Figure 4-9) can be relied upon. This is important to report as models are increasing being used to inform policy decisions. There was however, a positive correlation (0.52) between the number of data points used in training and the R-squared of the models suggesting that models built with more data performed better.

Most deep-sea SDM studies are built on much less data than used in the current study and validated with test datasets that are not independent from the training sets (Vierod et al., 2014). This is a result of the scarcity of data, and expense of collecting additional data. The few studies that have attempted to externally validate their data models with independently acquired data also reported poor performance of the models, contrary to what their internal validation indicated (Anderson et al., 2016, Rooper et al., 2016). The large difference between internal and external validation in this study supports these previous observations and suggests that, although models do benefit from a large amount of data in training, gathering thousands of data points from a restricted area is not sufficient to produce an accurate model.

We can only speculate on the reasons for the dramatic differences between internal and external validation. Models do need to be trained on data encompassing the whole range of environmental conditions in which the target species can live (Barbet-Massin et al., 2010, Yates et al., 2018). It is likely that none of the transects, combinations of 3 transects nor the whole dataset encompass the niche occupied by *S. fragilissima.* Considering the MESS maps for each model trained and tested, it is not clear if the relative novelty or absence of novelty of some testing datasets can be linked to the external validation performances. Indeed, model performances were systematically low regardless of the level of novelty. The models may thus not be able to effectively represent the relationship between *S. fragilissima* density and environmental driver. The instantaneous nature of the environmental data acquired by the AUV provides no understanding of temporal variation, and thus omits an important aspect of the environmental variability.

As knowledge of *S. fragilissima* ecology is still lacking, the true drivers of their distribution are unknown and it is likely that adding other variables, particularly those related to hydrography, would improve the model as Rengstorf et al. (2014) or Mohn et al. (2014) observed on other benthic habitat building species. A better understanding of their ecology in general would help in choosing appropriate predictors and modelling methods (Yates et al., 2018). Furthermore, the distribution of *S. fragilissima* may not be entirely environmentally driven and could be shaped by biotic interactions that can only be observed if the other organisms of the community were considered in the analysis after their presence or absence was recorded in the first place.

There is a large diversity of methods available to predict species distributions (Elith and Graham, 2009). The methods we used here were relatively generic and could be further tuned and optimised. However, the very stark contrast between internal and external validation is unlikely to be overcome with fine-tuning in the model-building framework and the currently available dataset. It is worth noting that Wenger and Olden (2012) observed how complex algorithms tended to produce good internal validation performances but were not as good as linear methods when tested on independent data. The ensemble model could possibly benefit from the inclusion of such methods. More generally, modelling algorithms selected for their quality for internal validation may not be the most appropriate for modelling species distribution in novel environments. Algorithms should be considered not only on the basis of their internal validation performances but on their external validations too. Furthermore, there is growing evidence that complex models and algorithms should be used with care and with ecological justification and more attention should be paid to the potential overfitting of the model to local conditions and how it can decrease its transferability (Bell and Schlaepfer, 2016, Yates et al., 2018, Merow et al., 2014).

Finally, a certain variability exists in the response variable itself even within the spatial autocorrelation distance. Artefacts also exist in the environmental predictors collected by the AUV, which could artificially introduce error in to the model and use of finer resolution multibeam data may produce layers more faithful to the local environment at station 26.

4.5.3 Benefit of large datasets and use of AUVs

This study used data obtained from a single dive of an AUV and the images analysed were gathered in less than 3 hours on the seabed. The size of the dataset available allowed a robust description and quantification of a phenomenon that would have been visible with a smaller dataset (e.g. patchy distribution) but without the certainty offered by the replication. Moreover, the size of the dataset seemed to improve the performances of SDM in internal validation, while allowing a more objective measure of model performance using external validation. Had we had only one of the transects available, as would have been the case if we used a slower sampling gear, like an ROV, external validation would likely not have been conducted.

The near-full coverage of the seabed within the transects has allowed us to precisely quantify SAC, which has rarely been achieved for deep-sea communities. The variability of deep-sea communities at fine scale has previously been observed and documented (Rex and Etter, 2010, McClain and Rex, 2015). Overall, ecologists currently agree that, although SAC is probably high at all spatial scales, its extent and influence on total deep-sea diversity and ecosystem functioning is not known (McClain et al., 2011, McClain and Schlacher, 2015). This is hampering conservation efforts that need precise information on species distribution and abundance (Danovaro et al., 2016). This variability is inherently difficult to study, as are most biological phenomena in the deep sea (Brind'Amour et al., 2009), and spatially limited datasets are probably underestimating it (Morris et al., 2016). In this context, AUVs can efficiently collect enough data to robustly quantify spatial variability and will thus provide a valuable contribution to deep-sea science.

However, the poor performance of our models in predicting the distribution of *S. fragilissima* in areas without data clearly demonstrate that size of the training set alone does not result in an accurate model. This confirms what other modellers and ecologists have warned the community against: the over-reliance on models built on spatially biased, restricted in size or in coverage, datasets may lead to wrong assumptions on spatial distribution, which could lead to misinformed conservation measures (Anderson et al., 2016, Yates et al., 2018).

In our dataset, focusing analysis effort on a reduced number of transects rather than spreading over the whole survey area allowed us to detect the spatial structure of *S. fragilissima*, but it may have hampered our ability to model its distribution. Deep-sea ecology sampling designs are usually a compromise between the suitability of the dataset for a given question while retaining some ubiquity in order to facilitate use in multiple studies (Foster et al., 2014). AUVs offer an advantage in that both good design and wide coverage can be achieved in terms of data collection. The challenge comes in the analysis of that data (Schoening et al., 2017). This will be the focus of **chapter 5** of this thesis.

The lack of generalisability of conclusions drawn from a restricted sample is an issue in the deep sea, where large areas are studied with comparatively very small samples. AUVs will not help address this issue if they are used to gather more samples from the same limited number of stations. However, as they are quick, efficient and cost-effective sampling tools, ecologists should not only aim at increasing the amount of data collected at each station but also cover a wider range of environmental gradients, as well as investigate larger spatial and temporal scales. In short, survey more stations. We therefore consider that there is benefit in acquiring large amounts of data but we also urge scientists to not be deceived by a false sense of statistical robustness it appears to offer. It is important to remember that spatially aggregated samples do not offer generalisable insights nor are they true replicates.

4.5.4 Future work

The increased availability of large datasets will open several interesting research opportunities on both short and longer terms. In the immediate future, annotating more data to sample a wider range of environmental conditions could contribute to a greater understanding of the environmental drivers of species distributions, and as a result, better accuracy of externally validated models. But more efficient annotation methods are desirable to take advantage of the AUVs sampling capacity. These new samples can be small groups of images spread across the whole sample area rather than entire transects.

At station 26, this would enable a more robust study of the interaction between the spatial distribution of *S. fragilissima* and other taxa and the local environmental variables. A more in detailed investigation of SAC patterns could also be implemented with other metrics, particularly indices of Local Indicators of Spatial Association (LISA) rather than whole transects (Anselin, 1995, Barrell and Grant, 2013), as it would enable a better visualisation of the variability of the SAC and help better understand the poor model performances.

In the longer term, more consideration is required on the best sampling designs to be used for AUVs, whose greater freedom of movement offers much more flexibility and the possibility to optimise the sampling, increase coverage, avoid biases in the datasets and account for the now known distribution of the target species. Other AUV users have also stressed the need for these improvements (Foster et al., 2014, Foster et al., 2017, Foster et al., 2018, Woodall et al., 2018).

Sampling designs should account for SAC in the spacing between the transects as well as their length so that some data points can be guaranteed independent (Brind’Amour et al., 2018, Brind’Amour et al., 2009). It is worth considering that transects are an appropriate way to use gears like trawls and ROVs that have reduced mobility and manoeuvrability but, as AUVs are less constrained in their movement, they can implement more sophisticated and better spatially optimised designs (Foster et al., 2014).

4.6 Conclusion

Our results have shown that *S. fragilissima* is patchily distributed in patches around 300 m across and, in some places can reach densities of up to 26 ind.m^{-2}. The observed structure could be explained by environmental parameters, mostly turbidity, broad-scale BPI and backscatter. We found that that model performances were positively correlated with the number of data points used in training but the external validation poor performances (no higher Rsquared than 0.33) showed that the exact relationship between predictors and density of *S. fragilissima* was not constant across transects. We concluded that the model built on that dataset, although statistically valid judging by the high internal validation scores (up to 0.76) should not be generalised to other areas.

This showed that dense sampling of small areas is no substitute for good survey design. This work demonstrated the potential of AUVs. These vehicles do offer significant advantages to the future study of deep-sea benthic ecosystems at fine scale and specifically SDM modelling through enabling more robust testing of relationships, and in providing larger datasets for modelling studies.

4.7 Acknowledgments

The authors acknowledge the officers and crew of RV ‘James Cook’ as well as the Autosub support team who assisted in the collection of the data. The DeepLinks project is funded by NERC NE/K011855/1. We thank Kirsty Morris, Erik Simon Lledo and Marcus Shirley for their advice on processing the images prior to analysis. Help from the developers of Biigle and MAIA, Daniel Langenkämper and Martin Zurowietz was equally instrumental and appreciated. We also thank Brian Bett for his help and code for the image size calculation.

Chapter 5: Automated identification of benthic epifauna with computer vision

This chapter was published as:

Piechaud, Nils, Christopher Hunt, Phil F. Culverhouse, Nicola L. Foster, and Kerry L. Howell. (2019) "Automated identification of benthic epifauna with computer vision." *Marine Ecology Progress Series* 615 (2019): 15-30. https://www.int-res.com/abstracts/meps/v615/p15-30/

The full text is attached to this thesis

5.1 Abstract

Benthic ecosystems are chronically undersampled, particularly in environments >50 m depth. Yet a rising level of anthropogenic threats makes data collection ever more urgent. Currently, modern underwater sampling tools, particularly autonomous underwater vehicles (AUVs), are able to collect vast image datasets, but cannot bypass the bottleneck formed by manual image annotation. Computer vision (CV) offers a faster, more consistent, cost effective and sharable alternative to manual annotation. We used TensorFlow to evaluate the performance of the Inception V3 model with different numbers of training images, as well as assessing how many different classes (taxa) it could distinguish. Classifiers (models) were trained with increasing amounts of data (20 to 1000 images of each taxa) and increasing numbers of taxa (7 to 52). Maximum performance (0.78 sensitivity, 0.75 precision) was achieved using the maximum number of training images but little was gained in performance beyond 200 training images. Performance was also highest with the least classes in training. None of the classifiers had average performances high enough to be a suitable alternative to manual annotation. However, some classifiers performed well for individual taxa (0.95 sensitivity, 0.94 precision). Our results suggest this technology is currently best applied to specific taxa that can be reliably identified and where 200 training images offers a good compromise between performance and annotation effort. This demonstrates that CV could be routinely employed as a tool to study benthic ecology by non-specialists, which could lead to a major increase in data availability for conservation research and biodiversity management.

5.2 Introduction

Marine ecosystems cover the majority of Earth's surface but benthic ecologists and biodiversity mangers have long been confronted with a shortage of data (Borja et al., 2016, Jongman, 2013) regarding its composition and functioning. With increasing anthropogenic pressure, management measures need to be implemented urgently (Danovaro et al., 2017a, Van Dover et al., 2014). These conservation measures must be based on a solid understanding of taxonomic diversity and ecological dynamics of habitats considered (Hernandez et al., 2006). In many cases, that knowledge is lacking and specialists agree that data collection must be increased to tackle the challenge (Borja et al., 2016, Costello et al., 2010). The amount of data currently available on benthic ecosystems is always limited by how many samples can be collected, stored, and processed at a time. Since the 19th century, various technological innovations have attempted to bypass this bottleneck.

Benthic ecosystems are traditionally sampled by trawls, cores and other physical means. These physical samples are costly to collect and process, and logistically challenging to store (Clark et al., 2016b). While physical samples remain the mainstay of benthic surveys, use of underwater imaging technologies is increasingly popular among marine ecologists (Bicknell et al., 2016, Brandt et al., 2016, Romero-Ramirez et al., 2016, Solan et al., 2003). These technologies offer a less invasive, more cost effective method of survey, and storage space for image data is virtually unlimited (Mallet and Pelletier, 2014). Underwater imaging is now regularly utilised alongside other sampling tools to provide a comprehensive view of the marine environment.

Modern underwater sampling vehicles, and particularly Autonomous Underwater Vehicles (AUV), have great potential in providing the step-change in the rate of data gathering that is needed to support sustainable marine environmental management. They are capable of collecting large numbers of images of the sea bed in a single deployment (Lucieer and Forrest, 2016, Williams et al., 2016). For example, a 22 hour AUV dive can deliver more than 150,000 images of the seafloor along with other types of environmental data (Wynn et al., 2012). Comparatively, trawls and Remotely Operated Vehicles (ROV) cover less ground per dive and the ship and its crew are unable to operate any other benthic equipment while they are deployed (Clark et al., 2016b, Brandt et al., 2016).

To translate the information contained in images into semantic data that can then be used in statistical analysis, a step of manual analysis (or annotation) is conducted by trained scientists. Human observers, even highly-trained, do not achieve 100% correct classification rates and are highly inconsistent across time and across annotators (Culverhouse et al., 2003, Durden et al., 2016a, Beijbom et al., 2015, Culverhouse et al., 2014). Besides, manual image annotation results are subject to observer bias, meaning interpretations vary depending on the annotator’s experience and their mood changes across the analysis process (tiredness, boredom or stress, etc…) (Durden et al., 2016a, Culverhouse et al., 2003). The results (format, taxonomic resolution and nomenclature) of these analyses also tend to differ from one institution, project or individual annotator to another. This lack of standardisation makes merging and comparing datasets difficult (Althaus et al., 2015, McClain and Rex, 2015, Bullimore et al., 2013), and the data quality is not always consistent. More importantly, manual analysis is a time-consuming process, which forms the current bottleneck in image-based marine ecological sampling (Schoening et al., 2017, Beijbom et al., 2015, Edgington et al., 2006). The growing trend towards use of AUVs for seafloor biological surveys will only increase the scientific challenge.

Artificial intelligence (AI) and computer vision (CV) provide potential means by which to both accelerate and standardise the interpretation of image data (Beijbom et al., 2012, Culverhouse et al., 2003, Favret and Sieracki, 2016, MacLeod et al., 2010). Although using AI for biological research has a long history (Rohlf and Sokal, 1967, Jeffries et al., 1984, Gaston and O'Neill, 2004), it has always been challenging to implement for non-specialists and requires skills and materials that most biologists do not have access to (Rampasek and Goldenberg, 2016, Gaston and O'Neill, 2004).

CV has been successfully applied to benthic species identification by a growing number of studies (Beijbom et al., 2015, Marburg and Bigham, 2016, Manderson et al., 2017, Edgington et al., 2006, Schneider et al., 2018, Norouzzadeh et al., 2018, Marini et al., 2018b) but has yet to be made into an easy to use tool that any biologist in the field can implement as an alternative to manual image annotation and integrate with previous analysis. Multiple potential commercial applications, the availability of new tools as open software, as well as the improvement of hardware capacity are driving new developments in AI (e.g. neural networks and deep learning). This is likely to change how CV can be employed in the field of scientific research (Weinstein, 2018, Rampasek and Goldenberg, 2016). In parallel, new image analysis and data science software allow easier and more efficient integration of various tools into the research process, from data collection to final scientific or public outreach material (Gomes-Pereira et al., 2016). These new technologies are potentially enabling full automation of the annotation process and could revolutionise ecological research (Weinstein, 2018).

While the principle of automated classification (automated assignation of pre-established classes to objects on images) has been validated, few practical examples exist of AI-based methods used to identify benthic organisms from images acquired by AUV. Consequently, implementing an automated species classifier is a potentially time-consuming investment for an uncertain return. Relying on proven manual methods remains the safe option for researchers. Practical guidance is needed to help ecologists decide whether adopting AI and CV is feasible and would fit their dataset and scientific objectives.

To make that decision, benthic ecologists need to know:

- What level of accuracy and uncertainty can be expected from CV annotation and does it match or approximate the accuracy of human annotators.
- How much material is needed to train a classifier and is a limited amount obtained from a single study sufficient.
- How to assess their own dataset to decide whether use of CV is appropriate.

In this study, we investigate these issues by using an open access algorithm to build a Convolutional Neural Network (CNN) to identify benthic organisms in seafloor images, obtained from a single deployment of the UK's Autosub6000 AUV. Technically speaking, we seek to train an automated classifier that is able to determine which taxa an animal on an image most likely belongs to, using a list of pre-defined taxa (or classes). Specifically, we ask, 1) what impact does the number of images, on which the classifier is trained, have on its performance? and 2) What impact does the number of classes, on which the classifier is trained, have on its performance? In addition, we provide a case study in the application of CV to an unbalanced ecological dataset.

5.3 Method

5.3.1 Study area and data collection:

All the images used in this study were collected by the UK's national AUV Autosub6000 in May 2016 as part of the NERC funded DeepLinks (JC136) research cruise. The images were taken as part of an 1880 m long transect at station 26 of that cruise at 1200 meters depth on the north-east side of Rockall Bank, N.E. Atlantic. This region was selected for the study due to the flat topography and low likelihood of disturbance, making it ideal for AUV deployment. The AUV was equipped with a downward facing Grasshopper2 GS2-GE-50S5C camera from Point Grey Research. The AUV was flown at 1.1ms^{-1} speed, at 3m ±0.1 m off bottom and took images every second, resulting in near overlapping image coverage. The surface area of each image is between 1 and 2.5 m^{2}, and the resolution is 2448 x 2048 pixels.

In total, 1165 raw photos of the seabed were manually annotated by a single observer with the Biigle 2.0 software (Langenkämper et al., 2017) using a regional catalogue of Operational Taxonomical Units (OTU) developed by Howell and Davies (2016). Within the Biigle 2.0 software, location (X and Y coordinates in pixels within the photo for point annotations, or X, Y and radius for individuals marked using a circle) and identity of individual OTUs annotated within each image were recorded and stored.

For each OTU, all individual annotations were visually inspected using the “Largo” evaluation tool in Biigle 2.0, to maximize consistency in identification and reduce error. Later, an assessment of 75% (around 28000) of the annotations in the final dataset used in the model found 41 identification errors. By that assessment, we concluded that the accuracy of identification was above 99%.

5.3.2 Image data

Manual image annotation resulted in a dataset consisting of 41208 individuals belonging to 148 OTUs. Each individual was then cropped from the raw image, together with its assigned OTU label, using a custom Python (www.Python.org) script. For each annotation, a square of 240 pixels or more, positioned manually on X and Y coordinates of the centre of the animal, was fitted and cropped out. For organisms bigger than 40 pixels, the size of the square was manually set to encompass the whole individual. These cropped image slices and associated OTU labels (to become classes in the model training design) formed the input used in the CNN.

5.3.3 Tensorflow and transfer learning

Rather than train our own neural network, we used transfer learning (Pan and Yang, 2010) to retrain the Inception V3 model (Szegedy et al., 2016), a CNN built in the freely available library Tensorflow (Abadi et al., 2016).

CNNs are a particular architecture of neural networks, more specifically, deep neural networks, particularly suited to image analysis (Krizhevsky et al., 2012, LeCun et al., 2015). A CNN has the capacity to detect and match patterns in images thereby “learning” what features are relevant to differentiate objects and, subsequently, classify them accordingly.

Tensorflow (TF) is a C++ based library but has a Python Application Programming Interface (API) that makes it easier to train, tune and deploy neural networks. Transfer learning is a method allowing a CNN built on a large dataset to be repurposed into a classifier capable of distinguishing between classes it was not initially trained on. The strength of this method is that the dataset on which it is transferred does not need to be as large as it should be to train a CNN from the beginning. Here, we were able to train a classifier with tens to hundreds of images per class (in our case, OTUs) instead of millions.

5.3.4 Classifier training and testing

A random 75-25% split was applied to every OTU in order to separate images used for training the classifier and those used for testing. The training and test data sets for all OTUs were then combined into single ‘training’ and ‘test’ datasets.

The OTUs the classifier was trained to identify are referred to as classes and only those OTUs for which there were a sufficient number of image slices (individual observations) available were selected for use in training. The minimum number of images needed for training was set to 20. This means that for an OTU to be included in the study at least 27 image slices were needed, 20 for training and 7 for testing. Out of the 148 OTUs observed, 52 were above that threshold. The remaining 96 OTUs represented 3.19% of the total number of individual annotations and were removed from the dataset.

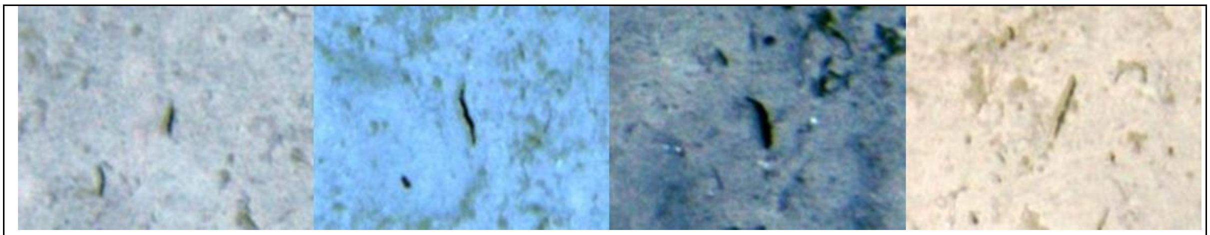
The classifier was trained on the training dataset and then predictions were made on the test dataset. For each cropped image slice in the test dataset, TF gave a score for each of the possible OTU classes for which it had been trained. The scores range from 0 to 1 (the sum of scores for all classes being 1) and represent the model’s confidence that the slice belongs to the corresponding class. The final prediction was the OTU class that received the highest score. The prediction was then compared to the manually assigned OTU class.

To measure the effect of the number of training images (or limit) on the accuracy and confidence of the predictions, the training data set was filtered so each OTU class was represented by 20, 50, 100, 200, 500, and 1000 images (Table 5-1)

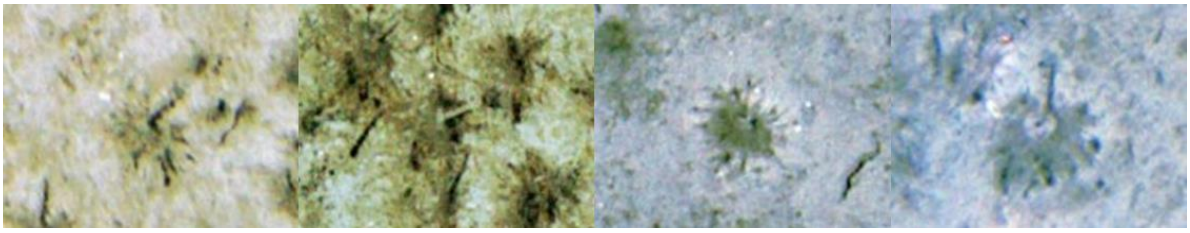
Table 5-1: Nomenclature of classifiers names and characteristics. The different classifiers names are a combination of group name and image numbers per Operational Taxonomical Units (OTU) in training. Groups are defined by the number of different OTUs (or classes) in the training set. In the different groups, the OTUs used are those for which the minimum number of images indicated are available. Within each group, treatments refer to the number *of images of each class in training. The same treatments (20, 50, 100, 200, 500 and 1000 images per OTU in training) were applied to each group but only the classifiers names* ***in bold*** *are balanced (equal number of images for every class). In unbalanced designs, the maximum number of available images is used and is therefore different for each OTU.*

	Groups		
	A	B	C
Number of classes	7	27	52
Minimum number of images available for the OTU to be in the group	1000	100	20
Classifiers names in group ***(balanced classifiers in bold)***	**A20, A50, A100, A200, A500, A1000**	**B20, B50, B100**, B200, B500, B1000	**C20**, C50, C100, C200, C500, C1000

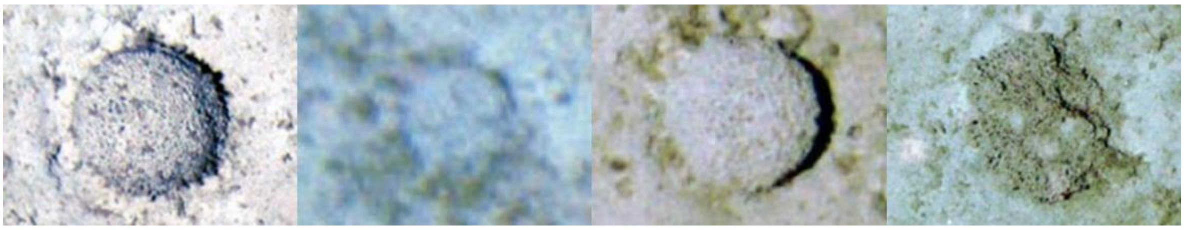
A classifier was then trained on each of these six pools of images and tested using the test data set. Only seven OTUs were observed frequently enough to be used with these six limits (Figure 5-1).



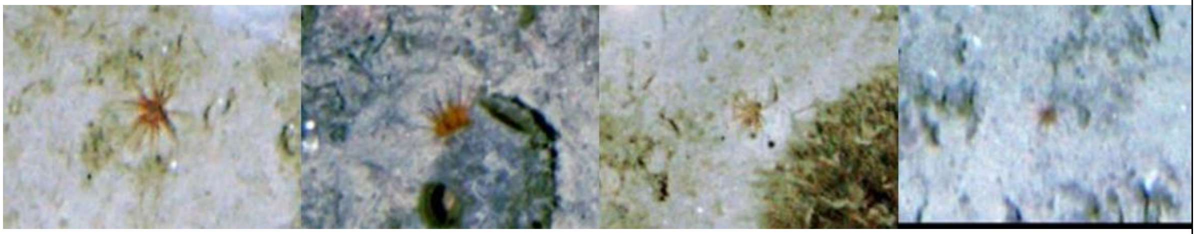
OTU603: Very small elongated sponge. Shape is constant.



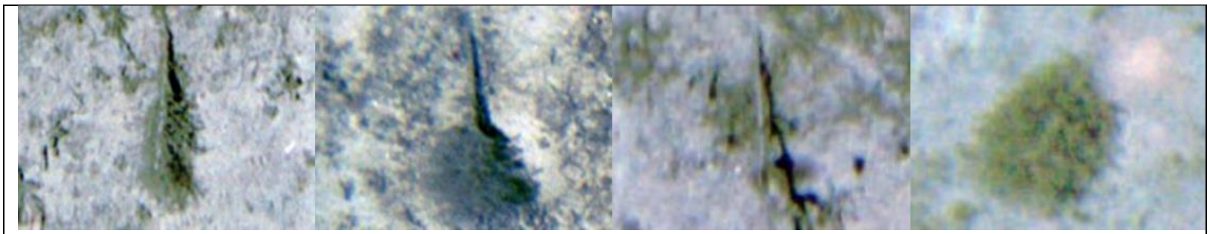
OTU375: Small tube worm. The gills can hide the tube



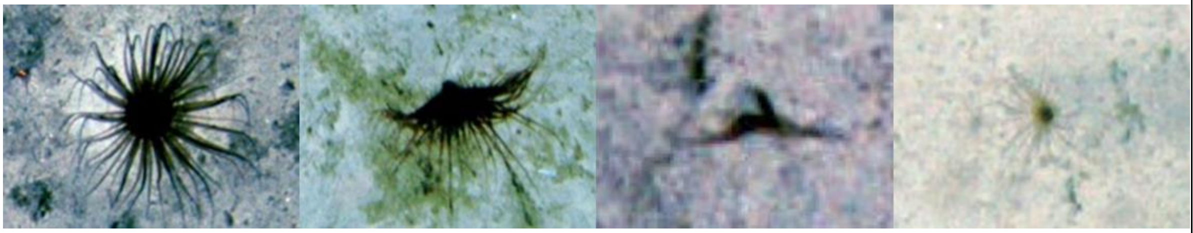
OTU261: The xenophyophore *Syringammina fragilissima*



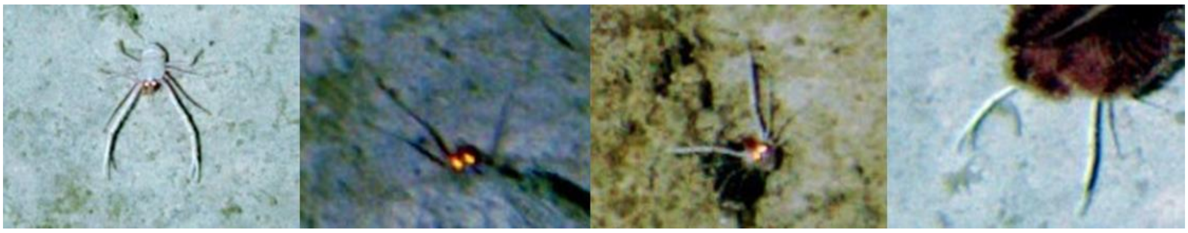
OTU23: Small halcampid/edwardsiid anemone



OTU995: Unknown animal, possibly a Chrysogorgiidae



OTU2: Cerianthid anemone of various size



OTU339: The squat lobster *Munida sarsi/tenuimana*

Figure 5-1 Example images and description of OTUs abundant enough to be in group A. Scale varies. OTUs are ordered by abundance in the original dataset.

The combination of groups and limits is referred to as treatments and designation of each treatment follows the nomenclature in Table 5-1 (e.g. A1000 is group A, limit 1000). Each treatment was repeated 10 times with different random splits between testing and training data for cross-validation.

To measure the effect of the number of OTU classes used to train the CNN on its capacity to correctly classify the test dataset, we used three training datasets each with different numbers of classes (referred to as groups) (Table 5-1). The number of classes is defined by the number of available images per OTU so classifiers can be trained on a set number of images for every class while retaining enough images for testing. Group A contained 7 classes for which more than 1000 images was available; group B contained 27 classes for which more than 100 images were available; and group C contained 52 classes for which more than 20 images were available. Within each group, classifiers were trained with all six pools of images (Table 5-1).

Note that when the limit is above the available number of images, the classes with less images were trained with the maximum number available regardless of the limit. This results in class imbalance in the model training for some treatments in group C with more than 20 images and in B with more than 100 images (balanced treatments are listed in Table 5-1). To assess the effect of the number of OTU classes used to train the CNN on its capacity to correctly classify the test dataset, only balanced designs were used.

In total, 180 (3x6x10) classifiers were trained and tested. All the CNNs were trained in the Google Cloud ML (https://cloud.google.com/) remote computing facility.

To be applied to a “real-life” ecological study, the classifiers have to maximize performances while minimizing the initial effort needed to build the training dataset. To assess appropriate use of CV on a ‘real-life’ dataset we considered all possible combinations of numbers of training image and numbers of OTU classes in an unbalanced design. Average performances and individual OTU performances were assessed.

5.3.5 Analysis and performances evaluation

Considering each class, the observation can be a presence (the OTU is present on the image) or an absence (the OTU is not on the image and another OTU is). The different possible outcomes or predictions of the classifier are detailed in Table 5-2. The respective number of each outcome type (the confusion matrix) was used to calculate performance metrics.

Table 5-2: Possible outcomes of the classifiers. It indicates how the classifiers predictions compare to the manual annotation (the labels) and if it identifies the Operational Taxonomical Unit (OTU) present on an image correctly.

Outcome	***Description***
True Positives	Label is OTU and class predicted is OTU ► Classifier correctly identified the OTU
True Negatives	Label is not OTU and class predicted is not OTU ► Classifier correctly recognized the OTU is not in the image
False Negatives	Label is OTU but class predicted is not OTU ► Classifier misidentified the OTU
False Positives	Label is not OTU but class predicted is OTU ► Classifier misidentified another OTU

The classification accuracy is the percentage of predictions that are correct (prediction matches observation) and is often used to evaluate performances in ML studies. This measure ignores the differences between classes, thus we used two model evaluation metrics which rely on a confusion matrix (Manel et al., 2001) explained in Table 5-2.

- **Sensitivity,** also referred to as *true positives rate* or *recall*. It varies between 0 and 1. It quantifies the proportion of individuals of a given OTU in the testing set that are correctly identified. A value of 1 means that all individuals of a given OTU are identified as such.

$$Sensitivity = \frac{True\ Positives}{True\ Positives + False\ Negatives}$$

- **Precision,** *or Positive Predictive Value*. It varies between 0 and 1. It quantifies the proportion of true positives among the individual identified as a given OTU. A value of 1 means all the individuals identified as a given OTU class are indeed that OTU.

$$Precision = \frac{True\ Positives}{True\ Positives + False\ Positives}$$

Average and standard deviation for all metrics were calculated for each class within each treatment and then averaged over other grouping factors. This gave an estimation of the overall performance of the classifiers. The performances of the classifiers for each individual class were also carefully analysed.

Differences in metrics were statistically tested with a permutation-based analysis of variance in the “lmPerm” package in R (Wheeler and Torchiano, 2010). We report p-values classified with five levels of significance: more than 0.05 or non-significant, less than 0.05, less than 0.01, less than 0.001 and less than 0.0001. Relationships between number of images and performance were extrapolated with a neural network regression in the “nnet” package in R (Ripley et al., 2016) projected over 1000 to 10000 images. All data analyses were carried out in R (Team, 2014) using the “tidyverse” package (Wickham, 2017).

5.4 Results

The results are presented in three sections. First, questions related to the impact of the number of training images are addressed, then the effect of the number of classes in the training set is assessed, and finally the results relevant to choosing the best method in our case study are presented.

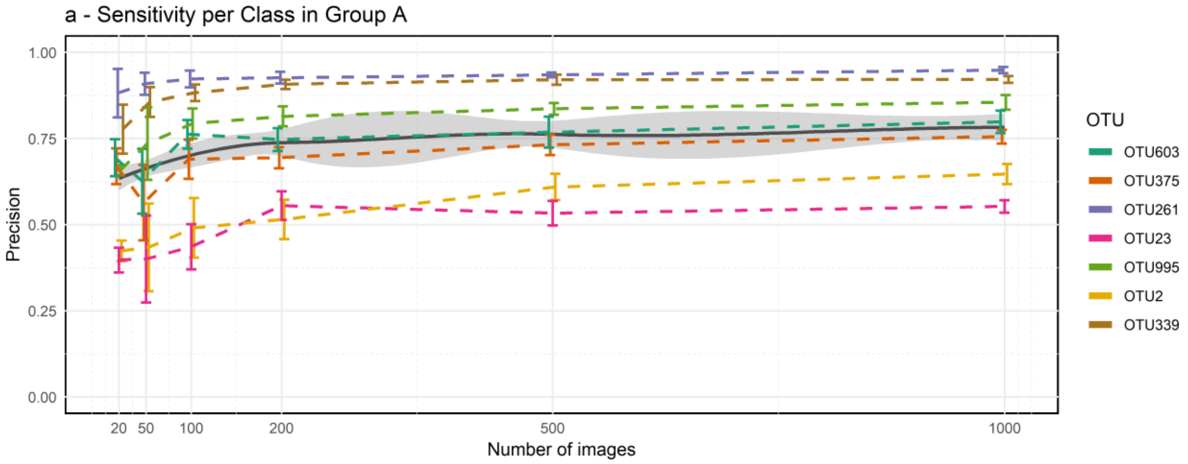
5.4.1 Impact of the number of training images on performance

Average performance, measured as both sensitivity and precision, increases with an increasing number of training images used (Figure 5-2). For sensitivity, there is an average increase from 0.64 to 0.78 when moving from 20 to 1000 images, respectively. This is mirrored by increases in precision from 0.63 to 0.75 when moving from 20 to 1000 images, respectively. Non-linear extrapolations of average sensitivity and precision show that performances reached with 1000 training images may be close to an asymptote and performances obtained with additional training material probably plateau below 0.78 for sensitivity and 0.75 for precision (Figure 5-2). This suggests that the model is unable to achieve perfect performance regardless of how many additional images are used in training.

Figure 5-2: Classifier performances (sensitivity and precision) per number of training images measured (20 – 1000) and extrapolated (1000 – 10000). Grey dots show averaged values across all OTUs for each classifier.

The number of training images has a clear positive effect on performance. For almost all pairs of models compared (Figure A5-1), performance values are statistically significantly different (p <0.05) and very often, significance is very high (p-value <0.0001). There are a few exceptions, like between the A20 and A50 classifiers where the p-value >0.05 for sensitivity and between 0.01-0.05 for precision, or the B1000 classifier, for which there is no significant difference in sensitivity between this classifier and the B500 and B200 classifiers. However, measured difference in performance between sequential models becomes vanishingly small at higher numbers of training images, such that the difference between A200 and A1000 classifiers is 0.04 for sensitivity and 0.05 for precision. This suggests little to no improvement is gained in model performance by using more than 200 training images.

There are strong between-OTU differences in classifier performance (Figure 5-3). All classifiers have high sensitivity for OTU261 and OTU339, even the A20 classifier (0.88 and 0.77, respectively). For OTU2 and OTU23, classifiers have more variable and lower sensitivity regardless of the number of training images used.



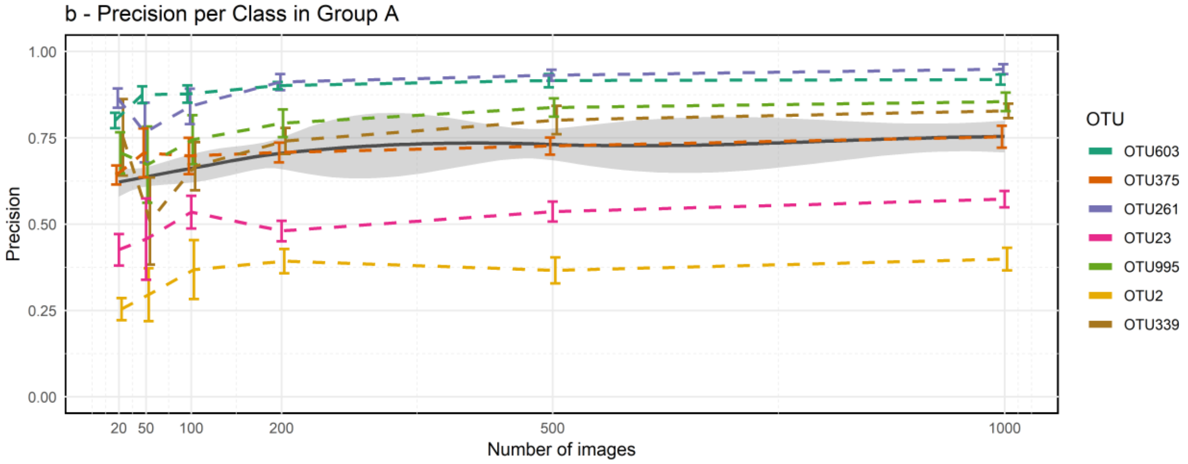


Figure 5-3: a) Evolution of Sensitivity in Group A classifier trained with an increasing number of images. b) Differences in Precision in Group A classifier trained with an increasing number of images. The black line is 'loess' smoothed curve of the average of all the classes and greyed area is a t-based approximation of the standard error.

The OTUs for which precision is highest are not necessarily those for which sensitivity is highest. The highest precision observed was for OTU261 but the second highest precision observed was for OTU603, which has a lower sensitivity. For some classes (OTU261 or OTU339), precision is lower with 50 training images compared to 20 training images.

5.4.2 Impact of the number of classes on classifier performance

Classifiers trained with 7 classes (group A) had significantly better sensitivity (Figure 5-A1) and precision than equivalent classifiers trained on more classes but the same number of images (Figure 5-4). Variability in performance was also lower for classifiers trained with fewer classes. Average sensitivity decreased from 0.71 to 0.38, and average precision decreased from 0.69 to 0.32, when moving from 7 to 27 classes. This suggests a negative effect of the number of classes on performance; however, on average, there is only a minor drop in performance (0.018 in sensitivity and 0.035 in precision) between classifiers trained on 27 and 52 classes. Interestingly, B100 and C100 both have sensitivity of 0.38 (no statistical difference) and C20 has higher (+ 0.02) sensitivity than B20.

OTUs that perform well in one group tend to perform well in other groups. OTU261 and OTU339 are in the top 10 for each group although their performances are lower in groups B and C.

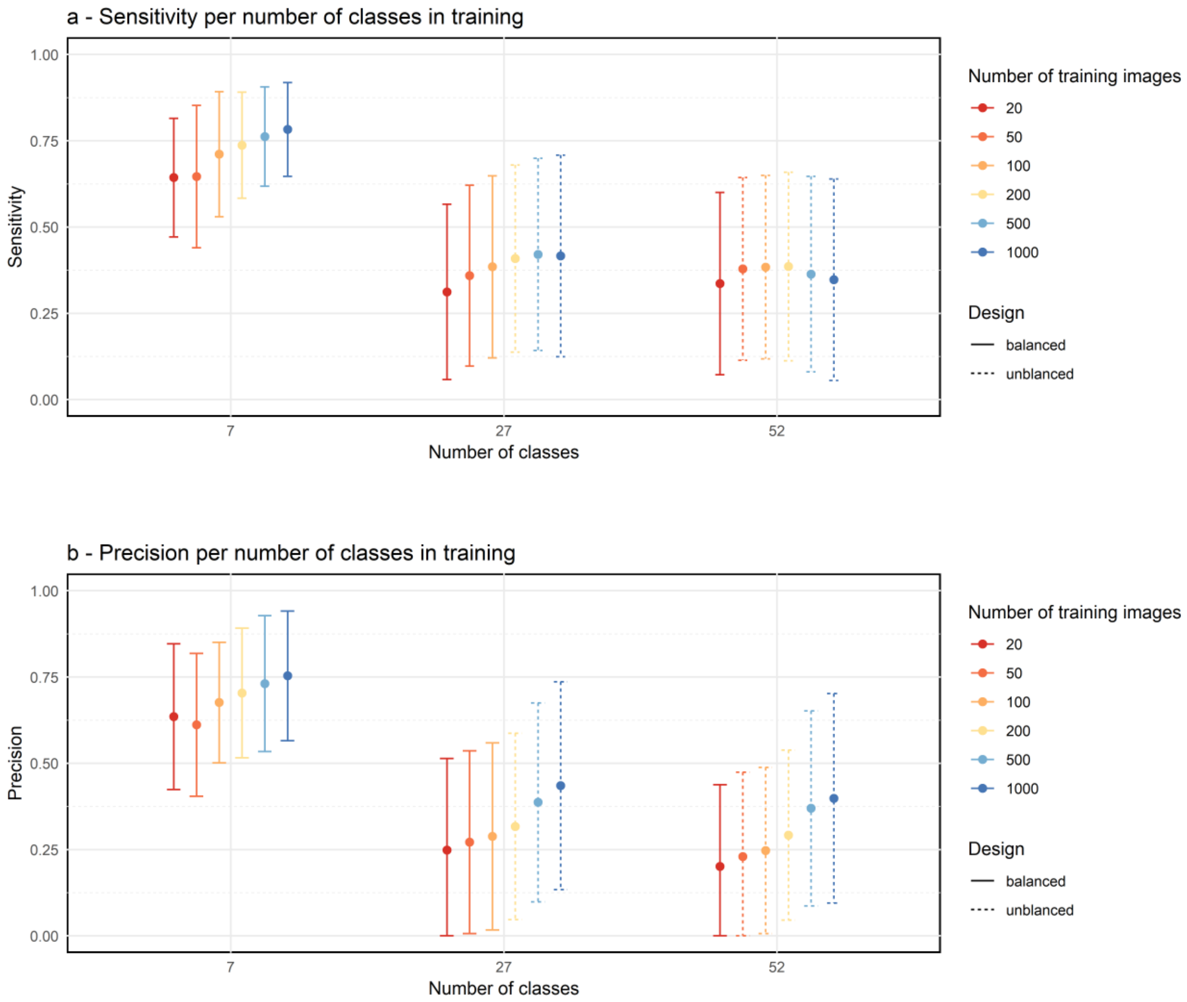


Figure 5-4: a) Differences in sensitivity in classifiers trained with different number of classes and images (7 for group A, 27 for group B and 52 for group C). b) Differences in precision in classifier trained with different number of classes and images. Error bars are standard deviation of the 10 random splits.

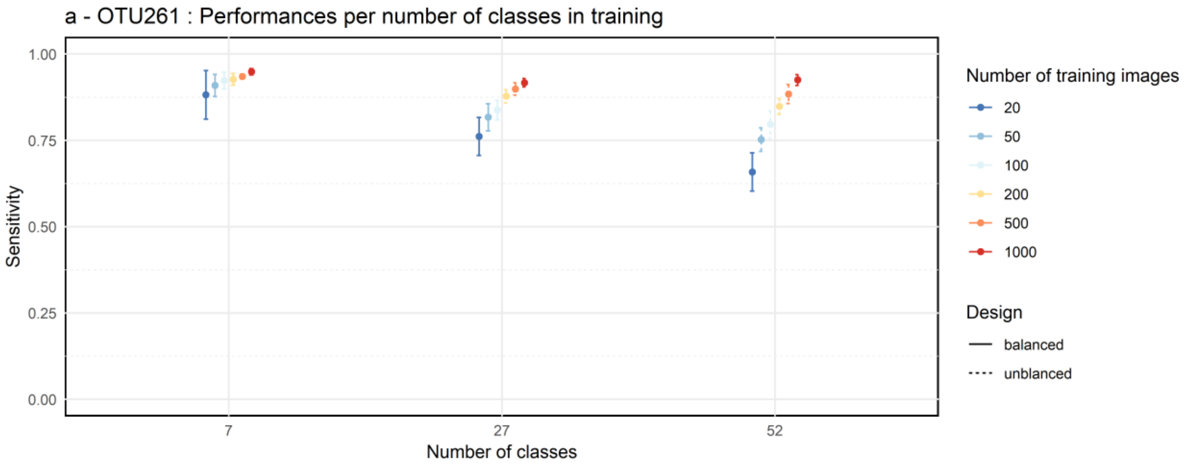
5.4.3 Application of CV to an unbalanced ecological dataset

When considering all treatments in an unbalanced design (Figure 5-4), the average sensitivity per treatment ranges from 0.32 to 0.78. The highest sensitivity was achieved by the A1000 classifier (7 classes, with 1000 training images in each class) while the lowest was achieved by the B20 and C20 classifiers (27 and 52 classes, respectively, and 20 images in each class). A1000 also had the highest precision (0.75), with the lowest precision observed in the C20 classifier (0.20). Sensitivity of the

C1000 classifier (where class imbalance is highest) was lower than in the C100 and C200 classifiers but precision simply increases with the number of training images, although this could be an artefact driven by the improvement of precision on the most abundant classes.

When considering individual OTUs, performance was unacceptably low for most, but not all as some had sensitivity and precision greater than 0.85. Based on average sensitivity across all treatments, the top 10 and the bottom 10 OTU classes were identified. The top 10 classes were large organisms with consistent or distinctive shape, colour and patterning. They were not necessarily the most abundant classes as six of them were only present in group C, for which there are less than 100 training images, and only two in A, for which there are at least 1000 training images. Of these OTUs, the two present in group A had better average precision than any other OTU class in the top 10. The OTU classes with the worst performances are generally those for which there are fewer training images (group C). They also tend to be smaller organisms, have colours similar to the background and have very variable shapes and sizes.

In this dataset, CV could be applied to OTU261 and OTU339. These OTUs were both very abundant in the study area, justifying automated annotation, and they both had very high performances, making their identification by the classifier reliable (Figure 5-5).


a - OTU261 : Performances per number of classes in training

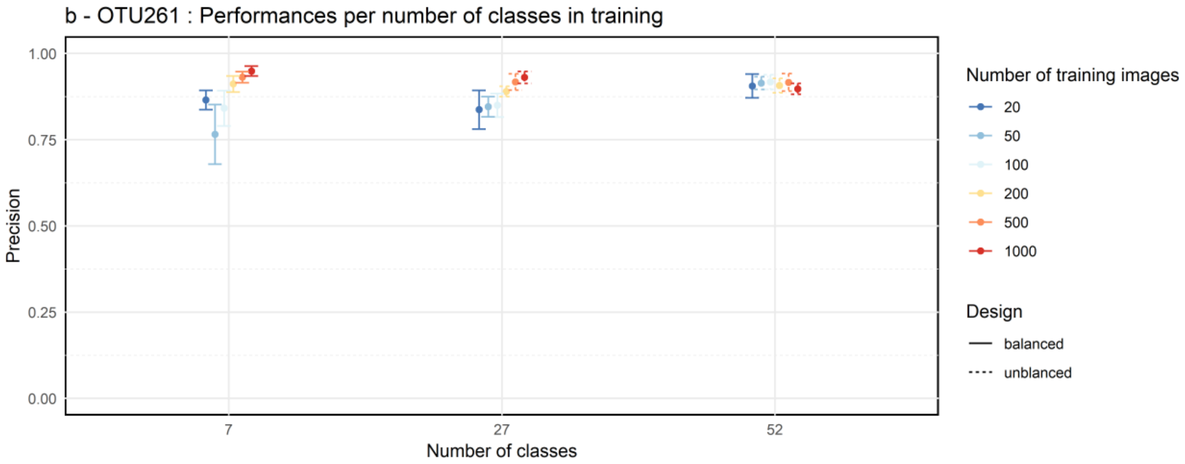

b - OTU261 : Performances per number of classes in training

Figure 5-5: a) Differences in sensitivity for OTU261 in classifier trained with different number of classes and images (7 for group A, 27 for group B and 52 for group C). Error bars are standard deviation calculated from the 10 random splits. b) Differences in precision for OTU261 in classifier trained with different number of classes and images. Error bars are standard deviation of the 10 random splits

The performance of CV for OTU261 and OTU339 was maximised in the A1000 classifier with only 7 classes and 1000 training images. The A200 classifier also achieved performances close to A1000, despite being trained on five times less images. For OTU261, even the A20 and A50 classifiers achieved sensitivity and precision greater than 0.86, and differences between the A20, A50 and A100 classifiers were not statistically significant (Figure 5-5).

Sensitivity in the C1000 classifier was 0.92 and 0.89 for OTU261 and OTU339, respectively, which is significantly lower than the A1000 classifier (p-value <0.0001 for both – Figure A5-2 and A5-3) but only a marginal difference (0.03 each). For OTU261, the C200 classifier achieved lower sensitivity than the A200 but they had equal precision. For OTU339, precision is also the same in the A200 classifier and all classifiers in group C (Figure A5-4). Note that for both OTUs, precision of all treatments in group C were either not significantly or barely significantly different (p-value above 0.01). Thus, classifiers in group C (with 52 classes) achieve performances almost as good as classifiers in group A when training on 200 or less images.

Group B classifiers tended to show slightly lower sensitivity than group A classifiers and slightly lower precision than group C classifiers, although often not with significant differences.

5.5 Discussion

In this study, our purpose was to test the capacity of a transferred CNN classifier (partially trained on a different dataset) to identify benthic organisms and, by extension, to test if this methodology can be successfully applied in ecology by non-specialists with a relatively small data set, open-source software and libraries, as well as a short investment in time after manual image annotation.

5.5.1.1 Overall performances

Our classifiers achieved a maximum average performance of 78% (0.78) in sensitivity and 75% (0.75) in precision. In other studies, performances achieved through manual annotation range from 50 to 95% for benthic fauna (Durden et al., 2016a, Beijbom et al., 2015) and 84 to 94% accuracy for plankton (Culverhouse et al., 2003). There is no consensus on what an acceptable error rate in the ecological literature is but, to be competitive with experts, automated identification performances should be towards the higher end of those achieved manually. In this regard, Culverhouse et al. (2014) report an anecdotal value of 90% correct classification cited by experts. Previous studies on marine ecosystems sampled via images that have attempted to automatically classify multiple benthic megafaunal taxa with various methods sometimes achieve performances comparable to those of experts. Beijbom et al. (2012) found that different coral species in shallow reefs were correctly identified 97% of the time. Schoening et al. (2012) found an average sensitivity of 87% and precision of 67% when classifying deep benthic megafauna in the Arctic. Marburg and Bigham (2016) found 89% accuracy when classifying benthic mobile megafauna off the Oregon coast. When considering other faunal groups, CV can achieve even higher performances, for example, Siddiqui et al. (2018) automatically identified various fish species and were correct 96.7% of the time on average.

Even at their best performances, our classifiers would misclassify more than one out of 5 observations if they were used to make novel predictions. This is not good enough to be considered a suitable replacement for manual annotation. To be the tool benthic ecologists need, average performances need to be increased by at least 10 or 15%.

5.5.1.2 Impact of the number of images in training on performances

In our study, average performance measured as both sensitivity and precision increased with the number of images used in training. Performances obtained with 1000 training images are significantly better than those obtained with fewer images, for example, those obtained with 200 images (five times less). This difference, however is marginal. Extrapolation of the data suggests that performances may never greatly exceed those obtained with 1000 training images regardless of how many images are used.

It has been generally demonstrated that more data is preferable when modelling (Enric et al., 2013) and training classifiers (Lu and Weng, 2007, Maxwell et al., 2018). Unsurprisingly then, our results suggest that the number of training images has a clear positive effect on performance, particularly on sensitivity. Sun et al. (2017) tested their generalist object classifiers with 10, 30 and 100 million images and observed a clear increase in performance. Siddiqui et al. (2018) also found that increasing the size of a dataset by 25% (20000 to 25000 images) resulted in a 6.6% increase in performance of the same CNN.

More data, however, is not a simple solution to low performance as the relationship between the amount of training data and performance is not linear. Sun et al. (2017) report a logarithmic relationship between the size of the training set and performance. These authors gained less than 20% increase in performance by adding 90 million images to their training set. This logarithmic relationship has also been reported by Favret and Sieracki (2016) in their fly species classifiers. These authors note a diminishing return of adding more training data and observed little gain when doubling their training size from 50 to 100 images. Cho et al. (2015), who classified computed tomography images of six human body parts, found the same logarithmic relationship and, although it was 95.7% with 200 training images, their desired 99.5% accuracy target was only reached with 4092 images. Thus, there is an optimal size to every dataset beyond which more training data results in very little gain. This point can be determined by the goal of the study and what is considered acceptable performance. With our methodology, this point occurs at 200 images for the dataset we used, and represents a reasonable amount of manual work for ecologists aiming to build the dataset to train a CNN.

5.5.1.3 Impact of the number of OTU classes in training on performances

We observed that classifiers with a small (7) number of classes had better performances than those trained with 27 or 52 classes. The difference in performance between the latter two was marginal, although significant.

The number of classes in machine learning studies is usually driven by the dataset and the research question rather than maximizing performance by limiting the number of classes. Thus, few studies have assessed the effect of that number on their performance. In the 24 CV-based animal identification studies cited by Favret and Sieracki (2016) and Weinstein (2018), no significant correlation exists between the number of classes used in each classifier and their respective performances. In their large dataset experiment, Sun et al. (2017) also found no difference when training with 1000 or 18000 classes. But in contrast, Favret and Sieracki (2016) observed a counterintuitive increase in performance as more insect species were included into their training set. They hypothesised that, although a higher number of possible outcomes could increase confusion, the higher number of comparison points helped determine the important features of each category. Further tests are needed to disentangle the effect of the number of classes in training or the relative difference in morphology of these classes on performance. In general, practical applications of CV in ecology would benefit from more information on this effect. A deeper investigation of the results of experiments like the one presented here could also shade some lights on this aspect of application of CV to real world data.

5.5.1.4 Potential application of CV to a real ecological dataset

To deploy classifiers such as these in a “real-life” ecological study, reasonable performances must be achieved while retaining time and cost effectiveness of building the training set.

In our study, no classifier achieved average performance above 78% (0.78), which would mean one misidentification out of 5 predictions, at best. We also observed high interclass variability as some OTUs were consistently well identified while others were, on the contrary, always misclassified. Even if the measured average performances were considered acceptable, it would introduce completely false appreciation of the distribution of some OTUs and local diversity. For example, false detection of rare taxas could drastically misrepresent their prevalence which (shown in chapter 3) could have important consequences on perception of sampling completeness.

This variability in both expert and machine classification performance between classes or taxa has been observed by other authors (Beijbom et al., 2015, Cho et al., 2015). Experts in Durden et al. (2016a) had various annotation successes for different taxa and Schoening et al. (2012) found that human observers and their semi-automated classifier had variable success at detecting and identifying different taxa but agreed on which one had the best performance. It is therefore sensible to consider the predictions of each OTU class separately and only rely on those for which the classifier achieves good performances.

Good performance obtained by our classifier with some specific OTU classes is encouraging and automated annotations could be an appropriate method to study these OTUs. The top 10 best and worst OTUs ranked by sensitivity shows that the classifiers are better at identifying large sized organisms exhibiting a low intra-class morphological variability.

The majority of the top 10 OTUs were rare (e.g. less than 100 training images). If CV were applied to these rare taxa, there would be a proportionally higher impact of any misidentification or false positives (predictions of presence that are in fact another OTU) on the results. Yet, given their relatively low number of occurrences (tens to a few hundreds), a manual verification step (or semi-automated identification), as performed by Schoening et al. (2012) and suggested by Marburg and Bigham (2016), would be easy to perform for a reasonable time investment and to ensure the reliability of the predictions. For example, the Largo tool in Biigle, used in this study to validate the training set, makes a visual check of a large number of annotations much faster than going over the raw images again. It would be an efficient way to validate the CNN's predictions and make results usable. On the other hand, OTU261 and OTU339, both among the top 10 OTU classes, were very abundant in the study area (above 1200 individuals). In a larger dataset, manual validation of identifications of these OTUs would be impractical and, to some extent, cancel the gains in speed and objectivity of CV. Ideally, their identification should be fully automated if the classifier is to be deployed in these conditions.

CNNs are considered as "black boxes" whose internal prediction and decision process are difficult to visualize and understand (Samek et al., 2017), yet, we can speculate on the reasons why some organisms are better identified than others. OTU261 is very constant in shape and colour and has a distinctive pattern on its outside: this homogeneity probably makes it easily identifiable. OTU339 can be in different pose or orientation within an image but has a number of distinguishing features, such as its reflective eyes, and its long, often spread-out, limbs. These features are not found in other OTUs, probably making confusions rarer.

OTU2 and OTU23 are both anemones. OTU2 is a cerianthid (a tube anemone) of various size and orientation and OTU23 is a Halcampidae/Edwardsiidea like anemone of very small size. They are similar in shape and size, hence distinguishing them is difficult even for human annotators. This could explain the lower performances of the classifier on them. The fact that, during annotation, the smaller OTUs were localized with point coordinates and an arbitrary radius was used for slicing is a potential source of bias. The 240 pixels square used by default leaves a large surface of the image as background. This feature, common to several small OTUs, including the small anemones, could be a cause for higher rates of confusion in the predictions. With OTU261 and OTU339, high sensitivity (up to 0.95 and 0.92, respectively) and high precision (up to 0.95 and 0.82, respectively) were achieved by the classifiers, meaning they were usually correctly identified and false positives (another OTU wrongly identified as one of them) were relatively rare. These performances are equivalent to those of human experts working on a very similar ecosystem (Durden et al. 2016) without the inconsistency over time by individual observers reported by these authors. Moreover, based on the speed of manual annotation in this study, we estimate that building the training set, validating it, training the classifier and testing it could be achieved in a matter of days rather than months. Therefore, these classifiers can be applied to the remaining un-annotated images in our dataset and provide useful presence records of these specific OTUs. This would be a valuable contribution to this study of deep-sea ecosystems.

Classifier A1000 had the best performance of all classifiers and would detect almost all individuals of OTU261 and OTU339, but it needs a large training set, while the A200 classifier has very similar performances but needs five time less training material and is therefore more cost-effective. These group A classifiers however, risk producing a high number of false positives if they encounter too many individuals of an OTU they have not been trained on. Thus, it is only applicable if diversity at the study site is low or it is predominantly represented by a small number of OTUs. These classifiers would not be suitable to survey very diverse ecosystems, like coral reefs.

In the long term, classifiers able to identify as many OTUs as possible, even semi-automatically, are undoubtedly more desirable, even if they perform slightly less well. In our study, the C classifiers had marginally lower performances than A classifiers, particularly if training with 200 images, but both sensitivity and precision were above 0.9 for OTU261, which is still comparable to manual annotation. Thus, although this design is still valid for identifying specific OTUs, it has the advantage, as it is trained on 52 classes, to be able to automatically identify more OTUs. Even if some of these identifications need to be manually validated, it is more representative of real field studies where many OTUs could be encountered.

Based on our observations on classifier performances, we recommend the following approach to the use of CV in small-scale benthic ecological studies: 1) Build a general classifier to identify OTUs that achieve good performance and quantify the error rate associated with each. This can be an unbalanced design with many OTUs, like group C in the current study. A large number of classes potentially allows more OTUs to be tested. The number of training images should preferably be above 200 plus a 20% or more surplus so the classifier can be tested with independent data. 2) Only use the presence prediction of those OTUs that have good performances and regard any other predictions as unknown or an absence of those. 3) Consider all remaining OTUs as “unidentified” and leave for manual identification or for later, more efficient, automated classifiers. Alternatively, a one-vs.-all classifier could be trained and deployed for each of the target OTUs (Siddiqui et al., 2018) but this approach would become logistically challenging for a large number of target OTUs.

Even if the presence records of some OTUs are not sufficient to understand the composition and dynamics of an ecosystem, it will still contribute to it and more importantly, it will take-on some of the annotation time, leaving experts free to perform other tasks while providing useful insights in ecology. In the specific case of this study, the automated identification of OTU261 and OTU339 would be useful for deep-sea ecologists, especially if it only requires a few days of work. Indeed, very little is known about the fine scale distribution of these OTUs. *Syringammina fragillissima* (OTU261) is considered habitat forming, enhances local metazoan abundance (Levin and Thomas, 1988, Levin et al., 1986, Gooday, 1984) and is a Vulnerable Marine Ecosystem under United Nations General Assembly Resolution 61/105 (Assembly, 2003). The squat-lobsters *Munida sarsi* or *M. tenuimana* may play an important role in the benthic community as predators or scavengers (Hudson and Wigham, 2003) and are suited to examining ecological patterns (Rowden et al., 2010). Extracting the location of these two taxa from a vast dataset would be a valuable way to study or map their extent and distribution as other studies have done with other faunal groups at fine (Milligan et al., 2016) and broad scale (Rex and Etter, 2010, Wei et al., 2010a). Besides, this would complement the studies carried out by trawling, which can underestimate diversity of benthic crustaceans (Ayma et al., 2016, Cartes and Sarda,

1992) and destroy xenophyophores (Roberts et al., 2000). Furthermore, appropriate data is currently lacking to study rhythmic diel and seasonal movements or behavioural changes of megabenthos, including squat-lobsters (Aguzzi et al., 2013, Aguzzi and Company, 2010). By providing more data on abundance and distribution of *Munida*, this method could greatly help this field of research. Also, assuming that benthic decapods can easily be counted with CV, and abundance differences reliably measured, the stock assessment of *Nephrops norvegicus*, an important and carefully monitored commercial species (ICES, 2010, Sardà and Aguzzi, 2012) could be achieved at greater speed, cost-efficiency and more objectively than by trawl.

This study only deals with the identification of animals and not with their detection within the images, which was performed manually in Biigle before cropping images around each individual. Detection is an essential step in automated image analysis and many solutions have been explored (Hollis et al., 2016, Sorensen et al., 2017, Cheng and Han, 2016). A step for object detection needs to be added to the protocol described here to completely automate the process. This study also did not deal with the behaviour of the classifiers when presented with novel OTUs. This situation is unavoidable in real-life ecological datasets, and although methods exist for novelty detection (Pimentel et al., 2014), this remains to be integrated into our methodology.

5.6 Conclusion

Our results demonstrate that CV based image annotation cannot entirely replace manual annotation of benthic images at present, but that usable results can be obtained for specific taxa with open-source software, very little tuning and optimisation of the model itself and a relatively small training dataset (200 images). These results can inform the distribution of these specific taxa in a more robust way than currently possible. In general, monitoring the abundance of a single taxon for novel research or in routine stock assessment could greatly benefit from this method. It offers greater speed, cost-efficiency, objectivity and consistency than trawl surveys or manual image analysis.

This does not immediately solve the many challenges of marine ecology but could initiate momentum and catalyse further development of CV based methods in this area as these tools are becoming more accessible to non-specialists. The development of fully automated image annotation, or pragmatic combinations of manual and automated annotation protocols (Matabos et al., 2017), is likely to continue across different platforms capable of gathering large image datasets (Marini et al., 2018b, Marini et al., 2018a). Indeed, there is still much room left for improving classifier performance with better image pre-processing prior to the training or better tuning of the model, and more research could lead to game-changing methodological development. In the age of big data and global open research, the participation of many different actors of research contributing data (Hussey et al., 2015, Hampton et al., 2013), computing power, and above all, taxonomic and informatics expertise (Weinstein, 2018) could be synthesised in the development of CV tools able to take on some of the workload of human researchers and increase the pace at which the oceans are explored and sampled and, ultimately, how they are preserved.

5.7 Acknowledgments

The authors would like to acknowledge the officers and crew of RV James Cook as well as the Autosub support team who assisted in the collection of the data. The DeepLinks project is funded by NERC NE/K011855/1. Piechaud’s PhD, Hunt’s time and the processing of the data in Google cloud were funded by University of Plymouth. We also would like to thank Kirsty Morris, Erik Simon Lledo and Marcus Shirley for their advice on processing the images prior to analysis.

Chapter 6: General discussion and conclusion

This thesis aimed to investigate how Autonomous Underwater Vehicles (AUVs) could contribute to addressing long-standing challenges in deep-sea benthic ecology. Deep-sea data is difficult to acquire and it takes a lot of effort and financial resources to gather evidence of any phenomena. So far, although much progress has been made in recent decades, the basic questions in deep-sea ecology remain unanswered. AUVs may have the potential to change this but there are few practical examples of their application and what sets them apart from other sampling gear is not yet clear to benthic ecologists.

This work revealed a number of new elements on the ecology of the ecosystem studied. More importantly, it has provided a certain level of understanding of the capacities and limitations of the cruising-class AUV Autosub6000 in this specific context. Based on that experience, I attempt to provide guidance on what researchers should and should not expect from these technologies and how their current and future capacities can help deep-sea benthic ecology.

6.1 On AUV contribution to addressing the challenges of deep-sea ecology

In **chapter 1**, I listed some of the challenges that AUV data could help address. In the following chapters, I evaluated if the data collected by an AUV could indeed shed new light on species diversity, species distribution and the environmental drivers of that distribution. I also investigated how manual and automated image analysis methods were able to exploit the AUV data.

Chapter 2 provided details on the origin of the data, on the vehicle used to obtain it and on the methods used to process and prepare the data for several case studies of deep-sea benthic ecology. Once the data was ready, I analysed it to assess what ecological knowledge it could bring.

I found in **chapters 3, 4 and 5** that AUVs could, at least partially, address the challenges faced by deep-sea benthic ecologists and have advantages that could, in the future, greatly contribute to the improvement of our knowledge of this ecosystem. The following section discuss the results of this thesis in more detail to describe the pros and cons of AUVs in the context of each individual challenge.

6.1.1 Species richness

The challenge of measuring the number of species present in a given location comes down to the amount of sampling effort accomplished in the field. In an ideal theoretical setting, the sampling can continue until no more new species are discovered and the sampling can thus be considered complete. **Chapter 2 and 3** showed that Autosub had the capacity to gather the necessary data to evaluate the local species diversity of a station, in a relatively short time. It supplied enough images to identify the totality of the OTUs present at the station while quantifying the variability of the measures of density and diversity. The same protocol could be used again in the same station to detect a potential change in diversity or density. Other communities in different locations within the Rockall Trough and over environmental gradients (like depth or proximity to Rockall Bank) could also be compared to the one described here, if samples of recommended sizes are used.

Chapter 3 also identified the relative importance of rare species in the measures of species richness and showed that, in theory, the AUV had the potential to encounter them all offering the possibility to better study them. Many of the species and morphotypes found in deep-sea benthic ecosystems are rare or at low density but their importance in its conservation and ecology is increasingly acknowledged (Turner et al., 2017, Connolly et al., 2014, Chapman et al., 2018, Goineau and Gooday, 2019, Gray et al., 2005). The rarity of these taxa makes them vulnerable to environmental changes while their possible disappearance can easily go unnoticed as gathering information about them is hard (Costello and Chaudhary, 2017). More data is needed to determine their extent and distribution in order to include them in conservation strategies (Goineau and Gooday, 2019). If the data analysis requirements can be met, AUVs can facilitate the study of these rare species as part of the deep-sea diversity. The data analysed in **Chapter 3** was not sufficient to truly replicate the measures of species richness. That has been gathered by Autosub and is available for further analysis, therefore, a truly replicated measure of diversity at station 26 could be obtained. Replication of sampling transects is highly desirable (Underwood, 1994, Morrisey et al., 1992, Halpern, 2003) in order to evaluate the variability of local diversity and density. Furthermore, replication should preferably be conducted with independent samples rather than random resampling with replacement as those do not give independent replicates (Hurlbert, 1984, Noble-James et al., 2017). Again, Autosub and AUVs of its class are capable of cost-effectively delivering the necessary amount of data, hence contributing to the improvement the quality of deep-sea ecological datasets.

6.1.2 Fine-scale species distribution

The study of fine-scale distribution of epibenthic species is complex and poorly understood. It is challenging to study because it needs both precise spatial positioning of the samples, at least relative to each other, and large amounts of data to ensure statistical robustness of the conclusions. in **Chapter 4**, the analysis of 6500 samples (images) of the density of a single benthic species (chosen for its ecological and conservation importance) collected by the AUV was able to not only detect important variations in that density but also measure the scale at which it varied and, more importantly, replicate this measure four times.

These results showed that AUVs are able to quickly collect a dataset large enough to detect and quantify spatial heterogeneity, where other gear would have been less conclusive. This could result in substantial progress towards a better understanding of how communities are distributed in the deep sea, as well as the way individual species are distributed within these communities. That knowledge can, in turn, benefit mapping studies (Perkins et al., 2019), facilitate monitoring (Foster et al., 2018) and eventually efficiently inform conservation (Danovaro et al., 2016).

The study of one species is useful but not sufficient to understand the dynamic of an entire ecosystem even locally. Generalising the responses of one species to other species or entire communities is only appropriate if supported with relevant ecological knowledge, which does not yet exist. Ultimately, data on all species are needed (Eaton et al., 2018, Turner et al., 2017, Chapman et al., 2018). The target species, *Syringammina fragilissima*, in **chapter 4** is known to enhance the local abundance and diversity of meiofauna (Levin, 1991), which gives value to its abundance as a proxy for local diversity. This chapter did not however inform on the distribution of other habitat building species present in the area, like the large sponges, nor on interspecies species interactions, which would be infinitely more valuable for conservation. This confirms how the advantages of the AUV are somewhat reduced by the impossibility to fully use the data it provides.

6.1.3 Drivers of species distribution

Knowledge of the drivers of species distribution are an essential element to know in order to understand and predict ecosystem dynamics. As shown in **chapter 2**, AUVs can measure and record many environmental parameters, hence, giving a very detailed picture of the habitat. This increases the chances of identifying the parameter or combination of parameters that correlates best with the distribution of a species. Autosub proved to be an efficient tool to swiftly perform these measures over large areas.

In **chapter 4**, it appeared that the variation in density of *S. fragilissima* was locally correlated with the environmental predictors collected by the AUV. This is a very useful discovery, since the correlation allows the use of species distribution modelling to make predictions of *S. fragilissima* density where no direct measurements are possible. However, replication of this experiment and validation of the modelled relationships established in one area to another area nearby showed that these local apparent relationships did not hold across transects, despite their close proximity.

In that sense, even this important effort of AUV data analysis was not sufficient to determine precisely what drives the distribution of *S. fragilissima*, but it nonetheless showed how a small, spatially concentrated dataset was inappropriate to study the drivers of *S. fragilissima* distribution and was, in fact, potentially misleading. This highlights another strength of the AUV as it provided a dataset that allowed a very thorough testing of the analysis method, which proved fruitful by detecting the flaws in its results.

Unlike **chapter 3**, the limitation of the dataset in **chapter 4** was probably not the amount of data per se, but the way the data points were distributed and it is likely that, if the data were better spread-out across the survey area, the same amount of data would have resulted in better model results. The layout of samples used in this chapter was designed to detect species distribution patterns rather than represent local relationships between species abundance and environmental predictors, and even a large dataset may give a biased answer if used for a study it has not been designed for. The manual analysis bottleneck prevented the analysis of additional data as it did in **chapter 3** but the entire survey could accommodate a different, sampling design covering larger environmental gradients if it had been entirely annotated (see the map in figure 2-4 in **chapter 2**). This, again demonstrates that Autosub had provided a dataset able to address a number of challenging question although data analysis bottleneck prevented its use.

With their capacity to collect large amount of both images and environmental data, AUVs can bring more certainty to the results of studies on the drivers of species distribution provided the right amount of data are collected and, more importantly, analysed. Care must also be taken that the sampling design is appropriate to the question of the study (Brind’Amour et al., 2009, Foster et al., 2014) and the size of these dataset can be exploited by implementing better verification and validation of the ecological patterns they reveal.

6.1.4 Data analysis methods

Both **chapters 3 and 4** observed that the additional data collected by the AUV but not analysed, (not only the images but also the environmental parameters) would have improved the knowledge that can be gained from one deployment of the sampling gear over a period of 20 hours. However, the results of both these chapters were limited in their scope by the bottleneck formed by manual analysis. The image analysis detailed in **chapter 2** is very time consuming, particularly, because a lot of time was invested in ensuring its quality. This illustrates why data analysis needs to be more efficient.

Chapter 4 saw the use of a partially automated object detection algorithm MAIA (Zurowietz et al., 2018), which proved to be a significant yet insufficient gain of time and still required careful verification to ensure the data quality was comparable to fully manual annotations. This has confirmed the limitations of manual analysis and semi-automated analysis.

One of the latest developments in the field of computer vision was explored in **chapter 5**. The results of this chapter indicated that CV could match manual annotation accuracy on a small number of abundant OTUs if trained with enough data. It confirmed that CNN trained (by transfer learning) in Tensorflow, could have direct practical applications in ecology, provided objects could be detected and cropped out of images prior to identification. This chapter presents no technical development but illustrates the availability of this complex technology to a small team using a relatively restricted dataset (by computer science standards). It is an important point to make: if AI can be operated at a scale relevant to individual projects, this flexibility makes it more accessible than it has been in the past.

Interestingly, the species targeted in **chapter 4,** due in part to its ease of detection and identification, is also accurately identified by the CNN, probably for the same reasons. This technology is capable of technical prowess as demonstrated by Krizhevsky et al. (2012), Beijbom et al. (2015), Marini et al. (2018a), Siddiqui et al. (2018), and Schneider et al. (2018). If these classifiers could be applied to AUV data by any team able to operate the vehicle and collect the data in the first place, benthic ecology would be durably changed. This would truly unlock the potential of AUVs and make their advantage in sampling efficiency truly worthwhile compared to other sampling methods.

Chapter 5 also gave possible hints on why CV is not more commonly used despite repeated calls by ecologists and computer scientists (Gaston and O'Neill, 2004, MacLeod et al., 2010). A large part of the time taken to complete this study was spent on understanding the method and learning how to successfully apply it before the proper analysis of the results could start.

The lack of adaptability and the initial difficulty in implementing CV, coupled with the absence of a guarantee of success within reasonable time-frames are probably responsible for its slow adoption by marine ecologists in routine benthic surveys and the mistrust ecologists tend to have in these technologies. Indeed, it requires skills that are not commonly part of the background of benthic ecologists (mostly programming in several languages, algebra and optimisation of computing power) and are currently expensive to hire. It is worth noting that implementation of automated analysis is rarely budgeted for in grant proposals although they are no more costlier than ROV pilots, hence a lack of application in practical case-studies. A change in the mind-set of the community has to take place so that more research is conducted on applying AI and CV to AUV data.

In a more practical consideration, image analysis in this thesis also illustrated how the interface between the raw data and user are the real enablers of complex analysis. The processing of a large number of annotations would have been harder without Biigle and its clear interface. For example, it was quick and easy to ensure the quality of the annotation with the various validation tools that Biigle implemented. It was also easy to crop images around individual annotations to form the AI's training set. The accessibility of CV was greatly increased by Biigle and **Chapter 5** was not only made possible because of the progress made in CV technologies but also because of the data management, manipulation and sharing capacity of this ergonomic and user-friendly program.

Overall, although automated analysis was not achieved in this thesis, possible ways towards it have been identified and there is hope for the future of benthic image analysis. A combination of methodological improvements, better accessibility of non-specialists to advanced analysis tools, and additional experience in the use these tools by the scientific community can eventually build-up towards full automation of image analysis. Meanwhile, the experience and needs of each individual researcher can feed directly into this process, leading to an improved output of data analysis as it combined ecological knowledge and technical efficiency to address very specific and practical problems.

Also, regarding analysis methods other than image annotation, **Chapter 4** also highlighted how the AUV data failed to produce a model of species distribution able to accurately make predictions in nearby areas despite good performances in internal validation. It showed how a large dataset could improve performances of models, more importantly could give a more objective assessment of their capacity to make predictions in data-poor areas. Use of AUVs could therefore lead to more research aimed at producing guidance for ecologists using SDM. Problems, such as algorithm choice, effect of spatial autocorrelation, effect of overfitting and model complexity, can be more thoroughly investigated with larger datasets and should contribute to a better utilization of these tools in the field.

6.2 On the future of AUVs in deep-sea benthic ecology

6.2.1 The future of AUVs

In this thesis, the data obtained from one specific “cruising” AUV has been able to provide valuable insights in deep-sea ecology. In general, however, individual AUVs are not a providential tool whose wide adoption can solve every facet of this research field. They seem to multiply and diversify the disciplines required to process of all the data they collect (geophysics, hydrography, oceanography and now informatics), which have to be matched with equally qualified analysts and could be a strain on some research teams budgets. Furthermore, AUVs need careful data management procedures as the quantity of data can be confusing, difficult to keep track of and challenging to use correctly. AUVs are also perceived as expensive to acquire, complex to operate and not reliable enough on routine deployments (Przeslawski et al., 2018). They cannot collect physical samples on the seabed or perform in situ experiments. Finally, cruising AUVs like Autosub are still not suited to studies of topographically complex habitats which are hosts of some of its most charismatic communities like cold water coral reefs and coral gardens. Other hovering AUVs have increased mobility which enable these ecosystems (Singh et al., 2004, Armstrong, 2016, Armstrong et al., 2019) but can cover less ground (Huvenne et al., 2018).

Technological development in the coming years will address some these disadvantages and quick progress can be made (Lucieer and Forrest, 2016). The new generation of commercial AUVs are smaller and cheaper than previous models, and are thus more accessible and more expandable (Phillips et al., 2017). Their navigation will improve as they are paired with surface vessels enabling satellite positioning (Huvenne et al., 2018). Their endurance will extend as better batteries and more efficient power management is implemented (Roper et al., 2017). They may also be deployed in fleets of cooperating units (Phillips et al., 2017) and their behaviour will become more and more adaptable during their mission with the help of AI (Quintana et al., 2018). Better cameras will also compensate the reduced taxonomic resolution and soon, they will cover more ground, survey more types of habitat, produce even larger datasets and fully bring benthic ecology into the world of “big data”.

As far as deep-sea benthic ecology with image samples is concerned, AUVs are, in a sense, ahead of their time. As seen in this thesis, their capacity is currently not fully exploitable because of the manual analysis bottleneck. Hence, the immediate answer to the challenges I listed may not be to collect more data - although this will be needed eventually – but to analyse more data and better. Therefore, the future of AUVs probably depends on the improvement and automation of analysis methods rather than the improvement of the AUVs themselves.

This calls for more research in that field of AI applied to CV so that these innovations trickle into biological studies. In the future, if the costs of acquiring and operating AUVs decrease and their reliability improves to a point where it becomes easier to operate them than using divers, trawls, drop cameras and ROVs, then the higher demand for performance data analysis might push towards further development of AI-based data analysis methods. In the meantime, AI tools developed for other purposes can be adapted to the context of benthic ecology, thus enabling better use of the AUVs. This is uncertain however. AUVs and AI have co-existed for long enough but research on their combined potential remains rare and, for a long time, the community has only showed limited interest in it.

6.2.2 The future of AI

The future of AI is also full of opportunities. There are now a number of programs that integrate some degree of automation in detection and identification of objects and animals on underwater images: AVED (Cline et al., 2007), MAIA in Biigle (Zurowietz et al., 2018) or VIAME (Dawkins et al., 2017). Research on these tools has gained traction recently and their accessibility is improving fast (Weinstein, 2018). Ecological studies of the deep sea may soon routinely use these applications, regardless of the sampling gear, and solve the manual analysis bottleneck. Further use of CV in biology will not only improve sampling efficiency, but also improve consistency and comparability across time and research groups.

As I experienced for this thesis, challenges remain for ecologists to either acquire the necessary skills to use tools as complex as AI and CV, or secure funding to hire those who can. Both these options are costly and will probably encourage collaborative investment for the development and maintenance of these tools. More research groups sharing their analysis software and methods will also positively impact the comparability of the data, although probably at the cost of flexibility. More data also calls for a change to data management practices towards better curation, sharing and storage strategies (Schoening et al., 2018). This will in turn facilitate greater standardisation of annotation practices, as advocated by a number of ecologists in recent years (Amon et al., 2016, McClain and Rex, 2015, Althaus et al., 2015, Howell et al., 2019).

Provided all their data can be analysed, coupled with possibilities of combining data from different sources, the sampling efficiency of AUVs could lead to important methodological changes in marine ecology when data cease to be a limiting factor. More complex and elaborate sampling designs will be implemented. They will be able to account for biases, like spatial autocorrelation, and have an appropriate level of replication and controls as advocated by Brind'Amour et al. (2009) and Foster et al. (2014). In the future field sampling could even take the shape of complete coverage of an area within which sub-samples are taken at subsequent analysis stages or in later studies with different objectives (Woodall et al., 2018).

If, instead of a day, it takes 10 or 5 hours to representatively sample one station, then more stations can be sampled within the time frame of an oceanographic cruise and coverage at broad scale can also be increased, which will help all fields of benthic ecology. The increased freedom in sampling designs will also allow the nesting of samples and stations within environmental parameter gradients (particularly by depth bands) so the drivers of benthic species distribution can be better understood.

In short, thanks to AUVs and CV, the pace of exploration of the deep sea could increase dramatically.

6.2.3 A word of caution

I found in several chapter of the thesis that collecting a lot of data from one place failed to comprehensively describe the local ecosystem and thus generalisations could not be made outside of the surveyed area, even though the number of data points alone was superior to many published SDM studies. This research undertaken here should be reproduced in other stations within the Rockall Basin in order to confirm or refute the conclusions reached here regarding the ecology of the *S. fragilissima* as well as validate observations made on this specific dataset. As this is unlikely to happen in the near future, it is important to keep in mind how the heterogeneity of the deep sea is constantly re-evaluated (Ramirez-Llodra et al., 2010) and how generalising from a restricted area or unique taxa should be done very carefully (Keith et al., 2012). Large datasets collected by AUVs are useful but size alone is not a sign of representativity of biological phenomena and the value of these datasets should not be overestimated. This word of caution could be generalised to other new technologies, including AI and CV. Although these techniques will provide more information, it does not exempt the use of methodological rigor in data interpretation and a solid understanding of the underlying biological processes shaping the observed patterns.

It seems more research is being done on developing better AUVs, to make them more reliable, more durable, to give them more power and autonomy, to improve navigation, manoeuvrability, and equip them with better sensors. However, in the meantime, all the data are still manually analysed and too little progress has been made in this area and it receives less attention. If this trend is not altered, an exponential increase in data availability will only result in more hard drives pilling up on researchers desks and drawers while they try to manually extract small amounts of information from it (Schoening et al., 2018, Schoening et al., 2017).

More generally, automation will not make people redundant. The algorithms used in CV have fundamental needs for manually annotated data, they will need validation by specialists and they will not be able to deal with novelty for a while. Besides, the need for competent experts in manual annotation remains, as the AI only builds up on the quality of the data it is trained on. Taxonomic research and expert training should be a priority (Schoening et al., 2017, Howell et al., 2019). One can even be hopeful that

AI could relive researchers from repetitive tasks and give them the opportunity to focus on more stimulating questions like rare or novel phenomena.

6.3 Limitations of this thesis and future work

The practical limitations that are likely responsible for the lack of popularity of AUVs and AI, compared to manual methods that are more reliable, have also impacted this thesis. Implementing these new tools is difficult and relies on a copious amount of trial and error. Due to these difficulties, the large pool of images collected by the AUV was not fully exploited here since I analysed 10% of the images at best for one species, and less than 3% if considering all OTUs. However, this was already more than could be obtained from an ROV and, in that sense, formed a large dataset. Nevertheless, this work does not represent the full potential of an AUV to study a single station.

Related to the issue of the amount of data analysed is the limited use of environmental data. Many of the variables measured by the AUV (particularly the hydrographic data recorded by the ADCP) were not used in any of the case studies here. They could have improved the results as the strength and direction of the current is a well-known driver of species distribution in the deep-sea (Henry et al., 2013).

Although combining AI and CV could revolutionize benthic ecology in the very near future, this thesis just falls short of it. Application of the automated analysis method developed here in another ecological case-study was planned but did not materialize due to time constraints. The evident challenge remains to develop the pipeline in which an object detection algorithm, a CNN identifying OTUs, and a thorough manual verification procedure could be implemented. The most promising practical research lead identified in this thesis, in my opinion, lies in the combination of existing algorithms and image analysis software to build such a pipeline. This is a much larger task than was possible in a single PhD chapter, and could form the basis of a post-doctoral research proposal.

Biologically speaking, the scope of this thesis is, regretfully, limited. It only presents the results obtained from one location. These results may or may not be representative of the broader area (the bathyal part of Rockall Basin). Whether the ecological insights gained at station 26 can be applied to other areas is an important modulator of the value of this work. This limitation in scope is partly due to the structure of the fieldwork whose priority was to maximise sample collection with the ROV. Several unsuccessful AUV deployments effectively limited the number stations from which usable images were gathered to a single one. Note that, although it limits the generalisability of the ecological finds of this thesis, the data from one station proved too large by, at least, an order of magnitude to manually analyse. Undoubtedly, splitting the effort into serval stations would have resulted in a dataset limited in other ways.

Many more interesting aspects of the ecology of station 26 could have been investigated, even with the annotations used this thesis and much more with all the images Autosub6000 collected. Transect t2 that I annotated manually included more than 110 OTUs of which several were other VME indicator taxa, for example the sponge *Pheronema carpenteri* (Thomson, 1869), or the bamboo coral *Acanella arbuscula* (Johnson, 1862). Their study is needed for their conservation. How are they distributed? Are the drivers of that distribution the same as *S. fragilissima*? Beyond these species, the other organisms present and their interactions also need to be studied. Are there one or more communities within station 26? Are there a number of statistically and ecologically meaningful assemblages, as there are at a larger scale (Parry et al., 2015, Howell et al., 2010a)? Or, on the contrary, is there continuous spatial turnover of assemblages and an infinite number of local communities depending on the scale considered?

Finally, the availability of more large datasets on the distribution of benthic megafauna could be the occasion to make progress on the understanding of SAC in benthic communities and its effect in SDM. The exact effect of SAC on model performances is not very well understood and subject to much debate among specialists (Dormann, 2007, Gaspard et al., 2019, Miller, 2012). Likewise, the methods to account for it are not agreed upon within the community and the most common answer is to resample the training dataset so that no samples are within the minimum spatial autocorrelation distance (Fortin and Dale, 2009). Clear guidance on the ways to ensure SAC is not biasing the results are needed in the future and more research can be carried out in the field with the aid of AUV data with appropriate designs.

6.4 Concluding remarks

What this thesis achieves with several case studies is a proof of concept that AUVs can bring new insight to benthic ecology. It also showed that AI could produce useful results, even in a small-scale project. The potential of AUVs can only be realised if AI can bypass the manual analysis bottleneck and this research area must gather momentum. Although I feel that more questions than answers have come from these years of study, the way to address some of the challenges of deep-sea ecology is clearer.

AUVs and AI can be tools to scale-up and widen the range of options available to benthic ecologists and help them make better use of their limited time and resources. Their integration into the existing research structure will take time, while their accessibility and reliability is still in development. But the trend towards more autonomy, automation and big data is general in society and goes beyond deep-sea benthic ecology, and even science.

It is important for those researchers using these methods to retain basic knowledge of their inner workings, biases and limitations, to factor them into the interpretation of their data. Many challenges remain on the road, not least, for the community to acknowledge they could be a solution to the need for more and better data. I hope the thesis will contribute to the recognition of the potential of these technologies and that research on their applications are worth considering in future research to study the largest and most mysterious ecosystem on the planet.

Appendices

7.1 Appendix A2:

Appendix accompanying **chapter 2.**

7.1.1 A2.1: Download data and species catalogue with Biigle API

This document explains how to take the CSV report from Biigle, as well as the species catalogue, so that the two tables can be used in an ecological study.

For this tutorial you will need R and R studio installed, as well as the packages “Tidyverse”, ” magrittr”, “plyr”, “fromJSON” and “worms”.

If they need to be installed, run:

```
install.packages(“Tidyverse”, “magrittr”, ”plyr”, “fromJSON”, ”worms” )
```

You also need a Biigle account and have access to the project you want to download data from, and be logged in within your favorite web browser.

Download data from Biigle

In the volume overview window, enter the “request report for this volume” tab, select “annotation report”. Request the annotation CSV variant in the drop down menu. You will be sent an email with a link to download them all in a zip archive. If you click the link, it should initiate download into your default DL folder (part of your browser parameters) and is usually in the windows “Download” folder.

If necessary, like if you need the surface of the image as calculated by Biigle, download the full reports as well. It should be another zip file.

Take both your zip files and put them in an appropriately named folder (the name of the project is a good choice).

Use the species catalogue tree from Biigle

The species catalogue can be downloaded from Biigle using the API. To do that, enter this line in the address bar of your web browser:

https://biigle.de/api/v1/label-trees/[code number of the catalogue you want to download] - Example: https://biigle.de/api/v1/label-trees/25

You will be brought to a page where you can download a JSON document of the species catalogue that the following script can turn into a table.

This JSON document has to be named “tree” and placed in the same folder as your Biigle report. This folder will be your working directory and the output of this script will be imported there. Place the pathway to that folder on the 4th line of the script and run it.

```r
library(tidyverse)

library(magrittr)

library(fromJSON)

# set environments
wd <- "~/AUV PhD/Biigle Data"
setwd(wd)
jsonlite::fromJSON("tree.json") -> d # feel free to change the name


d$labels %>% as_tibble() -> tree
tree %<>% arrange(id) # arrange them in id order (sort of time of creation order?)

#  cycle through all levels

# add a level1 column. it is the first parent id
tree %>% mutate(last_level = parent_id) -> tree.i
i = 0 # start at 0 so it can be updated to 1 at first iteration

repeat {
  i = i + 1
  print(head(tree.i))

  # take the labels that have no parent_id
  tree.i %>%  filter(is.na(last_level))  -> tree_f_out


  # remove the levels that have no parents
  tree.i %>%  filter(!is.na(last_level)) -> tree_f

  # stop if there is no level left to add
  if (nrow(tree_f) < 1) {
    break
  }

  # make it a list to get through levels
  split(tree_f, tree_f$id) -> tree.l

  # to each label, attach the paren ID of the parent ID
  tree.l %>% map(
    function(x)
      mutate(
        x,
        new_level = tree %>% filter(id == x$last_level) %>% pull(parent_id),
        # replace the last level ids by the name of the taxa
        last_level = tree %>% filter(id == x$last_level) %>% pull(name)
      )
  ) -> tree.l

  # remake a table
  tree.l %>%  map_df(bind_rows) -> tree_f

  # attach with hte labels excluded before
  # it will automatically assign an NA at the new level column for the labels out
```

```
  bind_rows(mutate(tree_f_out, last_level = as.character(last_level)),
            tree_f) -> tree.i

  # change the names
  tree.i %<>% rename_at(vars("last_level"),  funs(paste0("level_", i)))
  tree.i %<>% rename(last_level = new_level)

  # next level
}# next level

# give it the same format as the one out of biigle
tree.i -> tree_leveled

tree_leveled %>% split(tree_leveled$id) -> tree.l


tree.l %<>% map(function(x)
  mutate(
    x,
    label_hierarchy =  x %>% select(contains("level")) %>%
      select_if(!is.na(.)) %>% # remove nas
      select(rev(names(.))) %>%  # get Biota as first level
      paste(collapse = " > ")  # past into one vector
  ))
# reform table
tree.l %>%  map_df(bind_rows) -> tree_leveled_h

# label hiereachy without the final OTU name
# number of hierarchy levels
tree_leveled_h %<>% select(-contains("level")) %>%
  mutate(n_taxonomy_levels = label_hierarchy %>%
           str_split(pattern = " > ") %>%
           map(function(x)
             return(length(x))) %>%
           unlist)
# change top labels to 0 levels of taxonomy
tree_leveled_h[is.na(tree_leveled_h$parent_id), "n_taxonomy_levels"] <-
  0

# make the taxonomy column a table
taxonomies <-  tree_leveled_h %>%
  pull(label_hierarchy) %>%
  str_split(pattern = " > ")
# add names to thins list
names(taxonomies) <- tree_leveled_h %>% pull(name)
# maximum number of levels
taxonomies %>% map(length) %>% unlist() %>% max() -> max_levels

taxonomies %<>% map(function(x)
  c(x, rep("OTU", max_levels - length(x)))  %>%
    t() %>%  as_tibble(.name_repair = "unique"))  %>%
  map_df(bind_rows)
# change the names in the table
names(taxonomies) <- paste0("level", 1:max_levels)


tree_leveled_h  %<>% bind_cols(taxonomies)


# look for duplicates
tree_leveled_h   %>% count(name) %>% filter(n > 1)

# remove duplicated names
tree_leveled_h  %<>% distinct(name, .keep_all = T)
```

```
# export to CSV
 tree_leveled_h %>% write_csv("species catalogue Biigle.csv")
```

Add the taxonomic levels with the “Worms” package

```
 tree_leveled_h %>%  filter(!is.na(source_id))  -> worm_tree


# make a worm table for these id_s
worm_tree %>% split(worm_tree$id) %>%
   map(function(x) worms::wormsbyid(x = as.numeric(pull(x,source_id))) ) %>% map_df(bind_ro
ws) -> worm_table

# merge that table with our biigle catalogue
worm_table %>% as_tibble(.name_repairR = "universal" ) %>%
   select( source_id = AphiaID , rank,phylum,class,order,family ) %>%
   mutate(source_id = as.character(source_id)) %>%
  left_join( worm_tree,., by="source_id")  -> worm_tree


bind_rows(worm_tree  ,  tree_leveled_h %>%  filter(is.na(source_id))   ) -> worm_tree_level
ed



# for each OTU, look up through the levels and the first one with a phylum gives its name


worm_tree_leveled %>% split(worm_tree_leveled$id) -> l

l$`3474` -> x

   for(i in seq_along(l)){
      l[[i]] -> x
      # get levels 1 to 3
         x %>%  select(level1:level4) %>% unlist() %>% as.vector() -> phyls
      # get the phylum list for
         worm_tree_leveled %>% filter(name %in% phyls) %>% pull(phylum) %>% na.omit() %>% u
nique -> phyls
      #    special treatment if it is foraminifera
      if( str_detect(x$label_hierarchy, pattern =  "Rhizaria") ) {
               x$phylum <- "Foraminifera"
               print(paste("found some chromy in ", x$name))
         } else if (TRUE %in% is.na(x$phylum )){ # if there is no phylum with label, add th
e phylum form the above levels
            x[,"phylum"] <- phyls
             }

        l[[i]] <- x
   }

# Number of OTU per phylum

map_df(l,bind_rows) -> worm_tree_leveled

worm_tree_leveled %>% count(phylum)

# Which label are related to Chromista?

worm_tree_leveled %>% filter( str_detect(label_hierarchy,pattern =  "Chrom"))

# export the table
worm_tree_leveled %>% write_csv("species catalogue WORMS Biigle.csv" )
```

7.1.2 A2.2: R code used to process the Output of Biigle

This document explains how to take the CSV report from Biigle and turn it into table that can directly be used in most ecological analysis.

- It should create:
  - A table of OTU abundance per image (image counts)
  - A OTU x Samples contingency table (Bio)
  - A contingency table with info on individual images (BioMeta)
  - A table of info per OTU (OTU meta)
  - A plot of the abundance of individuals in each phylum
  - A plot of a multivariate analysis of the images.
  - A plot of the rarefaction curve
  - A table of the diversity estimates

It has been written so that it should run if you only supply the name of your working directory (the name of the folder if which your data is and want to work from) and the name of the Biigle CSV report, which itself is the name of the volume. It needs the table of the OTU catalogue used on Biigle. Instructions on how to download this catalogue are attached to the same appendix of my thesis. This table needs to be in your working directory.

This is a generic approach and you are encouraged to change, modify, customize and improve the code.

Data preparation

Load packages

Make sure they are all installed and up to date

```
library(tidyverse)

## Warning: package 'dplyr' was built under R version 3.6.1

library(magrittr)
library(vegan)
library(goeveg)
library(iNEXT)
```

Create a folder that will contain all the processed data

Create objects for folder name so that R can export the results in the right place. You should have the data in your working directory (wd)

Your wd should contain: * A Biigle CSV report named after the Biigle volume it is from (it is the default name) * A species catalogue table formatted as instructed in A2-1.

```
# set environments
wd <- "~/AUV PhD/Biigle Data" # insert folder name of WD
#enter the name of the project here
project_name <- "341-nrb-1200m-t3-framegrabs-1min"# insert project - Biigle

# set a directory where species catalogues are on your computer
"species catalogue Biigle JAN2019" -> species_cat.name

# Import the the Biigle species catalogue
read_csv(paste0(wd, "/", species_cat.name,  ".csv"),
         col_type = cols()) %>%
  # remove some column in that table so they are not added twice
  select(-contains("level"),-label_hierarchy,) -> Biigle_catalogue

setwd(wd)

# make a folder with the results
res.dir <- paste0(wd, "/", project_name, "_res")
# if it doesnt exists, create a folder for your project results
if (dir.exists(res.dir) == FALSE) {
  print("creating folder")
  dir.create(res.dir)
}
```

Open the Biigle report

Here, the script will also print the number of images, the number of OTU and the list of OTU present in the volume so users can verify these numbers match with their expectations.

```
# open the Biigle project
    readr::read_csv(paste0(wd, "/",project_name,".csv"),col_type = cols()) -> D

# make an image key table
    unique(select(D, image_id,filename)) -> d_imageNames
    print(paste("number of images: ",nrow(d_imageNames)))

    print(paste("number of OTUs: ", D %>% pull(label_name) %>% n_distinct() ) )

    print(paste("OTUs and annotations include: ", paste(
    D %>% distinct(label_name) %>% pull, collapse = ", ") ))

#
# list of species/OTU ( labels names)
    unique(D$label_name) -> OTUs
```

OTU list per image

Process each image and group the annotations per OTU

```
# split the table per image
      split(D,D$image_id) -> l_image

# make two lists
      l_image_meta <- list()
      l_image_count <- list()
```

```
    for(I in seq(names(l_image))){
      # select an image
      l_image[[I]] -> D_I

      # make a table to become contingency
      unique(select(D_I,
                    label_name,
                    user_id,
                    image_id,
                    filename,
                    label_hierarchy)) -> image_meta.I
      # collapse multiple shapes for a single OTU into 1 vector
          D_I %>%
            group_by(shape_name, label_name) %>%
            count() %>%
            group_by(label_name) %>%
            mutate(label_shapes = paste(shape_name,collapse = ",")) %>%
            ungroup() %>%
            distinct(label_name,.keep_all= TRUE) %>%
            select(label_name,label_shapes) %>%
            full_join(image_meta.I,by="label_name")  -> image_meta.I

      # concatenate each annotation by OTU and count them
        D_I %>% count(label_name) -> d.I
            mutate(d.I,annotation_label_id=unique(D_I$label_id))[,c(1,3,2)] -> d.I
            names(d.I) <- c("label_name","annotation_label_id","count")
      # make a count per OTU per image list
        full_join(image_meta.I,d.I,by="label_name") -> l_image_count[[I]]

    } #next image

    # combine the results into table and reshuffle
    map_df(l_image_count,bind_rows) -> image_meta
    select(image_meta,image_id,user_id,filename,label_shapes,
           annotation_label_id, label_name,label_hierarchy ,count)-> image_counts

    # remove the non living labels
    image_counts <- image_counts %>%
      filter(label_name != "Natural structure") %>%
      filter(label_name != "rock") %>%
      filter(label_name != "Laser Point") %>%
      filter(label_name != "Human artefacts")
    # skip if there is no label
    if(nrow(image_counts) < 1) {
      print("no annotation here")
      next}

    # reorder names and rename surface
    image_counts <- image_counts %>%
      select(image_id, user_id, filename,
             label_shapes,
             annotation_label_id,label_name, count,
                                             label_hierarchy)
```

Export the results into a useful format

```
    # export image count
readr::write_csv(image_counts,
                 paste0(res.dir,"/",project_name ,
                        "_image_counts.csv"))
```

Table of information for each OTU

```r
# Table of metadata per OTU
D %>%
  distinct(label_name, label_hierarchy, label_id) -> OTU_meta
# comput a few stats for that project
image_counts %>%
  group_by(label_name) %>%
  summarise(Tot_abundance = sum(count)) %>% # total abundance of that OTU
  right_join(OTU_meta, ., by = "label_name")  %>%
  arrange(desc(Tot_abundance)) -> OTU_meta # rearrange the table by abundance
# number of images wher OTU is present
# cummulative sum from
OTU_meta$cumulSum <- cumsum(OTU_meta$Tot_abundance)
#
OTU_meta <-
  mutate(OTU_meta, percent_total = round(cumulSum / (sum(
    OTU_meta$Tot_abundance
  )), digits = 5) * 100)
```

Add the taxonomic levels to that table

```r
# Label hierarchy without the final OTU name
# Number of hierarchy levels
OTU_meta %<>%
  mutate(n_taxonomy_levels = label_hierarchy %>%
           str_split(pattern = " > ") %>%
           map(function(x)
             return(length(x))) %>%
           unlist)

# make the taxonomy column a table
taxonomies <- OTU_meta %>%
  pull(label_hierarchy) %>%
  str_split(pattern = " > ") %>%
  map(function(x)
    return(head(x, -1)))
# add names to thins list
names(taxonomies) <- OTU_meta %>% pull(label_name)
# maximum number of levels
taxonomies %>%
  map(length) %>%
  unlist() %>%
  max() -> max_levels

taxonomies %<>% map(function(x)
  c(x, rep("OTU", max_levels - length(x))) %>%
    t() %>% as_tibble(.name_repair = "unique")) %>%
  map_df(bind_rows)
# change the names in the table
names(taxonomies) <- paste0("level", 1:max_levels)
# add it to the OTU meta table
OTU_meta <- OTU_meta %>% bind_cols(taxonomies)
# export
paste0(res.dir, "/", project_name, "_OTU_meta.csv") %>%
  write_csv(OTU_meta, path = .)
```

Add the taxonomy form Biigle

If that table exist that is but, if not, skip that chunk

```r
OTU_meta %>% filter(level1 == "Biota") %>%

left_join(Biigle_catalogue, by = c("label_name" = "name")) %>%
  split(.$label_name) %>%
```

```
  # change parent_id the actual name of the taxa
  map(function(x)
    mutate(x, parent_name =
             Biigle_catalogue %>%
             filter(id == pull(
               distinct(x, parent_id), 1
             )) %>%
             pull(name))) %>% map_df(bind_rows) %>%
  arrange(desc(Tot_abundance)) -> OTU_meta

# export it again. It should overwite the previous version
paste0(res.dir, "/", project_name, "_OTU_meta.csv") %>%
  write_csv(OTU_meta, path = .)
```

Finally, export a table of the metadata associated to each image

```
D %>% select( "image_id","filename","image_longitude","image_latitude") %>% distinct() -> M
ETA
```

Basic ecological results exploration

All non-living annotations have to be removed.

Benthic ecologists have the habit of removing the most mobile taxa, as they may be a source of biases. That include fish and sharks.

Unknown annotations should also be removed

```
image_counts <- image_counts  %>%
  # no minerals or Unknown
  filter(!str_detect(label_hierarchy, "Natural structure")) %>%
  # no lebenspurren either
  filter(!str_detect(label_hierarchy, "Liebenspurren")) %>%
  # no "NEW" OTU (can be many things)
  filter(!str_detect(label_hierarchy, "NEW$")) %>% # here the '$' means finishes by
  # no fish
  filter(!str_detect(label_hierarchy, "Pisces"))  %>%
  # no unknown stuff
  filter(!str_detect(label_hierarchy, "Unknown"))


# make it an image*species matrix
image_counts %>%
  select(filename , label_name, count) %>%
  spread(key = label_name, value = count, fill = 0) -> Bio

# now export the table
paste0(res.dir, "/", project_name, "_Bio.csv") %>%
  write_csv(Bio, path = .)

# make a vector of OTUs - updated since you have removed many
image_counts %>% distinct(label_name) %>% pull -> OTUs

# add the name and other metadata avaialble to each image
inner_join(Bio, META, by = "filename") -> BioMeta

# now export the table
paste0(res.dir, "/", project_name, "_BioMeta.csv") %>%
  write_csv(BioMeta, path = .)
```

Create a folder where the plots will be exported

```
# make a folder with the results
plot.dir <- paste0(res.dir,"/","plots")
# if it doesnt exists, create a folder for your projects results
if(dir.exists(plot.dir) == FALSE) {
  print("creating folder")
  dir.create(plot.dir)
  }
```

Make plots of the diversity of phyla within the volume

```
# number of phyla
OTU_meta %>%
  group_by(phylum) %>%
  summarise(
    phylum_abundance.raw = sum(Tot_abundance),
    phylum_diversity.raw = n_distinct(label_name),
    phylum_abundance.avg = mean(Tot_abundance)
  ) %>%
  arrange(desc(phylum_abundance.raw)) %>%
  mutate(phylum_cumSum = cumsum(phylum_abundance.raw)) -> Phylum_meta


# Draw a plot of abundance per phylum and colour each phylum bar per number of OTU in that
phylum
OTU_meta %>%
  ggplot(aes(x = phylum, y = Tot_abundance, fill = label_name)) +
  geom_col(show.legend = FALSE) +
  theme(
    panel.background = element_rect(
      fill = "snow1",
      colour = "NA",
      size = 0.5,
      linetype = "solid"
    ),
    axis.text.x = element_text(angle = 45, hjust = 1),
    legend.position = "none"
  ) +
  ggtitle(label = "Abundance of each phylua and number of OTUs within phylum")

# export plots
ggsave(paste0(plot.dir, "/", "Phylums_", project_name , ".jpeg"), dpi = 500)
```

Multivariate analysis

Conduct a simple Principal coordinates analysis to see if there is any obvious pattern or structure in the volume.

```
Bio %>% select(-filename) -> d

# try a bray curtis capscale because it is fast
capscale(d~1,distance = "bray") -> dbCCA
limited <- ordiselect( d, dbCCA, fitlim = 0.05 )

jpeg(filename = paste0(plot.dir,"/"," Multivariate analysis ",project_name ,".jpeg"))
plot( dbCCA, type = "n",  main= "PcoA plot - with abundance",sub = "main driving species on
ly")
# add arrows of species effects
arrows(x0 = 0,y0 = 0,
       x1 = data.frame(scores(dbCCA)$species)[limited,"MDS1"] ,
       y1 = data.frame(scores(dbCCA)$species)[limited,"MDS2"],
       length = 0.01, lty = 2, col= "coral")
```

```
# add sites points
points(dbCCA,display = c("sites"), pch = 21, bg ="yellow",  col ="red4")

# add species names
points(dbCCA, display="species",
       select = limited, pch=3, col="red4", cex=0.7)
ordipointlabel(dbCCA, display="species",
               select = limited, col="red4", cex=0.8, add = TRUE)
# pimp it more
points(0,0, pch = 21, col="red4",bg= 1, cex = 2)
abline(h = 0, v= 0, lty = 3, col="grey")
dev.off()

## png
##   2
```

Plot the abundance of each OTU along the transect

Make a plot for each OTU and export it in that folder

Make a species accumulation plot by number of individuals

```
# format Bio table
tL <- Bio %>% select(-filename) %>% t

# calculate rarefaction
tL %>% rowSums() %>%
  iNEXT( datatype = "abundance",
         endpoint = ( sum(.)*3) , # extrapolate to 3 times the number of samples
         q=c(0)) -> sac
# plot the curve
  ggiNEXT(sac, type = 1) +
    theme_bw()+
    ggtitle("Rarefaction curve") +
  labs(x = "Individuals", y = "Richness Estimator") +
  # line of the maximum estimator
  geom_hline(yintercept=sac$AsyEst[1,"Estimator"], linetype="dashed",
                color = "red", size=0.8) +
    geom_hline(yintercept=sac$AsyEst[1,"Observed"], linetype="solid",
                color = "red", size=0.8)

# export plot
     ggsave(filename = paste0(plot.dir,"/"," Rarefaction curve in ",project_name ,".jpeg"),
dpi = 500)

## Saving 5 x 4 in image
```

Table of diversity indices

```
tL %>% rowSums() %>%
  ChaoRichness( datatype = "abundance") %>%
  mutate(transect = row.names(.)) %>% as_tibble() %>%
  mutate(   E = "Chao", order = "0")  -> ChaoRichnessEstimate
tL %>% rowSums() %>%
  ChaoShannon( datatype = "abundance") %>%
  mutate(transect = row.names(.)) %>% as_tibble() %>%
  mutate(Est_s.e. =  Est_s.e, E = "Shannon", order = "1") %>%
  select(-Est_s.e) -> ShannonEstimate
tL %>% rowSums() %>%
  ChaoSimpson( datatype = "abundance") %>%
  mutate(transect = row.names(.)) %>% as_tibble() %>%
```

```
  mutate(E = "Simpson", order = "2") -> SimpsonEstimate

# combine them

list(ChaoRichnessEstimate, ShannonEstimate, SimpsonEstimate) %>% map_df(bind_rows) -> EstD_
all

# export the table
          paste0(res.dir,"/",project_name,"_diversity.csv") %>%
                    write_csv(EstD_all,path = . )
```

7.2 Appendix A4:

Appendix accompanying chapter 4

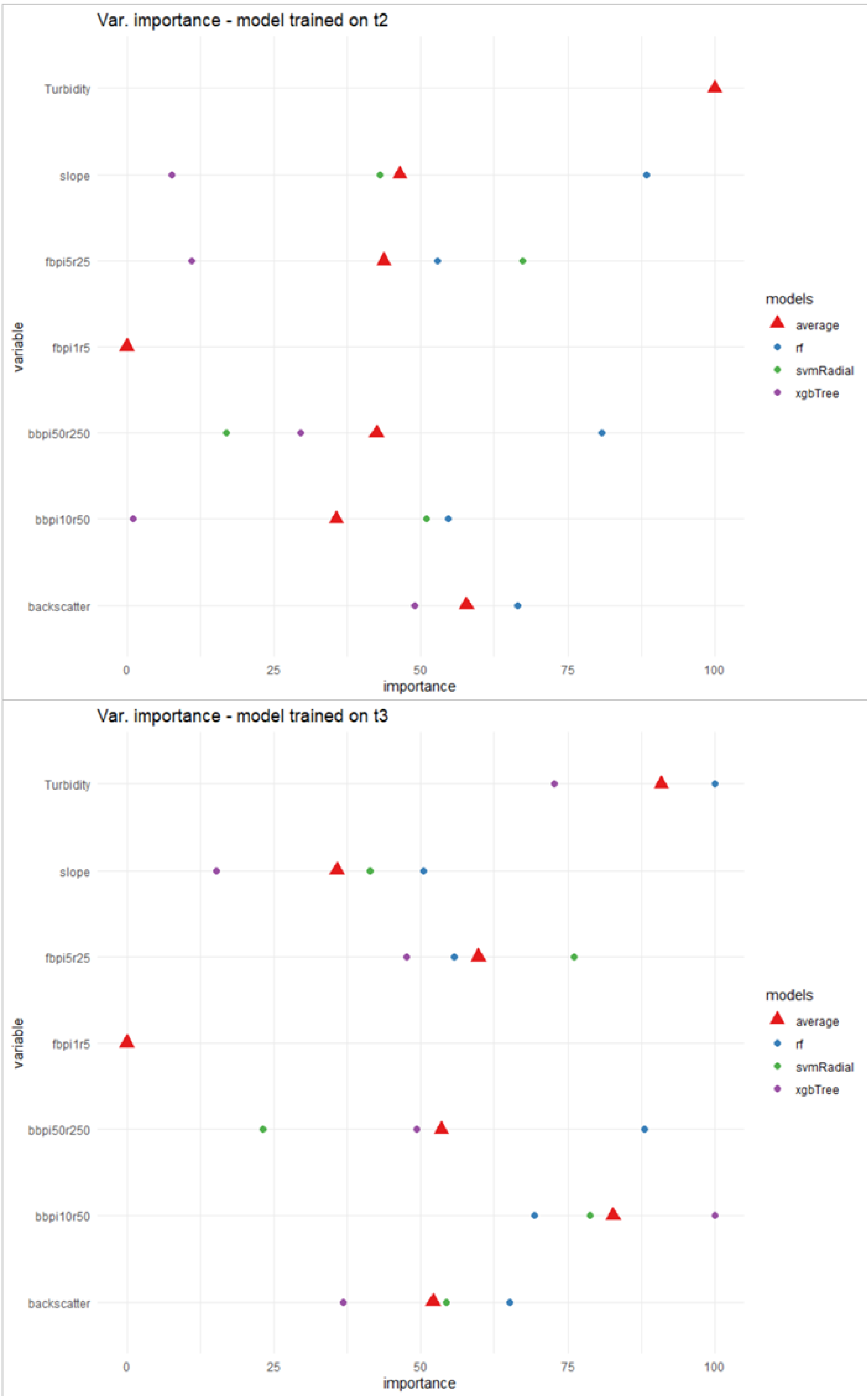


Figure A4-1: Variable importance plots for models trained on t2 and t3

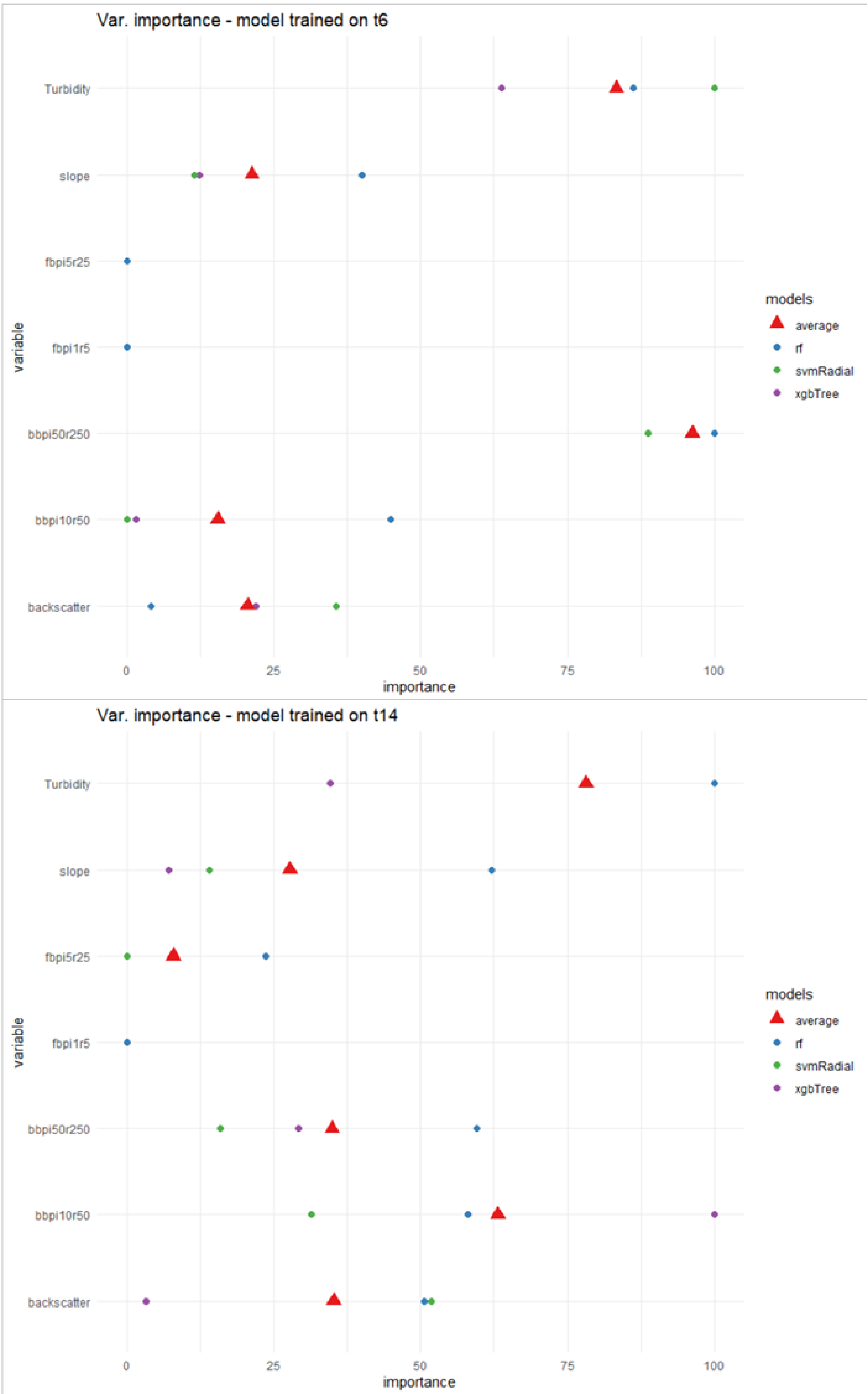


Figure A4-1: Variable importance plots for models trained on t2 and t3

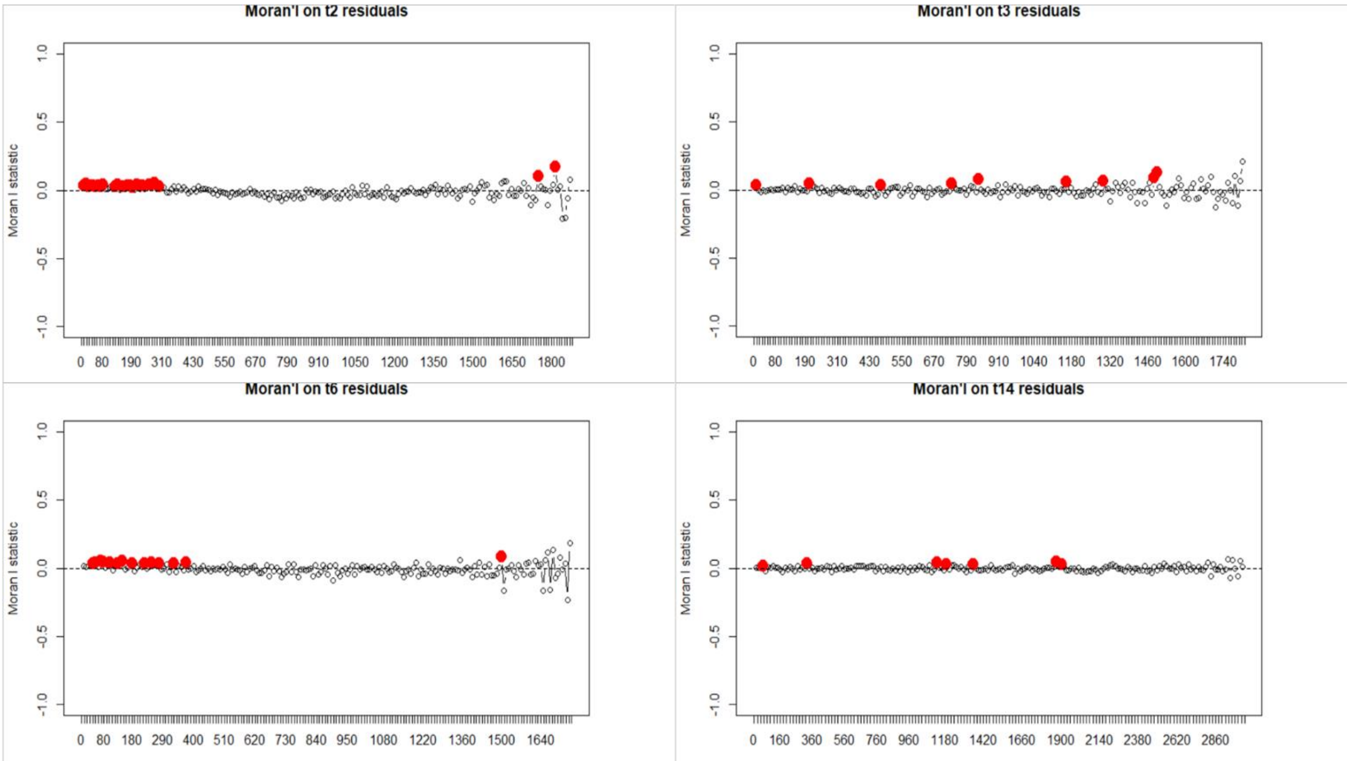


Figure A4-2: Moran's I correlograms on residuals of models trained on each individual transects. Red dots indicate the significant distant classes

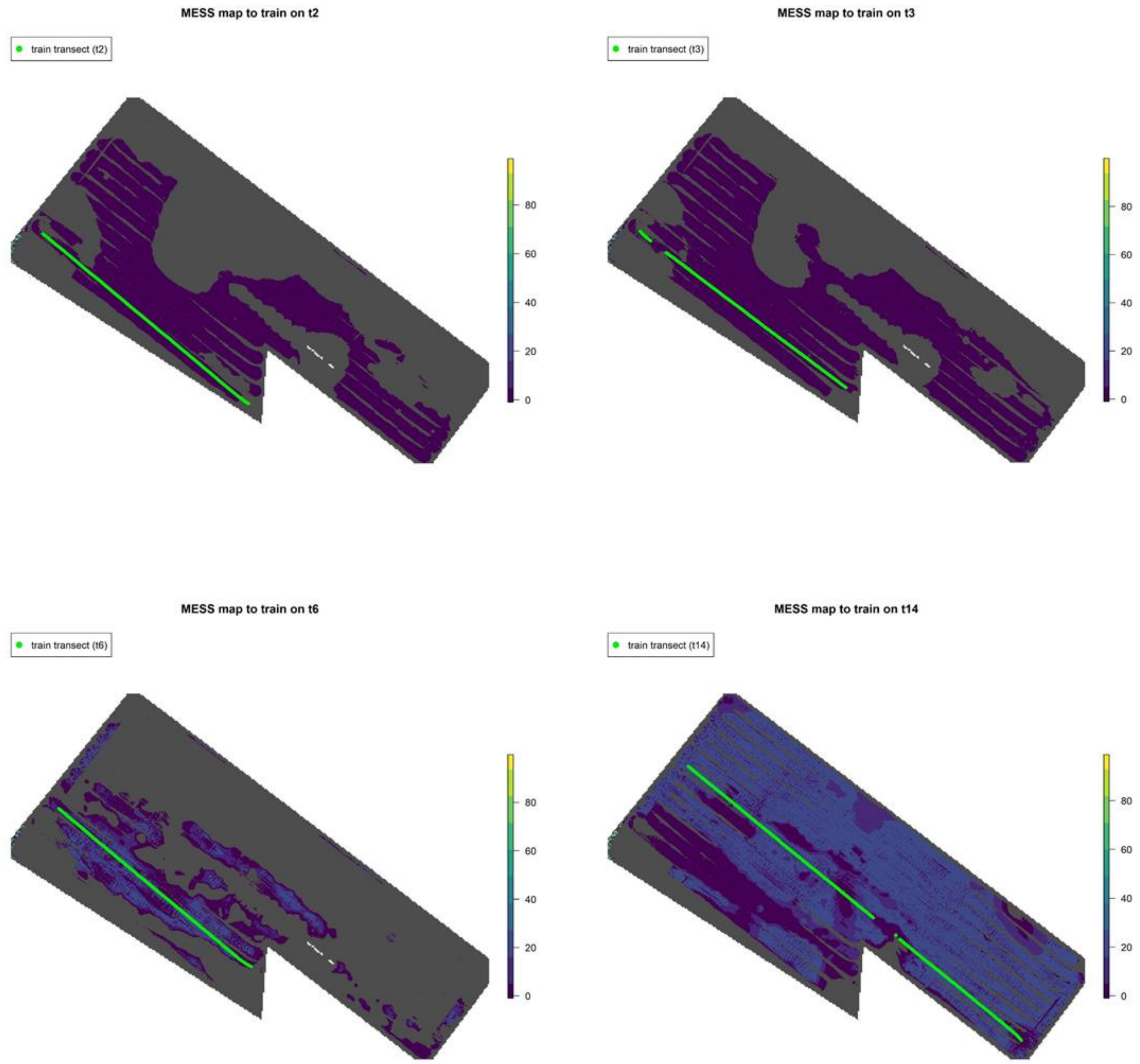


Figure A4-3: MESS maps obtained with the data in each individual transects . The data points in the transect are plotted in green. The Scale indicated the similarity index of the MESS map.

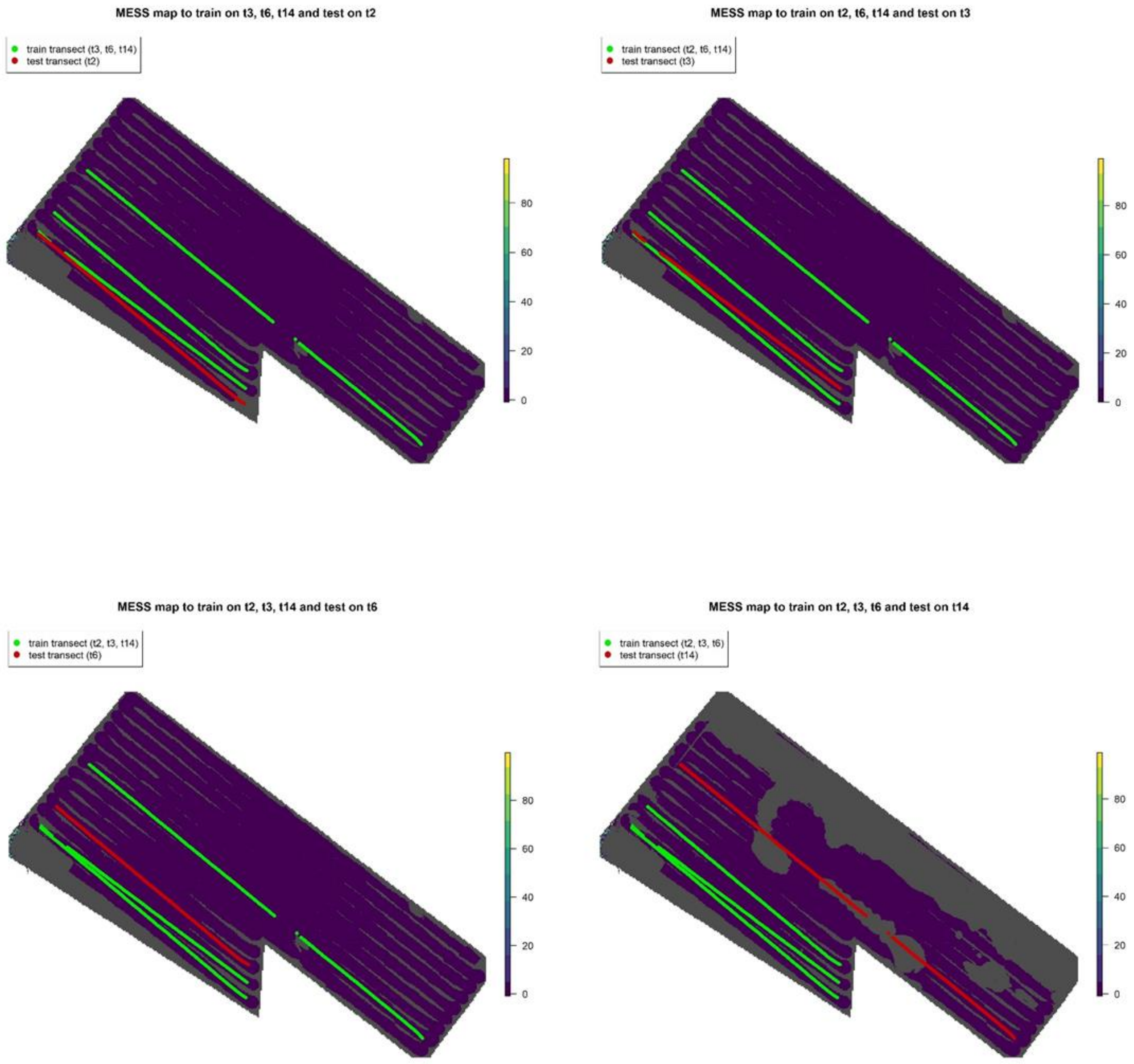


Figure A4-4: MESS maps of training sets of 3 transects (plotted in green). The remaining transect used for testing is (plotted in red). The Scale indicated the similarity index of the MESS map.

7.3 Appendix A5:

Appendix accompanying **chapter 5**

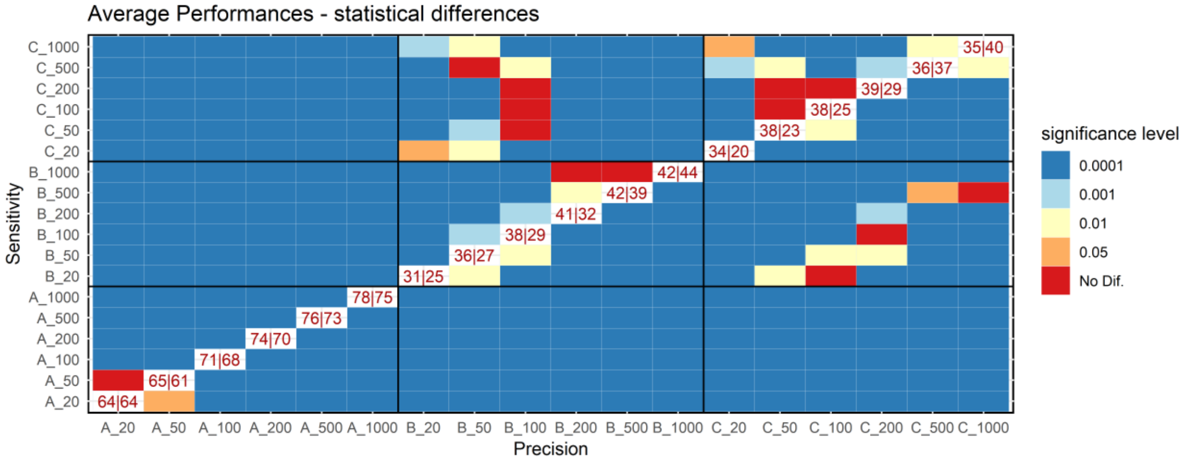


Figure A5-1: Pair-wise permutation-based analysis of variance of differences in sensitivity (upper left triangle of the matrix) and precision (lower right triangle of the matrix) between each treatment. The numbers in central cells indicates sensitivity (left) and precision (right) of corresponding treatments on the axis. Significance level indicate at which alpha threshold the two treatments are significantly different in percentages of maximal value (i.e. 1). No dif. indicates a p-value above 0.05.

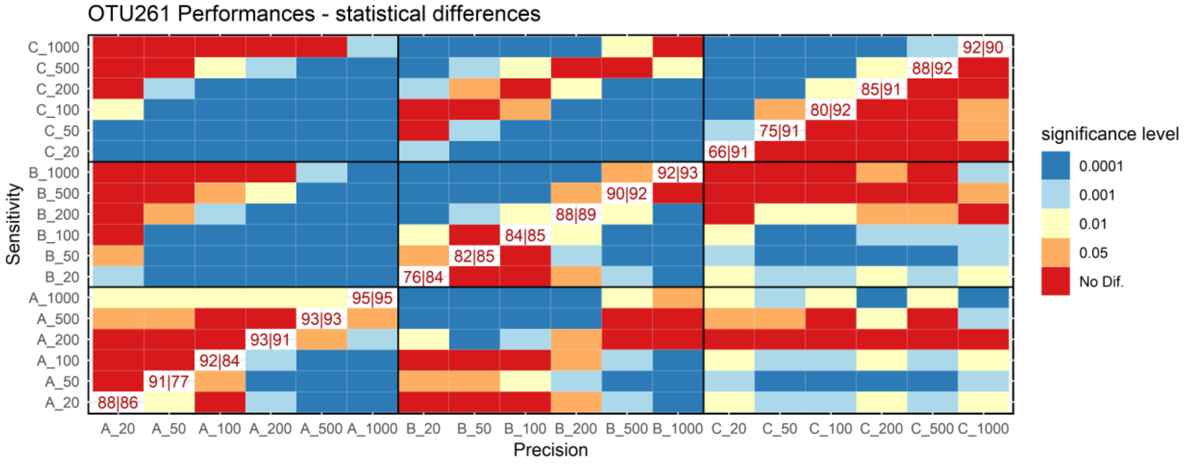


Figure A5-2: Pair-wise permutation-based analysis of variance of differences in sensitivity (upper left triangle of the matrix) and precision (lower right triangle of the matrix) between each treatment for OTU 261. The numbers in central cells indicates sensitivity (left) and precision (right) of corresponding treatments on the axis in percentages of maximal value (i.e. 1). Significance level indicate at which alpha threshold the two treatments are significantly different. No dif. indicates a p-value above 0.05.

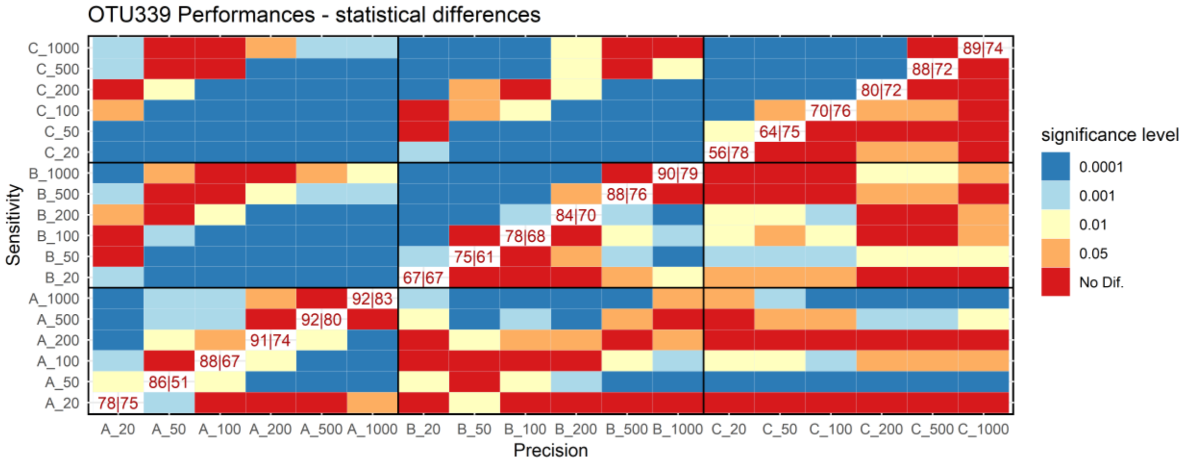


Figure A5-3: Pair-wise permutation-based analysis of variance of differences in sensitivity (upper left triangle of the matrix) and precision (lower right triangle of the matrix) between each treatment for OTU 339. The numbers in central cells indicates sensitivity (left) and precision (right) of corresponding treatments on the axis in percentages of maximal value (i.e. 1). Significance level indicate at which alpha threshold the two treatments are significantly different. No dif. indicates a p-value above 0.05.

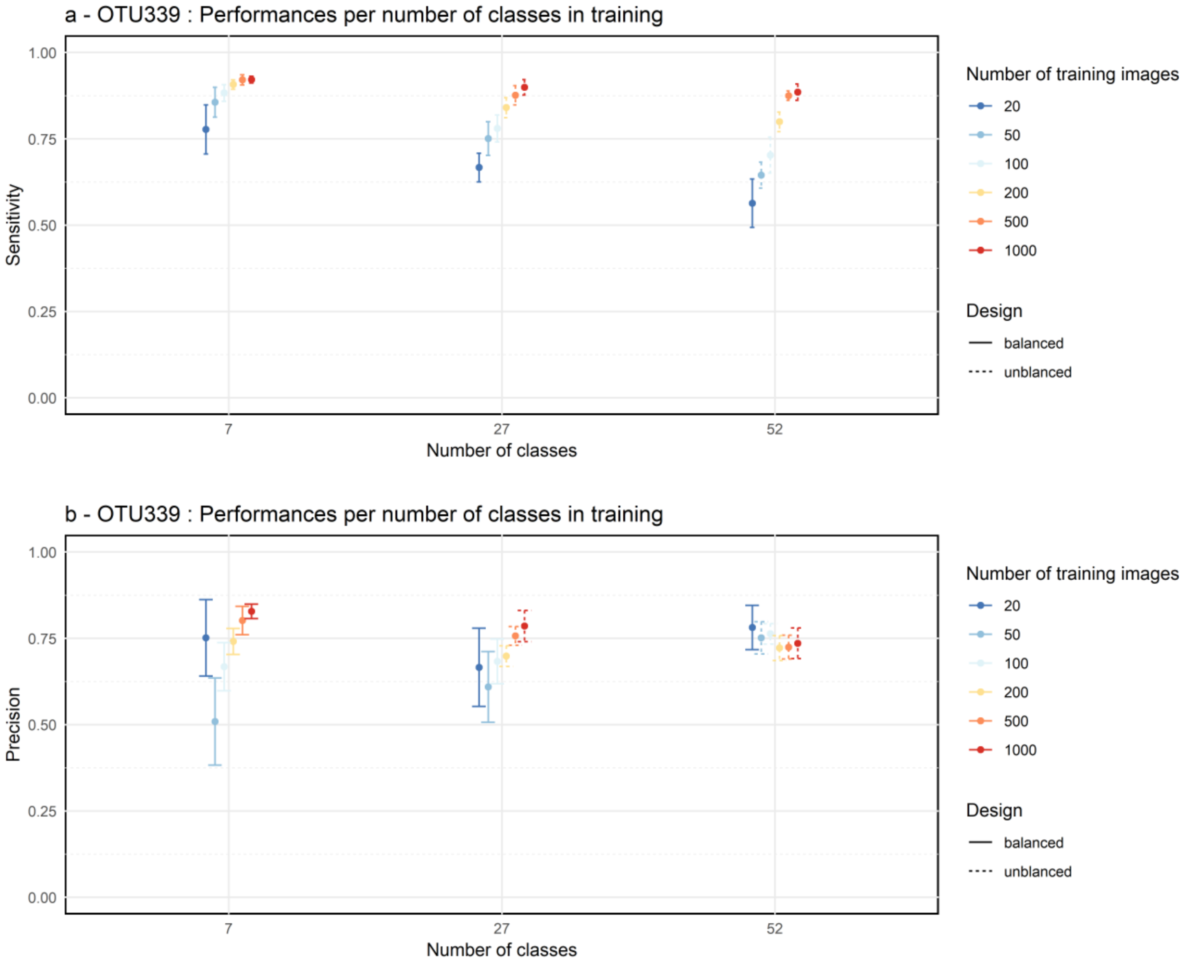


Figure A5-4 a) Differences in sensitivity for OTU 339 in classifiers trained with different number of classes and images (7 for group A, 27 for group B and 52 for group C). Error bars are standard deviation calculated from the 10 random splits. b) Differences in precision for OTU 339 in classifier trained with different number of classes and images. Error bars are standard deviation of the 10 random splits.

List of Acronyms

ADCP: Acoustic Doppler Current Profiler

AI: Artificial Intelligence

API: Application programming interface

AUV: Autonomous Underwater Vehicle

CNN: Convolutional Neural Network

CTD: Conductivity, Temperature, and Depth

CV: Computer Vision (not to be confused with cross validation)

DL: Deep Learning

GIS: Geographic Information System

HOV: Human Occupied Vehicle

MAIA: Machine learning Assisted Image Annotation

MB: Multibeam

MESS: Multivariate environmental similarity surface

nMDS: non-metric Multi-Dimensional Scaling

NN: Neural Network

OTU: Operational Taxonomical Unit

POC: Particulate Organic Carbon

RMSE: Root Mean Square Error

ROV: Remotely operated Vehicle

SAC: Spatial AutoCorrelation

SDM: Species Distribution Model

TF: Tensorflow

UAV: Unmanned Aerial Vehicle

USBL: Ultra-Short Baseline

VME: Vulnerable Marine Ecosystem

WoRMS: World Register of Marine Species

Reference list

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G. S., Davis, A., Dean, J. and Devin, M. (2016) 'Tensorflow: Large-scale machine learning on heterogeneous distributed systems', *arXiv preprint arXiv:1603.04467*.

Aguzzi, J. and Company, J. B. (2010) 'Chronobiology of deep-water decapod crustaceans on continental margins', *Advances in marine biology*: Elsevier, pp. 155-225.

Aguzzi, J., Costa, C., Fujiwara, Y., Iwase, R., Ramirez-Llorda, E. and Menesatti, P. (2009) 'A novel morphometry-based protocol of automated video-image analysis for species recognition and activity rhythms monitoring in deep-sea fauna', *Sensors,* 9(11), pp. 8438-8455.

Aguzzi, J., Costa, C., Ketmaier, V., Angelini, C., Antonucci, F., Menesatti, P. and Company, J. B. (2013) 'Light-dependent genetic and phenotypic differences in the squat lobster Munida tenuimana (Crustacea: Decapoda) along deep continental margins', *Progress in Oceanography,* 118, pp. 199-209.

Althaus, F., Hill, N., Ferrari, R., Edwards, L., Przeslawski, R., Schönberg, C. H., Stuart-Smith, R., Barrett, N., Edgar, G. and Colquhoun, J. (2015) 'A standardised vocabulary for identifying benthic biota and substrata from underwater imagery: the CATAMI classification scheme', *PloS one,* 10(10), pp. e0141039.

Amon, D. J., Ziegler, A. F., Dahlgren, T. G., Glover, A. G., Goineau, A., Gooday, A. J., Wiklund, H. and Smith, C. R. (2016) 'Insights into the abundance and diversity of abyssal megafauna in a polymetallic-nodule region in the eastern Clarion-Clipperton Zone', *Scientific Reports,* 6, pp. 30492.

Anderson, K. and Gaston, K. J. (2013) 'Lightweight unmanned aerial vehicles will revolutionize spatial ecology', *Frontiers in Ecology and the Environment,* 11(3), pp. 138-146.

Anderson, O. F., Guinotte, J. M., Rowden, A. A., Clark, M. R., Mormede, S., Davies, A. J. and Bowden, D. A. (2016) 'Field validation of habitat suitability models for vulnerable marine ecosystems in the South Pacific Ocean: implications for the use of broad-scale models in fisheries management', *Ocean & Coastal Management,* 120, pp. 110-126.

Anselin, L. (1995) 'Local indicators of spatial association—LISA', *Geographical analysis,* 27(2), pp. 93-115.

Araújo, M. B. and New, M. (2007) 'Ensemble forecasting of species distributions', *Trends in ecology & evolution,* 22(1), pp. 42-47.

Ardron, J. A., Simon-Lledó, E., Jones, D. O. B. and Ruhl, H. A. (2019) 'Detecting the Effects of Deep-Seabed Nodule Mining: Simulations Using Megafaunal Data From the Clarion-Clipperton Zone', *Frontiers in Marine Science,* 6(604).

Armstrong, R. A. (2016) 'Landscape-level imaging of benthic environments in optically-deep waters', *Seafloor Mapping along Continental Shelves*: Springer, pp. 261-272.

Armstrong, R. A., Pizarro, O. and Roman, C. (2019) 'Underwater robotic technology for imaging mesophotic coral ecosystems', *Mesophotic Coral Ecosystems*: Springer, pp. 973-988.

Ashford, O. S., Davies, A. J. and Jones, D. O. (2014) 'Deep-sea benthic megafaunal habitat suitability modelling: A global-scale maximum entropy model for xenophyophores', *Deep Sea Research Part I: Oceanographic Research Papers,* 94, pp. 31-44.

Assembly, U. N. G. (2003) 'Oceans and the Law of the Sea. Report of the Secretary General', A/58/65(26668(March)).

Ayma, A., Aguzzi, J., Canals, M., Lastras, G., Bahamon, N., Mecho, A. and Company, J. B. (2016) 'Comparison between ROV video and Agassiz trawl methods for sampling deep water fauna of submarine canyons in the Northwestern Mediterranean Sea with observations on behavioural reactions of target species', *Deep Sea Research Part I: Oceanographic Research Papers,* 114, pp. 149-159.

Balmaseda, M. A., Trenberth, K. E. and Källén, E. (2013) 'Distinctive climate signals in reanalysis of global ocean heat content', *Geophysical Research Letters*.

Ban, N. C., Bax, N. J., Gjerde, K. M., Devillers, R., Dunn, D. C., Dunstan, P. K., Hobday, A. J., Maxwell, S. M., Kaplan, D. M. and Pressey, R. L. (2013) 'Systematic conservation planning: A better recipe for managing the high seas for biodiversity conservation and sustainable use', *Conservation Letters,* 7(1), pp. 41-54.

Barbet-Massin, M., Thuiller, W. and Jiguet, F. (2010) 'How much do we overestimate future local extinction rates when restricting the range of occurrence data in climate suitability models?', *Ecography,* 33(5), pp. 878-886.

Barbier, E. B., Moreno-Mateos, D., Rogers, A. D., Aronson, J., Pendleton, L., Danovaro, R., Henry, L.-A., Morato, T., Ardron, J. and Van Dover, C. L. (2014) 'Protect the deep sea', *Nature,* 505(7484), pp. 475-477.

Barré, P., Stöver, B. C., Müller, K. F. and Steinhage, V. (2017) 'LeafNet: A computer vision system for automatic plant species identification', *Ecological Informatics,* 40, pp. 50-56.

Barrell, J. and Grant, J. (2013) 'Detecting hot and cold spots in a seagrass landscape using local indicators of spatial association', *Landscape ecology,* 28(10), pp. 2005-2018.

Bayley, D. T. I. and Mogg, A. O. M. (2019) 'Chapter 6 - New Advances in Benthic Monitoring Technology and Methodology', in Sheppard, C. (ed.) *World Seas: an Environmental Evaluation (Second Edition)*: Academic Press, pp. 121-132.

Beazley, L. I., Kenchington, E. L., Murillo, F. J. and del Mar Sacau, M. (2013) 'Deep-sea sponge grounds enhance diversity and abundance of epibenthic megafauna in the Northwest Atlantic', *ICES Journal of Marine Science: Journal du Conseil,* 70(7), pp. 1471-1490.

Begon, M., Townsend, C. R. and Harper, J. L. (2006) *Ecology: from individuals to ecosystems.*

Beijbom, O., Edmunds, P. J., Kline, D. I., Mitchell, B. G. and Kriegman, D. 'Automated annotation of coral reef survey images'. *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*: IEEE, 1170-1177.

Beijbom, O., Edmunds, P. J., Roelfsema, C., Smith, J., Kline, D. I., Neal, B. P., Dunlap, M. J., Moriarty, V., Fan, T.-Y. and Tan, C.-J. (2015) 'Towards automated annotation of benthic survey images: Variability of human experts and operational modes of automation', *PloS one,* 10(7), pp. e0130312.

Bell, D. M. and Schlaepfer, D. R. (2016) 'On the dangers of model complexity without ecological justification in species distribution modeling', *Ecological Modelling,* 330, pp. 50-59.

Benfield, M. C., Grosjean, P., Culverhouse, P. F., Irigoien, X., Sieracki, M. E., Lopez-Urrutia, A., Dam, H. G., Hu, Q., Davis, C. S. and Hansen, A. (2007) 'RAPID: research on automated plankton identification', *Oceanography,* 20(2), pp. 172-187.

Berk, R. A. (2006) 'An introduction to ensemble methods for data analysis', *Sociological methods & research,* 34(3), pp. 263-295.

Bett, B. J. (2001a) 'UK-Atlantic margin environmental survey: introduction and overview of bathyal ecology', *Continental Shelf Research,* 21, pp. 917-956.

Bewley, M., Nourani-Vatani, N., Rao, D., Douillard, B., Pizarro, O. and Williams, S. B. 'Hierarchical classification in AUV imagery'. *Field and service robotics*: Springer, 3-16.

Bicknell, A. W., Godley, B. J., Sheehan, E. V., Votier, S. C. and Witt, M. J. (2016) 'Camera technology for monitoring marine biodiversity and human impact', *Frontiers in Ecology and the Environment,* 14(8), pp. 424-432.

Bolnick, D. I. and Fitzpatrick, B. M. (2007) 'Sympatric speciation: models and empirical evidence', *Annu. Rev. Ecol. Evol. Syst.,* 38, pp. 459-487.

Borja, A., Elliott, M., Snelgrove, P. V. R., Austen, M. C., Berg, T., Cochrane, S., Carstensen, J., Danovaro, R., Greenstreet, S., Heiskanen, A.-S., Lynam, C. P., Mea, M., Newton, A., Patrício, J., Uusitalo, L., Uyarra, M. C. and Wilson, C. (2016) 'Bridging the Gap between Policy and Science in Assessing the Health Status of Marine Ecosystems', *Frontiers in Marine Science,* 3(175).

Borregaard, M. K. and Hart, E. M. (2016) 'Towards a more reproducible ecology', *Ecography,* 39(4), pp. 349-353.

Bowden, D. A., Rowden, A. A., Leduc, D., Beaumont, J. and Clark, M. R. (2016) 'Deep-sea seabed habitats: Do they support distinct mega-epifaunal communities that have different vulnerabilities to anthropogenic disturbance?', *Deep Sea Research Part I: Oceanographic Research Papers,* 107(Supplement C), pp. 31-47.

Brandt, A., Gutt, J., Hildebrandt, M., Pawlowski, J., Schwendner, J., Soltwedel, T. and Thomsen, L. (2016) 'Cutting the Umbilical: New Technological Perspectives in Benthic Deep-Sea Research', *Journal of Marine Science and Engineering,* 4(2), pp. 36.

Brattegard, T. and Fosså, J. (1991) 'Replicability of an epibenthic sampler', *Journal of the Marine Biological Association of the United Kingdom,* 71(01), pp. 153-166.

Brind'Amour, A., Mahévas, S., Legendre, P. and Bellanger, L. (2018) 'Application of Moran Eigenvector Maps (MEM) to irregular sampling designs', *Spatial Statistics,* 26, pp. 56-68.

Brind'Amour, A., Menot, L., Galéron, J. and Crassous, P. (2009) 'Spatial organization of a sedimentary macrobenthic community located on the West African Equatorial margin', *Deep Sea Research Part II: Topical Studies in Oceanography,* 56(23), pp. 2292-2298.

Brown, A. M., Warton, D. I., Andrew, N. R., Binns, M., Cassis, G. and Gibb, H. (2014) 'The fourth-corner solution – using predictive models to understand how species traits interact with the environment', *Methods in Ecology and Evolution,* 5(4), pp. 344-352.

Brown, C. J., Smith, S. J., Lawton, P. and Anderson, J. T. (2011) 'Benthic habitat mapping: a review of progress towards improved understanding of the spatial ecology of the seafloor using acoustic techniques', *Estuarine, Coastal and Shelf Science,* 92(3), pp. 502-520.

Bryson, M., Johnson-Roberson, M., Pizarro, O. and Williams, S. 'Automated registration for multi-year robotic surveys of marine benthic habitats'. *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems*, 3-7 Nov. 2013, 3344-3349.

Buhl-Mortensen, L., Buhl-Mortensen, P., Dolan, M. and Gonzalez-Mirelis, G. (2015) 'Habitat mapping as a tool for conservation and sustainable use of marine resources: Some perspectives from the MAREANO Programme, Norway', *Journal of Sea Research,* 100, pp. 46-61.

Buhl-Mortensen, L., Vanreusel, A., Gooday, A. J., Levin, L. A., Priede, I. G., Buhl-Mortensen, P., Gheerardyn, H., King, N. J. and Raes, M. (2010) 'Biological structures as a source of habitat heterogeneity and biodiversity on the deep ocean margins', *Marine Ecology,* 31(1), pp. 21-50.

Bullimore, R. D., Foster, N. L. and Howell, K. L. (2013) 'Coral-characterized benthic assemblages of the deep Northeast Atlantic: defining "Coral Gardens" to support future habitat mapping efforts', *Ices Journal of Marine Science,* 70(3), pp. 511-522.

Camps-Valls, G., Hickler, T. and König-Ries, B. 'Computer Science Meets Ecology (Dagstuhl Seminar 17091)'. *Dagstuhl Reports*: Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik.

Carney, R., Atkinson, R. and Gordon, J. (2005) 'Zonation of deep biota on continental margins', *Oceanography and marine biology: an annual review,* 43, pp. 211-78.

Carney, R. S. (1997) 'Basing conservation policies for the deep-sea floor on current-diversity concepts: a consideration of rarity', *Biodiversity & Conservation,* 6(11), pp. 1463-1485.

Cartes, J. E. and Sarda, F. (1992) 'Abundance and diversity of decapod crustaceans in the deep-catalan sea (Western mediterranean)', *Journal of Natural History,* 26(6), pp. 1305-1323.

Chaniotis, P. D., Robson, L. M., Lemasson, A. J., Cornthwaite, A. L. and Howell, K. L. (2020) 'UK deep-sea conservation: Progress, lessons learned, and actions for the future', *Aquatic Conservation: Marine and Freshwater Ecosystems,* 30(2), pp. 375-393.

Chao, A. and Chiu, C.-H. (2016) 'Species richness: estimation and comparison', *Wiley StatsRef: Statistics Reference Online,* 1, pp. 26.

Chao, A., Gotelli, N. J., Hsieh, T. C., Sander, E. L., Ma, K. H., Colwell, R. K. and Ellison, A. M. (2014) 'Rarefaction and extrapolation with Hill numbers: a framework for sampling and estimation in species diversity studies', *Ecological Monographs,* 84(1), pp. 45-67.

Chapman, A. S. A., Tunnicliffe, V. and Bates, A. E. (2018) 'Both rare and common species make unique contributions to functional diversity in an ecosystem unaffected by human activities', *Diversity and Distributions,* 24(5), pp. 568-578.

Chapman, M. G. and Underwood, A. J. (2008) 'Scales of variation of gastropod densities over multiple spatial scales: comparison of common and rare species', *Marine Ecology Progress Series,* 354, pp. 147-160.

Cheng, G. and Han, J. (2016) 'A survey on object detection in optical remote sensing images', *ISPRS Journal of Photogrammetry and Remote Sensing,* 117, pp. 11-28.

Chimienti, G., Angeletti, L., Rizzo, L., Tursi, A. and Mastrototaro, F. (2018) 'ROV vs trawling approaches in the study of benthic communities: the case of Pennatula rubra (Cnidaria: Pennatulacea)', *Journal of the Marine Biological Association of the United Kingdom,* 98(8), pp. 1859-1869.

Cho, J., Lee, K., Shin, E., Choy, G. and Do, S. (2015) 'How much data is needed to train a medical image deep learning system to achieve necessary high accuracy?', *arXiv preprint arXiv:1511.06348*.

Clark, M. R., Althaus, F., Schlacher, T. A., Williams, A., Bowden, D. A. and Rowden, A. A. (2016a) 'The impacts of deep-sea fisheries on benthic communities: a review', *ICES Journal of Marine Science: Journal du Conseil,* 73(suppl 1), pp. i51-i69.

Clark, M. R., Consalvey, M. and Rowden, A. A. (2016b) *Biological Sampling in the Deep Sea.* John Wiley & Sons.

Cline, D. E., Edgington, D. R. and Mariette, J. (2007) 'An Automated Visual Event Detection System for Cabled Observatory Video'. *OCEANS 2007*, 29 Sept.-4 Oct. 2007, 1-5.

Collins, P., Kennedy, B., Copley, J., Boschen, R., Fleming, N., Forde, J., Ju, S.-J., Lindsay, D., Marsh, L. and Nye, V. (2013) 'VentBase: Developing a consensus among stakeholders in the deep-sea regarding environmental impact assessment for deep-sea mining–A workshop report', *Marine Policy,* 42, pp. 334-336.

Colwell, R. K., Chao, A., Gotelli, N. J., Lin, S.-Y., Mao, C. X., Chazdon, R. L. and Longino, J. T. (2012) 'Models and estimators linking individual-based and sample-based rarefaction, extrapolation and comparison of assemblages', *Journal of plant ecology,* 5(1), pp. 3-21.

Connolly, S. R., MacNeil, M. A., Caley, M. J., Knowlton, N., Cripps, E., Hisano, M., Thibaut, L. M., Bhattacharya, B. D., Benedetti-Cecchi, L., Brainard, R. E., Brandt, A., Bulleri, F., Ellingsen, K. E., Kaiser, S., Kröncke, I., Linse, K., Maggi, E., O'Hara, T. D., Plaisance, L., Poore, G. C. B., Sarkar, S. K., Satpathy, K. K., Schückel, U., Williams, A. and Wilson, R. S. (2014) 'Commonness and rarity in the marine biosphere', *Proceedings of the National Academy of Sciences of the United States of America,* 111(23), pp. 8524-8529.

Cordes, E. E., Jones, D. O., Schlacher, T. A., Amon, D. J., Bernardino, A. F., Brooke, S., Carney, R., DeLeo, D. M., Dunlop, K. M. and Escobar-Briones, E. G. (2016) 'Environmental impacts of the deep-water oil and gas industry: a review to guide management strategies', *Frontiers in Environmental Science,* 4, pp. 58.

Costello, M. J., Bouchet, P., Boxshall, G., Fauchald, K., Gordon, D., Hoeksema, B. W., Poore, G. C., van Soest, R. W., Stöhr, S. and Walter, T. C. (2013) 'Global coordination and standardisation in marine biodiversity through the World Register of Marine Species (WoRMS) and related databases', *PloS one,* 8(1), pp. e51629.

Costello, M. J. and Chaudhary, C. (2017) 'Marine Biodiversity, Biogeography, Deep-Sea Gradients, and Conservation', *Current Biology,* 27(11), pp. R511-R527.

Costello, M. J., Coll, M., Danovaro, R., Halpin, P., Ojaveer, H. and Miloslavich, P. (2010) 'A census of marine biodiversity knowledge, resources, and future challenges', *PloS one,* 5(8), pp. e12110.

Courtene-Jones, W., Quinn, B., Gary, S. F., Mogg, A. O. and Narayanaswamy, B. E. (2017) 'Microplastic pollution identified in deep-sea water and ingested by benthic invertebrates in the Rockall Trough, North Atlantic Ocean', *Environmental Pollution,* 231, pp. 271-280.

Crowder, L. and Norse, E. (2008) 'Essential ecological insights for marine ecosystem-based management and marine spatial planning', *Marine Policy,* 32(5), pp. 772-778.

Culverhouse, P. F., Macleod, N., Williams, R., Benfield, M. C., Lopes, R. M. and Picheral, M. (2014) 'An empirical assessment of the consistency of taxonomic identifications', *Marine Biology Research,* 10(1), pp. 73-84.

Culverhouse, P. F., Simpson, R., Ellis, R., Lindley, J., Williams, R., Parisini, T., Reguera, B., Bravo, I., Zoppoli, R. and Earnshaw, G. (1996) 'Automatic classification of field-collected dinoflagellates by artificial neural network', *Marine Ecology Progress Series,* 139, pp. 281-287.

Culverhouse, P. F., Williams, R., Reguera, B., Herry, V. and González-Gil, S. (2003) 'Do experts make mistakes? A comparison of human and machine identification of dinoflagellates', *Marine Ecology Progress Series,* 247, pp. 17-25.

Danovaro, R., Aguzzi, J., Fanelli, E., Billett, D., Gjerde, K., Jamieson, A., Ramirez-Llodra, E., Smith, C. R., Snelgrove, P. V. R., Thomsen, L. and Dover, C. L. V. (2017a) 'An ecosystem-based deep-ocean strategy', *Science,* 355(6324), pp. 452-454.

Danovaro, R., Carugati, L., Berzano, M., Cahill, A. E., Carvalho, S., Chenuil, A., Corinaldesi, C., Cristina, S., David, R., Dell'Anno, A., Dzhembekova, N., Garcés, E., Gasol, J. M., Goela, P., Féral, J.-P., Ferrera, I., Forster, R. M., Kurekin, A. A., Rastelli, E., Marinova, V., Miller, P. I., Moncheva, S., Newton, A., Pearman, J. K., Pitois, S. G., Reñé, A., Rodríguez-Ezpeleta, N., Saggiomo, V., Simis, S. G. H., Stefanova, K., Wilson, C., Lo Martire, M., Greco, S., Cochrane, S. K. J., Mangoni, O. and Borja, A. (2016) 'Implementing and Innovating Marine Monitoring Approaches for Assessing Marine Environmental Status', *Frontiers in Marine Science,* 3(213).

Danovaro, R., Corinaldesi, C., Dell'Anno, A. and Snelgrove, P. V. (2017b) 'The deep-sea under global change', *Current Biology,* 27(11), pp. R461-R465.

Danovaro, R., Gambi, C., Dell'Anno, A., Corinaldesi, C., Fraschetti, S., Vanreusel, A., Vincx, M. and Gooday, A. J. (2008) 'Exponential decline of deep-sea ecosystem functioning linked to benthic biodiversity loss', *Current Biology,* 18(1), pp. 1-8.

Danovaro, R., Snelgrove, P. V. and Tyler, P. (2014) 'Challenging the paradigms of deep-sea ecology', *Trends in ecology & evolution,* 29(8), pp. 465-475.

Davies, A., Narayanaswamy, B., Hughes, D. and Roberts, J. 2006. An introduction to the benthic ecology of the Rockall-Hatton Area (SEA 7). Report for the Department of Trade and Industry.

Davies, J. S. (2012) 'Mapping Deep-Sea Features in UK Waters For Use in Marine Protected Area Network Design'.

Davies, J. S., Stewart, H. A., Narayanaswamy, B. E., Jacobs, C., Spicer, J., Golding, N. and Howell, K. L. (2015) 'Benthic assemblages of the Anton Dohrn Seamount (NE Atlantic): defining deep-sea biotopes to support habitat mapping and management efforts with a focus on vulnerable marine ecosystems', *PloS one,* 10(5), pp. e0124815.

Davies, T. E., Maxwell, S. M., Kaschner, K., Garilao, C. and Ban, N. C. (2017) 'Large marine protected areas represent biodiversity now and under climate change', *Scientific reports,* 7(1), pp. 9569.

Davison, J. J., van Haren, H., Hosegood, P., Piechaud, N. and Howell, K. L. (2019) 'The distribution of deep-sea sponge aggregations (Porifera) in relation to oceanographic processes in the Faroe-Shetland Channel', *Deep Sea Research Part I: Oceanographic Research Papers,* 146, pp. 55-61.

Dawkins, M., Sherrill, L., Fieldhouse, K., Hoogs, A., Richards, B., Zhang, D., Prasad, L., Williams, K., Lauffenburger, N. and Wang, G. 'An open-source platform for underwater image and video analytics'. *2017 IEEE Winter Conference on Applications of Computer Vision (WACV)*: IEEE, 898-906.

Dayton, P. K. and Hessler, R. R. (1972) 'Role of biological disturbance in maintaining diversity in the deep sea', *Deep Sea Research and Oceanographic Abstracts,* 19(3), pp. 199-208.

De'Ath, G. (2007) 'Boosted trees for ecological modeling and prediction', *Ecology,* 88(1), pp. 243-251.

Deane-Mayer, Z. A. and Knowles, J. E. 2016. caretEnsemble: Ensembles of caret models. R package version 2.0. 0.

Defra 2005. Safeguarding Sea Life: The Joint UK Response to the Review of Marine Nature Conservation. DEFRA London.

Diesing, M., Mitchell, P. and Stephens, D. (2016) 'Image-based seabed classification: what can we learn from terrestrial remote sensing?', *ICES Journal of Marine Science,* 73(10), pp. 2425-2441.

Dormann, C. F. (2007) 'Effects of incorporating spatial autocorrelation into the analysis of species distribution data', *Global Ecology and Biogeography,* 16(2), pp. 129-138.

Dunbabin, M. and Marques, L. (2012) 'Robots for environmental monitoring: Significant advancements and applications', *IEEE Robotics & Automation Magazine,* 19(1), pp. 24-39.

Durden, J. M., Bett, B. J., Jones, D. O. B., Huvenne, V. A. I. and Ruhl, H. A. (2015) 'Abyssal hills – hidden source of increased habitat heterogeneity, benthic megafaunal biomass and diversity in the deep sea', *Progress in Oceanography,* 137(Part A), pp. 209-218.

Durden, J. M., Bett, B. J., Schoening, T., Morris, K. J., Nattkemper, T. W. and Ruhl, H. A. (2016a) 'Comparison of image annotation data generated by multiple investigators for benthic ecology', *Marine Ecology Progress Series,* 552, pp. 61-70.

Durden, J. M., Schoening, T., Althaus, F., Friedman, A., Garcia, R., Glover, A. G., Greinert, J., Stout, N. J., Jones, D. O. and Jordt, A. (2016c) 'Perspectives in visual imaging for marine biology and ecology: from acquisition to understanding', *Oceanography and marine biology: an annual review,* 54, pp. 1-72.

Eaton, S., Ellis, C., Genney, D., Thompson, R., Yahr, R. and Haydon, D. T. (2018) 'Adding small species to the big picture: Species distribution modelling in an age of landscape scale conservation', *Biological Conservation,* 217, pp. 251-258.

Eckstein, M. P., Koehler, K., Welbourne, L. E. and Akbas, E. (2017) 'Humans, but Not Deep Neural Networks, Often Miss Giant Targets in Scenes', *Current Biology,* 27(18), pp. 2827-2832.

Edgington, D. R., Cline, D. E., Davis, D., Kerkez, I. and Mariette, J. 'Detecting, tracking and classifying animals in underwater video'. *OCEANS 2006*: IEEE, 1-5.

Elith, J. and Graham, C. H. (2009) 'Do they? How do they? WHY do they differ? On finding reasons for differing performances of species distribution models', *Ecography,* 32(1), pp. 66-77.

Elith, J., Graham, C. H., Anderson, R. P., Dudík, M., Ferrier, S., Guisan, A., Hijmans, R. J., Huettmann, F., Leathwick, J. R., Lehmann, A., Li, J., Lohmann, L. G., Loiselle, B. A., Manion, G., Moritz, C., Nakamura, M., Nakazawa, Y., Overton, J. M., Peterson, A. T., Phillips, S. J., Richardson, K., Scachetti-Pereira, R., Schapire, R. E., Sobero´n, J., Williams, S., Wisz, M. S. and Zimmermann, N. E. (2006) 'Novel methods improve prediction of species' distributions from occurrence data', *Ecography,* 29, pp. 129-151.

Elith, J., Kearney, M. and Phillips, S. (2010) 'The art of modelling range-shifting species', *Methods in Ecology and Evolution,* 1(4), pp. 330-342.

Elith, J. and Leathwick, J. R. (2009) 'Species Distribution Models: Ecological Explanation and Prediction Across Space and Time', *Annual Review of Ecology Evolution and Systematics Annual Review of Ecology Evolution and Systematics*. Palo Alto: Annual Reviews, pp. 677-697.

Enric, J. d. F., David, M. and Foster, P. (2013) 'Predictive Modeling With Big Data: Is Bigger Really Better?', *Big Data,* 1(4), pp. 215-226.

ESRI 2014. ArcMap 10.2. Redlands, USA: ESRI Redlands, CA.

Favret, C. and Sieracki, J. M. (2016) 'Machine vision automated species identification scaled towards production levels', *Systematic Entomology,* 41(1), pp. 133-143.

Ferrier, S. (2002) 'Mapping spatial pattern in biodiversity for regional conservation planning: where to from here?', *Systematic biology,* 51(2), pp. 331-363.

Folkersen, M. V., Fleming, C. M. and Hasan, S. (2018) 'The economic value of the deep sea: A systematic review and meta-analysis', *Marine Policy,* 94, pp. 71-80.

Fortin, M. J. and Dale, M. R. (2009) 'Spatial autocorrelation in ecological studies: a legacy of solutions and myths', *Geographical Analysis,* 41(4), pp. 392-397.

Foster, N. L., Foggo, A. and Howell, K. L. (2013) 'Using Species-Area Relationships to Inform Baseline Conservation Targets for the Deep North East Atlantic', *PLOS ONE,* 8(3), pp. e58941.

Foster, S. D., Hosack, G. R., Hill, N. A., Barrett, N. S. and Lucieer, V. L. (2014) 'Choosing between strategies for designing surveys: autonomous underwater vehicles', *Methods in Ecology and Evolution,* 5(3), pp. 287-297.

Foster, S. D., Hosack, G. R., Lawrence, E., Przeslawski, R., Hedge, P., Caley, M. J., Barrett, N. S., Williams, A., Li, J. and Lynch, T. (2017) 'Spatially balanced designs that incorporate legacy sites', *Methods in Ecology and Evolution,* 8(11), pp. 1433-1442.

Foster, S. D., Monk, J., Lawrence, E., Hayes, K. R., Hosack, G. R. and Przeslawski, R. (2018) 'Statistical Considerations for Monitoring and Sampling', *Field Manuals for Marine Sampling to Monitor Australian Waters,* 1, pp. 23-41.

Franklin, J. (2010) 'Mapping species distributions: spatial inference and prediction', *Cambridge University Press*.

Furlong, M. E., Paxton, D., Stevenson, P., Pebody, M., McPhail, S. D. and Perrett, J. 'Autosub long range: A long range deep diving AUV for ocean monitoring'. *2012 IEEE/OES Autonomous Underwater Vehicles (AUV)*: IEEE, 1-7.

Gafurov, S. A. and Klochkov, E. V. (2015) 'Autonomous unmanned underwater vehicles development tendencies', *Procedia Engineering,* 106, pp. 141-148.

Gage, J. (1986) 'The benthic fauna of the Rockall Trough: regional distribution and bathymetric zonation', *Proceedings of the Royal Society of Edinburgh, Section B: Biological Sciences,* 88, pp. 159-174.

Gage, J. (2002) 'Benthic biodiversity across and along the continental margin: patterns, ecological and historical determinants, and anthropogenic threats', *Ocean margin systems*: Springer, pp. 307-321.

Gage, J. D. and Tyler, P. A. (1991) *Deep-sea biology: a natural history of organisms at the deep-sea floor.* Cambridge University Press.

Gambi, C. and Danovaro, R. (2016) 'Biodiversity and life strategies of deep-sea meiofauna and nematode assemblages in the Whittard Canyon (Celtic margin, NE Atlantic Ocean)', *Deep Sea Research Part I: Oceanographic Research Papers,* 108, pp. 13-22.

Gaspard, G., Kim, D. and Chun, Y. (2019) 'Residual spatial autocorrelation in macroecological and biogeographical modeling: a review', *Journal of Ecology and Environment,* 43(1), pp. 19.

Gaston, K. J. and O'Neill, M. A. (2004) 'Automated species identification: why not?', *Philosophical Transactions of the Royal Society of London B: Biological Sciences,* 359(1444), pp. 655-667.

Gazis, I.-Z., Schoening, T., Alevizos, E. and Greinert, J. (2018) 'Quantitative mapping and predictive modeling of Mn nodules' distribution from hydroacoustic and optical AUV data linked by random forests machine learning', *Biogeosciences,* 15(23), pp. 7347-7377.

German, C. R., Yoerger, D. R., Jakuba, M., Shank, T. M., Langmuir, C. H. and Nakamura, K.-I. (2008) 'Hydrothermal exploration with the autonomous benthic explorer', *Deep Sea Research Part I: Oceanographic Research Papers,* 55(2), pp. 203-219.

Glover, A. G., Gooday, A. J., Bailey, D. M., Billett, D. S. M., Chevaldonne, P., Colaco, A., Copley, J., Cuvelier, D., Desbruyeres, D., Kalogeropoulou, V., Klages, M., Lampadariou, N., Lejeusne, C., Mestre, N. C., Paterson, G. L. J., Perez, T., Ruhl, H., Sarrazin, J., Soltwedel, T., Soto, E. H., Thatje, S., Tselepides, A., Van Gaever, S. and Vanreusel, A. (2010) 'TEMPORAL CHANGE IN DEEP-SEA BENTHIC ECOSYSTEMS: A REVIEW OF THE EVIDENCE FROM RECENT TIME-SERIES STUDIES', *Advances in Marine Biology, Vol 58,* 58, pp. 1-95.

Gobi, A. F. 'Towards generalized benthic species recognition and quantification using computer vision'. *OCEANS 2010 IEEE-Sydney*: IEEE, 1-6.

Goineau, A. and Gooday, A. J. (2019) 'Diversity and spatial patterns of foraminiferal assemblages in the eastern Clarion–Clipperton zone (abyssal eastern equatorial Pacific)', *Deep Sea Research Part I: Oceanographic Research Papers,* 149.

Gomes-Pereira, J. N., Auger, V., Beisiegel, K., Benjamin, R., Bergmann, M., Bowden, D., Buhl-Mortensen, P., De Leo, F. C., Dionísio, G. and Durden, J. M. (2016) 'Current and future trends in marine image annotation software', *Progress in Oceanography,* 149, pp. 106-120.

Gómez, A., Salazar, A. and Vargas, F. (2016) 'Towards automatic wild animal monitoring: identification of animal species in camera-trap images using very deep convolutional neural networks', *Ecological Informatics,* 41, pp. 24-32.

González-Rivero, M., Beijbom, O., Rodriguez-Ramirez, A., Holtrop, T., González-Marrero, Y., Ganase, A., Roelfsema, C., Phinn, S. and Hoegh-Guldberg, O. (2016) 'Scaling up Ecological Measurements of Coral Reefs Using Semi-Automated Field Image Collection and Analysis', *Remote Sensing,* 8(1), pp. 30.

Gooday, A., Bett, B. and Pratt, D. (1993) 'Direct observation of episodic growth in an abyssal xenophyophore (Protista)', *Deep Sea Research Part I: Oceanographic Research Papers,* 40(11-12), pp. 2131-2143.

Gooday, A. J. (1984) 'Records of seep-sea rhizopod tests inhabited by metazoans in the North-East Atlantic', *Sarsia,* 69, pp. 45-53.

Gooday, A. J., Aranda da Silva, A. and Pawlowski, J. (2011) 'Xenophyophores (Rhizaria, Foraminifera) from the Nazaré Canyon (Portuguese margin, NE Atlantic)', *Deep Sea Research Part II: Topical Studies in Oceanography,* 58(23), pp. 2401-2419.

Goral, F. and Schellenberg, J. 2017. Goeveg: functions for Community Data and Ordinations. R package version 0.3. 3.

Gotelli, N. J. and Colwell, R. K. (2011) 'Estimating species richness', *Biological diversity: frontiers in measurement and assessment,* 12, pp. 39-54.

Grasmueck, M., Eberli, G. P., Viggiano, D. A., Correa, T., Rathwell, G. and Luo, J. (2006) 'Autonomous underwater vehicle (AUV) mapping reveals coral mound distribution, morphology, and oceanography in deep water of the Straits of Florida', *Geophysical Research Letters,* 33(23).

Grassle, F. J. and Maciolek, N. J. (1992a) 'Deep-sea species richness: regional and local diversity estimates from quantitative bottom samples', *American naturalist,* 139(2), pp. 313-341.

Grassle, J. F. (1989) 'Species diversity in deep-sea communities', *Trends in Ecology & Evolution,* 4(1), pp. 12-15.

Grassle, J. F. and Sanders, H. L. 'Life histories and the role of disturbance'. *Deep Sea Research and Oceanographic Abstracts*: Elsevier, 643-659.

Gray, J. S. (2002) 'Species richness of marine soft sediments', *Marine Ecology Progress Series,* 244, pp. 285-297.

Gray, J. S., Bjørgesæter, A. and Ugland, K. I. (2005) 'The impact of rare species on natural assemblages', *Journal of Animal Ecology,* 74(6), pp. 1131-1139.

Guisan, A., Tingley, R., Baumgartner, J. B., Naujokaitis-Lewis, I., Sutcliffe, P. R., Tulloch, A. I., Regan, T. J., Brotons, L., McDonald-Madden, E. and Mantyka-Pringle, C. (2013) 'Predicting species distributions for conservation decisions', *Ecology letters,* 16(12), pp. 1424-1435.

Guisan, A. and Zimmermann, N. E. (2000) 'Predictive habitat distribution models in ecology', *Ecological modelling,* 135(2), pp. 147-186.

Halpern, B. S. (2003) 'The impact of marine reserves: do reserves work and does reserve size matter?', *Ecological applications,* 13(sp1), pp. 117-137.

Halpern, B. S., Lester, S. E. and McLeod, K. L. (2010) 'Placing marine protected areas onto the ecosystem-based management seascape', *Proceedings of the National Academy of Sciences,* 107(43), pp. 18312-18317.

Halpern, B. S., Selkoe, K. A., Micheli, F. and Kappel, C. V. (2007) 'Evaluating and ranking the vulnerability of global marine ecosystems to anthropogenic threats', *Conservation Biology,* 21(5), pp. 1301-1315.

Hampton, S. E., Strasser, C. A., Tewksbury, J. J., Gram, W. K., Budden, A. E., Batcheller, A. L., Duke, C. S. and Porter, J. H. (2013) 'Big data and the future of ecology', *Frontiers in Ecology and the Environment,* 11(3), pp. 156-162.

He, K., Gkioxari, G., Dollár, P. and Girshick, R. 'Mask r-cnn'. *Proceedings of the IEEE international conference on computer vision*, 2961-2969.

Henry, L.-A., Moreno Navas, J. and Roberts, J. (2013) 'Multi-scale interactions between local hydrography, seabed topography, and community assembly on cold-water coral reefs', *Biogeosciences,* 10(4), pp. 2737-2746.

Henry, L.-A. and Roberts, J. M. (2017) 'Global biodiversity in cold-water coral reef ecosystems', *Marine Animal Forests: The Ecology of Benthic Biodiversity Hotspots*, pp. 235-256.

Hernandez, P. A., Graham, C. H., Master, L. L. and Albert, D. L. (2006) 'The effect of sample size and species characteristics on performance of different species distribution modeling methods', *Ecography,* 29(5), pp. 773-785.

Hessler, R. R. and Sanders, H. L. 'Faunal diversity in the deep-sea'. *Deep Sea Research and Oceanographic Abstracts*: Elsevier.

Higgs, N. D. and Attrill, M. (2015) 'Biases in biodiversity: wide-ranging species are discovered first in the deep sea', *Frontiers in Marine Science,* 2, pp. 61.

Hilário, A., Metaxas, A., Gaudron, S. M., Howell, K. L., Mercier, A., Mestre, N. C., Ross, R. E., Thurnherr, A. M. and Young, C. (2015) 'Estimating dispersal distance in the deep sea: challenges and applications to marine reserves', *Frontiers in Marine Science,* 2, pp. 6.

Hobson, B. W., Bellingham, J. G., Kieft, B., McEwen, R., Godin, M. and Zhang, Y. 'Tethys-class long range AUVs - extending the endurance of propeller-driven cruising AUVs from days to weeks'. *2012 IEEE/OES Autonomous Underwater Vehicles (AUV)*, 24-27 Sept. 2012, 1-8.

Hollis, D. J., Edgington, D. and Cline, D. (2016) 'Automated Detection of Deep-Sea Animals', *Digital Commons, http://digitalcommons. calpoly. edu/star/370*.

Howell, K.-L., Piechaud, N., Downie, A.-L. and Kenny, A. (2016a) 'The distribution of deep-sea sponge aggregations in the North Atlantic and implications for their effective spatial management', *Deep Sea Research Part I: Oceanographic Research Papers,* 115, pp. 309-320.

Howell, K., Davies, J. and Narayanaswamy, B. (2010a) 'Identifying deep-sea megafaunal epibenthic assemblages for use in habitat mapping and marine protected area network design', *Journal of the Marine Biological Association of the United Kingdom,* 90(1), pp. 33-68.

Howell, K., Huvenne, V., Piechaud, N., Robert, K. and Ross, R. (2014a) 'Report 528 Analysis of biological data from the JC060 survey of areas of conservation interest in deep waters off north and west Scotland', *JNCC Monitoring and Surveillance Reports*.

Howell, K. L. (2010) 'A benthic classification system to aid in the implementation of marine protected area networks in the deep/high seas of the NE Atlantic', *Biological Conservation,* 143, pp. 1041-1056.

Howell, K. L., Billett, D. S. M. and Tyler, P. A. (2002) 'Depth-related distribution and abundance of seastars (Echinodermata: Asteroidea) in the Porcupine Seabight and Porcupine Abyssal Plain, NE Atlantic', *Deep Sea Research Part I: Oceanographic Research Papers,* 49(10), pp. 1901-1920.

Howell, K. L., Bullimore, R. D. and Foster, N. L. (2014b) 'Quality assurance in the identification of deep-sea taxa from video and image analysis: response to Henry and Roberts', *ICES Journal of Marine Science: Journal du Conseil,* 71(4), pp. 899-906.

Howell, K. L. and Davies, J. S. (2016) 'Deep-sea species image catalogue, On-line version 2', https://deepseacru.org/2016/12/16/deep-sea-species-image-catalogue/.

Howell, K. L., Davies, J. S., Allcock, A. L., Braga-Henriques, A., Buhl-Mortensen, P., Carreiro-Silva, M., Dominguez-Carrió, C., Durden, J. M., Foster, N. L., Game, C. A., Hitchin, B., Horton, T., Hosking, B., Jones, D. O. B., Mah, C., Marchais, C. L., Menot, L., Morato, T., Pearman, T. R. R., Piechaud, N., Ross, R. E., Ruhl, H. A., Saeedi, H., Stefanoudis, P. V., Taranto, G. H., Thompson, M. B., Taylor, J. R., Tyler, P., Vad, J., Victorero, L., Vieira, R. P., Woodall, L. C., Xavier, J. R. and Wagner, D. (2019) 'A framework for the development of a global standardised marine taxon reference image database (SMarTaR-ID) to support image-based analyses', *BioRxiv*, pp. 670786.

Howell, K. L., Davies, J. S. and Narayanaswamy, B. E. (2010b) 'Identifying deep-sea megafaunal epibenthic assemblages for use in habitat mapping and marine protected area network design', *Journal of the Marine Biological Association of the United Kingdom,* 90(1), pp. 33-68.

Howell, K. L., Holt, R., Endrino, I. P. and Stewart, H. (2011) 'When the species is also a habitat: Comparing the predictively modelled distributions of Lophelia pertusa and the reef habitat it forms', *Biological Conservation,* 144(11), pp. 2656-2665.

Howell, K. L., Taylor, M., Crombie, K., Faithfull, S., Golding, N., Nimmo-Smith, W. A., Perrett, J., Piechaud, N., Ross, R. E., Stashchuk, N., Turner, D., Vlasenko, V. and Foster, N. (2016b) 'RRS James Cook, Cruise No. JC136, 14th May – 23rd June, DEEPLINKS: Influence of population connectivity on depth-dependent diversity of deep-sea marine benthic biota.

Hsieh, T., Ma, K. and Chao, A. (2016) 'iNEXT: an R package for rarefaction and extrapolation of species diversity (H ill numbers)', *Methods in Ecology and Evolution,* 7(12), pp. 1451-1456.

Hudson, I. R. and Wigham, B. D. (2003) 'In situ observations of predatory feeding behaviour of the galatheid squat lobster Munida sarsi using a remotely operated vehicle', *Journal of the Marine Biological Association of the United Kingdom,* 83(3), pp. 463-464.

Hughes, D. J. and Gage, J. D. (2004) 'Benthic metazoan biomass, community structure and bioturbation at three contrasting deep-water sites on the northwest European continental margin', *Progress in Oceanography,* 63(1-2), pp. 29-55.

Hughes, J. A. and Gooday, A. J. (2004) 'Associations between living benthic foraminifera and dead tests of Syringammina fragilissima (Xenophyophorea) in the Darwin Mounds region (NE Atlantic)', *Deep-Sea Research Part I-Oceanographic Research Papers,* 51(11), pp. 1741-1758.

Hurlbert, S. H. (1984) 'Pseudoreplication and the design of ecological field experiments', *Ecological monographs,* 54(2), pp. 187-211.

Hussey, N. E., Kessel, S. T., Aarestrup, K., Cooke, S. J., Cowley, P. D., Fisk, A. T., Harcourt, R. G., Holland, K. N., Iverson, S. J., Kocik, J. F., Mills Flemming, J. E. and Whoriskey, F. G. (2015) 'Aquatic animal telemetry: A panoramic window into the underwater world', *Science,* 348(6240).

Huvenne, V., Bett, B., Masson, D., Le Bas, T. and Wheeler, A. (2016a) 'Effectiveness of a deep-sea cold-water coral Marine Protected Area, following eight years of fisheries closure', *Biological Conservation,* 200, pp. 60-69.

Huvenne, V. A., McPhail, S. D., Wynn, R. B., Furlong, M. and Stevenson, P. (2009) 'Mapping giant scours in the deep ocean', *Eos, Transactions American Geophysical Union,* 90(32), pp. 274-275.

Huvenne, V. A., Robert, K., Marsh, L., Iacono, C. L., Le Bas, T. and Wynn, R. B. (2018) 'ROVs and AUVs', *Submarine Geomorphology*: Springer, pp. 93-108.

Huvenne, V. A. I. (2011) *RRS James Cook Cruise 60, 09 May-12 June 2011. Benthic habitats and the impact of human activities in Rockall Trough, on Rockall Bank and Hatton Basin*, Southampton, UK: National Oceanography Centre Southampton.

Huvenne, V. A. I., Bett, B. J., Masson, D. G., Le Bas, T. P. and Wheeler, A. J. (2016b) 'Effectiveness of a deep-sea cold-water coral Marine Protected Area, following eight years of fisheries closure', *Biological Conservation,* 200, pp. 60-69.

ICES (2010) 'Report of the working group on the assessment of southern shelf stocks of hake, monk and megrim'.

Ingram, T. (2011) 'Speciation along a depth gradient in a marine adaptive radiation', *Proceedings of the Royal Society B: Biological Sciences,* 278(1705), pp. 613-618.

James, L. C., Marzloff, M. P., Barrett, N., Friedman, A. and Johnson, C. R. (2017) 'Changes in deep reef benthic community composition across a latitudinal and environmental gradient in temperate Eastern Australia', *Marine Ecology Progress Series,* 565, pp. 35-52.

Jamieson, A. J., Boorman, B. and Jones, D. O. B. (2013) 'Deep-Sea Benthic Sampling', *Methods for the Study of Marine Benthos*, pp. 285-347.

Jeffries, H., Berman, M., Poularikas, A., Katsinis, C., Melas, I., Sherman, K. and Bivins, L. (1984) 'Automated sizing, counting and identification of zooplankton by pattern recognition', *Marine biology,* 78(3), pp. 329-334.

Jones, D. O., Gates, A. R., Huvenne, V. A., Phillips, A. B. and Bett, B. J. (2019) 'Autonomous marine environmental monitoring: Application in decommissioned oil fields', *Science of the total environment,* 668, pp. 835-853.

Jones, D. O. B., Kaiser, S., Sweetman, A. K., Smith, C. R., Menot, L., Vink, A., Trueblood, D., Greinert, J., Billett, D. S. M., Arbizu, P. M., Radziejewska, T., Singh, R., Ingole, B., Stratmann, T., Simon-Lledó, E., Durden, J. M. and Clark, M. R. (2017) 'Biological responses to disturbance from simulated deep-sea polymetallic nodule mining', *PLOS ONE,* 12(2), pp. e0171750.

Jongman, R. H. G. (2013) 'Biodiversity observation from local to global', *Ecological Indicators,* 33(Supplement C), pp. 1-4.

Jost, L. (2006) 'Entropy and diversity', *Oikos,* 113(2), pp. 363-375.

Jumars, P. (1975a) 'Environmental grain and polychaete species' diversity in a bathyal benthic community', *Marine Biology,* 30(3), pp. 253-266.

Jumars, P. (1975b) 'Methods for measurement of community structure in deep-sea macrobenthos', *Marine Biology,* 30(3), pp. 245-252.

Jumars, P. A. (1976) 'Deep-sea species diversity: does it have a characteristic scale', *J. mar. Res,* 34(2), pp. 217-246.

Jumars, P. A. and Eckman, J. E. (1983) 'Spatial structure within deep-sea benthic communities', *The sea,* 8, pp. 399-451.

Keith, S. A., Webb, T. J., Böhning-Gaese, K., Connolly, S. R., Dulvy, N. K., Eigenbrod, F., Jones, K. E., Price, T., Redding, D. W. and Owens, I. P. 2012. What is macroecology? : The Royal Society.

Kenchington, E., Callery, O., Davidson, F., Grehan, A., Morato, T., Appiott, J., Davis, A., Dunstan, P., Du Preez, C. and Finney, J. (2019) *Use of Species Distribution Modeling in the Deep Sea.*

Kostylev, V. E. (2012) 'Benthic habitat mapping from seabed acoustic surveys: do implicit assumptions hold?', *Int. Assoc. Sedimentol. Spec. Publ,* 44, pp. 405-416.

Krizhevsky, A., Sutskever, I. and Hinton, G. E. 'Imagenet classification with deep convolutional neural networks'. *Advances in neural information processing systems*, 1097-1105.

Kroeger, K. and Johnston, C. 2016. The UK marine biodiversity monitoring strategy v4. 1.

Kuhn, M. (2008) 'The caret package', *Journal of Statistical Software,* 28(5), pp. 1-26.

Lacharité, M. and Brown, C. J. (2019) 'Utilizing benthic habitat maps to inform biodiversity monitoring in marine protected areas', *Aquatic Conservation: Marine and Freshwater Ecosystems,* 29(6), pp. 938-951.

Lacharité, M. and Metaxas, A. (2017) 'Hard substrate in the deep ocean: How sediment features influence epibenthic megafauna on the eastern Canadian margin', *Deep Sea Research Part I: Oceanographic Research Papers,* 126(Supplement C), pp. 50-61.

Lacharité, M., Metaxas, A. and Lawton, P. (2015) 'Using object-based image analysis to determine seafloor fine-scale features and complexity', *Limnology and Oceanography: Methods,* 13(10), pp. 553-567.

Langenkämper, D., Zurowietz, M., Schoening, T. and Nattkemper, T. W. (2017) 'BIIGLE 2.0 - Browsing and Annotating Large Marine Image Collections', *Frontiers in Marine Science,* 4(83).

LeCun, Y., Bengio, Y. and Hinton, G. (2015) 'Deep learning', *nature,* 521(7553), pp. 436.

Leonardsson, K., Blomqvist, M. and Rosenberg, R. (2016) 'Reducing spatial variation in environmental assessment of marine benthic fauna', *Marine pollution bulletin,* 104(1), pp. 129-138.

Levin, L. and Gooday, A. (1992a) 'Possible roles for xenophyophores in deep-sea carbon cycling', *Deep-sea food chains and the global carbon cycle*: Springer, pp. 93-104.

Levin, L. A. (1991) 'Interaction between metazoans and large, agglutinating protozoans: implications for the community structure of deep-sea benthos', *American Zoology,* 31, pp. 886-900.

Levin, L. A., Bett, B. J., Gates, A. R., Heimbach, P., Howe, B. M., Janssen, F., McCurdy, A., Ruhl, H. A., Snelgrove, P., Stocks, K. I., Bailey, D., Baumann-Pickering, S., Beaverson, C., Benfield, M. C., Booth, D. J., Carreiro-Silva, M., Colaço, A., Eblé, M. C., Fowler, A. M., Gjerde, K. M., Jones, D. O. B., Katsumata, K., Kelley, D., Le Bris, N., Leonardi, A. P., Lejzerowicz, F., Macreadie, P. I., McLean, D., Meitz, F., Morato, T., Netburn, A., Pawlowski, J., Smith, C. R., Sun, S., Uchida, H., Vardaro, M. F., Venkatesan, R. and Weller, R. A. (2019) 'Global Observing Needs in the Deep Ocean', *Frontiers in Marine Science,* 6(241).

Levin, L. A., DeMaster, D. J., McCann, L. D. and Thomas, C. L. (1986) 'Effect of giant protozoans (class:Xenophyophorea) on deep-seamount benthos', *Marine Ecology Progress Series,* 29, pp. 99-104.

Levin, L. A., Etter, R. J., Rex, M. A., Gooday, A. J., Smith, C. R., Pineda, J., Stuart, C. T., Hessler, R. R. and Pawson, D. (2001) 'Environmental influences on regional deep-sea species diversity', *Annual Review of Ecology and Systematics,* 32(1), pp. 51-93.

Levin, L. A. and Gooday, A. J. (1992b) 'Possible roles for Xenophyophores in deep-sea carbon cycling', in Rowe, G.T. & Pariente, V. (eds.) *Deep-Sea Food Chains and the Global Carbon Cycle*. Netherlands: Kluwer, pp. 93-104.

Levin, L. A. and Le Bris, N. (2015) 'The deep ocean under climate change', *Science,* 350(6262), pp. 766-768.

Levin, L. A. and Thomas, C. L. (1988) 'THE ECOLOGY OF XENOPHYOPHORES (PROTISTA) ON EASTERN PACIFIC SEAMOUNTS', *Deep-Sea Research Part a-Oceanographic Research Papers,* 35(12), pp. 2003-2027.

Lo Iacono, C., Robert, K., Gonzalez-Villanueva, R., Gori, A., Gili, J.-M. and Orejas, C. (2018) 'Predicting cold-water coral distribution in the Cap de Creus Canyon (NW Mediterranean): Implications for marine conservation planning', *Progress in Oceanography,* 169, pp. 169-180.

Lobet, G. (2017) 'Image Analysis in Plant Sciences: Publish Then Perish', *Trends in Plant Science,* 22(7), pp. 559-566.

Lu, B. and He, Y. (2017) 'Species classification using Unmanned Aerial Vehicle (UAV)-acquired high spatial resolution imagery in a heterogeneous grassland', *ISPRS Journal of Photogrammetry and Remote Sensing,* 128(Supplement C), pp. 73-85.

Lu, D. and Weng, Q. (2007) 'A survey of image classification methods and techniques for improving classification performance', *International journal of Remote sensing,* 28(5), pp. 823-870.

Lucieer, V. L. and Forrest, A. L. (2016) 'Emerging Mapping Techniques for Autonomous Underwater Vehicles (AUVs)', in Finkl, C.W. & Makowski, C. (eds.) *Seafloor Mapping along Continental Shelves: Research and Techniques for Visualizing Benthic Environments*. Cham: Springer International Publishing, pp. 53-67.

MacLeod, N., Benfield, M. and Culverhouse, P. (2010) 'Time to automate identification', *Nature,* 467(7312), pp. 154-155.

MacPherson, K., Walker, D., Devillers, R., Kennedy, N., King, P., Lewis, R., Munroe, J. and Vardy, A. 'The development of AUV strategies for multidisciplinary use'. *Oceans-St. John's, 2014*: IEEE, 1-7.

Mallet, D. and Pelletier, D. (2014) 'Underwater video techniques for observing coastal marine biodiversity: A review of sixty years of publications (1952–2012)', *Fisheries Research,* 154, pp. 44-62.

Manderson, T., Li, J., Dudek, N., Meger, D. and Dudek, G. (2017) 'Robotic Coral Reef Health Assessment Using Automated Image Analysis', *Journal of Field Robotics,* 34(1), pp. 170-187.

Manel, S., Williams, H. C. and Ormerod, S. J. (2001) 'Evaluating presence–absence models in ecology: the need to account for prevalence', *Journal of applied Ecology,* 38(5), pp. 921-931.

Marburg, A. and Bigham, K. 'Deep learning for benthic fauna identification'. *OCEANS 2016 MTS/IEEE Monterey*: IEEE, 1-5.

Marchese, C. (2014) 'Biodiversity hotspots: A shortcut for a more complicated concept', *Global Ecology and Conservation,* 3, pp. 297-309.

Marcos, M. S. A. C., Soriano, M. N. and Saloma, C. A. (2005) 'Classification of coral reef images from underwater video using neural networks', *Optics Express,* 13(22), pp. 8766-8771.

Marini, S., Corgnati, L., Mantovani, C., Bastianini, M., Ottaviani, E., Fanelli, E., Aguzzi, J., Griffa, A. and Poulain, P.-M. (2018a) 'Automated estimate of fish abundance through the autonomous imaging device GUARD1', *Measurement,* 126, pp. 72-75.

Marini, S., Fanelli, E., Sbragaglia, V., Azzurro, E., Del Rio Fernandez, J. and Aguzzi, J. (2018b) 'Tracking Fish Abundance by Underwater Image Recognition', *Scientific Reports,* 8(1), pp. 13748.

Marsh, L., Huvenne, V. A. and Jones, D. O. (2018) 'Geomorphological evidence of large vertebrates interacting with the seafloor at abyssal depths in a region designated for deep-sea mining', *Royal Society open science,* 5(8), pp. 180286.

Marzinelli, E. M., Williams, S. B., Babcock, R. C., Barrett, N. S., Johnson, C. R., Jordan, A., Kendrick, G. A., Pizarro, O. R., Smale, D. A. and Steinberg, P. D. (2015) 'Large-Scale Geographic Variation in Distribution and Abundance of Australian Deep-Water Kelp Forests', *PLOS ONE,* 10(2), pp. e0118390.

Matabos, M., Hoeberechts, M., Doya, C., Aguzzi, J., Nephin, J., Reimchen, T. E., Leaver, S., Marx, R. M., Branzan Albu, A. and Fier, R. (2017) 'Expert, Crowd, Students or Algorithm: who holds the key to deep-sea imagery 'big data'processing?', *Methods in Ecology and Evolution,* 8(8), pp. 996-1004.

Maxwell, A. E., Warner, T. A. and Fang, F. (2018) 'Implementation of machine-learning classification in remote sensing: an applied review', *International Journal of Remote Sensing,* 39(9), pp. 2784-2817.

McCauley, D. J., Pinsky, M. L., Palumbi, S. R., Estes, J. A., Joyce, F. H. and Warner, R. R. (2015) 'Marine defaunation: Animal loss in the global ocean', *Science,* 347(6219), pp. 1255641.

McClain, C. R., Lundsten, L., Barry, J. and DeVogelaere, A. (2010) 'Assemblage structure, but not diversity or density, change with depth on a northeast Pacific seamount', *Marine Ecology,* 31(s1), pp. 14-25.

McClain, C. R., Nekola, J. C., Kuhnz, L. and Barry, J. P. (2011) 'Local-scale faunal turnover on the deep Pacific seafloor', *Marine Ecology Progress Series,* 422, pp. 193-200.

McClain, C. R. and Rex, M. A. (2015) 'Toward a Conceptual Understanding of β-Diversity in the Deep-Sea Benthos', *Annual Review of Ecology, Evolution, and Systematics,* 46(1), pp. 623-642.

McClain, C. R. and Schlacher, T. A. (2015) 'On some hypotheses of diversity of animal life at great depths on the sea floor', *Marine ecology,* 36(4), pp. 849-872.

McPhail, S. (2009) 'Autosub6000: A Deep Diving Long Range AUV', *Journal of Bionic Engineering,* 6(1), pp. 55-62.

McPhail, S., Stevenson, P., Pebody, M., Furlong, M., Perrett, J. and Lebas, T. 'Challenges of using an AUV to find and map hydrothermal vent sites in deep and rugged terrains'. *2010 IEEE/OES Autonomous Underwater Vehicles*, 1-3 Sept. 2010, 1-8.

Merow, C., Smith, M. J., Edwards, T. C., Guisan, A., McMahon, S. M., Normand, S., Thuiller, W., Wüest, R. O., Zimmermann, N. E. and Elith, J. (2014) 'What do we gain from simplicity versus complexity in species distribution models?', *Ecography,* 37(12), pp. 1267-1281.

Miglietta, M. P., Faucci, A. and Santini, F. (2011) 'Speciation in the sea: overview of the symposium and discussion of future directions', *Integrative and Comparative Biology,* 51(3), pp. 449-455.

Miller, J. A. (2012) 'Species distribution models:Spatial autocorrelation and non-stationarity', *Progress in Physical Geography: Earth and Environment,* 36(5), pp. 681-692.

Milligan, R. J., Morris, K. J., Bett, B. J., Durden, J. M., Jones, D. O. B., Robert, K., Ruhl, H. A. and Bailey, D. M. (2016) 'High resolution study of the spatial distributions of abyssal fishes by autonomous underwater vehicle', *Scientific Reports,* 6, pp. 26095.

Mohn, C., Rengstorf, A., White, M., Duineveld, G., Mienis, F., Soetaert, K. and Grehan, A. (2014) 'Linking benthic hydrodynamics and cold-water coral occurrences: A high-resolution model study at three cold-water coral provinces in the NE Atlantic', *Progress in Oceanography,* 122, pp. 92-104.

Monk, J., Barrett, N., Bridge, T., Carroll, A., Friedman, A., Jordan, A., Kendrick, G. and Lucieer, V. (2018) 'Marine Sampling Field Manual for AUV's (Autonomous Underwater Vehicles)'.

Mora, C., Tittensor, D. P., Adl, S., Simpson, A. G. and Worm, B. (2011) 'How many species are there on Earth and in the ocean?', *PLoS Biol,* 9(8), pp. e1001127.

Morris, K. J., Bett, B. J., Durden, J. M., Benoist, N. M., Huvenne, V. A., Jones, D. O., Robert, K., Ichino, M. C., Wolff, G. A. and Ruhl, H. A. (2016) 'Landscape-scale spatial heterogeneity in phytodetrital cover and megafauna biomass in the abyss links to modest topographic variation', *Scientific reports,* 6, pp. 34080.

Morris, K. J., Bett, B. J., Durden, J. M., Huvenne, V. A., Milligan, R., Jones, D. O., McPhail, S., Robert, K., Bailey, D. M. and Ruhl, H. A. (2014) 'A new method for ecological surveying of the abyss using autonomous underwater vehicle photography', *Limnology and Oceanography: Methods,* 12(11), pp. 795-809.

Morrisey, D., Howitt, L., Underwood, A. and Stark, J. (1992) 'Spatial variation in soft-sediment benthos', *Marine ecology progress series. Oldendorf,* 81(2), pp. 197-204.

Nagendra, H. (2002) 'Opposite trends in response for the Shannon and Simpson indices of landscape diversity', *Applied geography,* 22(2), pp. 175-186.

Nicholls, K., Abrahamsen, E., Buck, J., Dodd, P., Goldblatt, C., Griffiths, G., Heywood, K., Hughes, N., Kaletzky, A. and Lane-Serff, G. (2006) 'Measurements beneath an Antarctic ice shelf using an autonomous underwater vehicle', *Geophysical Research Letters,* 33(8).

Noble-James, T., Jesus, A. and McBreen, F. (2017) 'Monitoring guidance for marine benthic habitats'.

Norouzzadeh, M. S., Nguyen, A., Kosmala, M., Swanson, A., Palmer, M. S., Packer, C. and Clune, J. (2018) 'Automatically identifying, counting, and describing wild animals in camera-trap images with deep learning', *Proceedings of the National Academy of Sciences*, pp. 201719367.

Oksanen, J., Kindt, R., Legendre, P., O'Hara, B., Stevens, M. H. H., Oksanen, M. J. and Suggests, M. (2007) 'The vegan package', *Community ecology package. Disponível em: http://www. R-project. org. Acesso em,* 10(01), pp. 2008.

Ospar (2008) 'OSPAR List of Threatened and/or Declining Species and Habitats', OSPAR Agreement 2008-06, pp. pp. 1–4.

OSPAR. (2008b) 'OSPAR List of Threatened and/or Declining Species and Habitats', OSPAR Agreement 2008-06, pp. pp. 1–4.

Palumbi, S. R. (1994) 'Genetic divergence, reproductive isolation, and marine speciation', *Annual review of ecology and systematics,* 25(1), pp. 547-572.

Pan, S. J. and Yang, Q. (2010) 'A survey on transfer learning', *IEEE Transactions on knowledge and data engineering,* 22(10), pp. 1345-1359.

Parmesan, C., Ryrholm, N., Stefanescu, C., Hill, J. K., Thomas, C. D., Descimon, H., Huntley, B., Kaila, L., Kullberg, J. and Tammaru, T. (1999) 'Poleward shifts in geographical ranges of butterfly species associated with regional warming', *Nature,* 399(6736), pp. 579-583.

Parry, M. E. V., Howell, K. L., Narayanaswamy, B. E., Bett, B. J., Jones, D. O. B., Hughes, D. J., Piechaud, N., Ellwood, H., Askew, N., Jenkins, C. and Manca, E. (2015) *A Deep-sea Section for the Marine Habitat Classification of Britain and Ireland, V15. 03.* Joint Nature Conservation Committee.

Paull, L., Saeedi, S., Seto, M. and Li, H. (2014) 'AUV Navigation and Localization: A Review', *IEEE Journal of Oceanic Engineering,* 39(1), pp. 131-149.

Pennington, J. T., Blum, M. and Chavez, F. P. (2016) 'Seawater sampling by an autonomous underwater vehicle: “Gulper” sample validation for nitrate, chlorophyll, phytoplankton, and primary production', *Limnology and Oceanography: Methods,* 14(1), pp. 14-23.

Perkins, N. R., Foster, S. D., Hill, N. A. and Barrett, N. S. (2016) 'Image subsampling and point scoring approaches for large-scale marine benthic monitoring programs', *Estuarine, Coastal and Shelf Science,* 176, pp. 36-46.

Perkins, N. R., Hosack, G. R., Foster, S. D., Hill, N. A. and Barrett, N. S. (2019) 'Spatial properties of sessile benthic organisms and the design of repeat visual survey transects', *Aquatic Conservation: Marine and Freshwater Ecosystems,* 29(1), pp. 59-71.

Peukert, A., Schoening, T., Alevizos, E., Köser, K., Kwasnitschka, T. and Greinert, J. (2018) 'Understanding Mn-nodule distribution and evaluation of related deep-sea mining impacts using AUV-based hydroacoustic and optical data', *Biogeosciences (BG),* 15(8), pp. 2525-2549.

Phillips, A. B., Gold, N., Linton, N., Harris, C. A., Richards, E., Templeton, R., Thuné, S., Sitbon, J., Muller, M., Vincent, I. and Sloane, T. 'Agile design of low-cost autonomous underwater vehicles'. *OCEANS 2017 - Aberdeen*, 19-22 June 2017, 1-7.

Piechaud, N., Downie, A., Stewart, H. A. and Howell, K. L. (2015) 'The impact of modelling method selection on predicted extent and distribution of deep-sea benthic assemblages', *Earth and Environmental Science Transactions of the Royal Society of Edinburgh,* 105(04), pp. 251-261.

Pierdomenico, M., Guida, V. G., Macelloni, L., Chiocci, F. L., Rona, P. A., Scranton, M. I., Asper, V. and Diercks, A. (2015) 'Sedimentary facies, geomorphic features and habitat distribution at the Hudson Canyon head from AUV multibeam data', *Deep Sea Research Part II: Topical Studies in Oceanography,* 121, pp. 112-125.

Pimentel, M. A. F., Clifton, D. A., Clifton, L. and Tarassenko, L. (2014) 'A review of novelty detection', *Signal Processing,* 99, pp. 215-249.

Pimm, S. L., Jenkins, C. N., Abell, R., Brooks, T. M., Gittleman, J. L., Joppa, L. N., Raven, P. H., Roberts, C. M. and Sexton, J. O. (2014) 'The biodiversity of species and their rates of extinction, distribution, and protection', *Science,* 344(6187).

Pinsky, M. L., Worm, B., Fogarty, M. J., Sarmiento, J. L. and Levin, S. A. (2013) 'Marine Taxa Track Local Climate Velocities', *Science,* 341(6151), pp. 1239-1242.

Poore, G. C., Avery, L., Błażewicz-Paszkowycz, M., Browne, J., Bruce, N. L., Gerken, S., Glasby, C., Greaves, E., McCallum, A. W. and Staples, D. (2015) 'Invertebrate diversity of the unexplored marine western margin of Australia: taxonomy and implications for global biodiversity', *Marine Biodiversity,* 45(2), pp. 271-286.

Pringle, R. M., Doak, D. F., Brody, A. K., Jocqué, R. and Palmer, T. M. (2010) 'Spatial pattern enhances ecosystem functioning in an African savanna', *PLoS biology,* 8(5), pp. e1000377.

Probert, P. K. (2017) *Marine Conservation.* Cambridge University Press.

Przeslawski, R., Foster, S., Monk, J., Langlois, T., Lucieer, V. and Stuart-Smith, R. (2018) 'Comparative assessment of seafloor sampling platforms. Report to the National Environmental Science Programme', *Marine Biodiversity Hub, Geoscience Australia*.

Puebla, O. (2009) 'Ecological speciation in marine v. freshwater fishes', *Journal of Fish Biology,* 75(5), pp. 960-996.

Qian, K., Guo, J., Ishida, K. and Matsuoka, S. (2017) 'Fast recognition of bird sounds using extreme learning machines', *IEEJ Transactions on Electrical and Electronic Engineering,* 12(2), pp. 294-296.

Quintana, J., Garcia, R., Neumann, L., Campos, R., Weiss, T., Köser, K., Mohrmann, J. and Greinert, J. 'Towards automatic recognition of mining targets using an autonomous robot'. *OCEANS 2018 MTS/IEEE Charleston*, 22-25 Oct. 2018, 1-7.

Author (2011) *R: A Language and Environment for Statistical Computing*. Available at: http://www.R-project.org.

Ramirez-Llodra, E. and Billett, D. S. (2006) '3. DEEP-SEA ECOSYSTEMS: PRISTINE BIODIVERSITY RESERVOIR AND TECHNOLOGICAL CHALLENGES', *The Exploration of Marine Biodiversity Scientific and Technological Challenges*, pp. 63.

Ramirez-Llodra, E., Brandt, A., Danovaro, R., De Mol, B., Escobar, E., German, C., Levin, L., Arbizu, P., Menot, L. and Buhl-Mortensen, P. (2010) 'Deep, diverse and definitely different: unique attributes of the world's largest ecosystem', *Biogeosciences,* 7(9), pp. 2851-2899.

Ramirez-Llodra, E., Tyler, P. A., Baker, M. C., Bergstad, O. A., Clark, M. R., Escobar, E., Levin, L. A., Menot, L., Rowden, A. A. and Smith, C. R. (2011) 'Man and the last great wilderness: human impact on the deep sea', *PLoS One,* 6(8), pp. e22588.

Rampasek, L. and Goldenberg, A. (2016) 'TensorFlow: Biology's Gateway to Deep Learning?', *Cell Systems,* 2(1), pp. 12-14.

Reisenbichler, K. R., Chaffey, M. R., Cazenave, F., McEwen, R. S., Henthorn, R. G., Sherlock, R. E. and Robison, B. H. 'Automating MBARI's midwater time-series video surveys: The transition from ROV to AUV'. *OCEANS 2016 MTS/IEEE Monterey*: IEEE, 1-9.

Rengstorf, A. M., Grehan, A., Yesson, C. and Brown, C. (2012) 'Towards High-Resolution Habitat Suitability Modeling of Vulnerable Marine Ecosystems in the Deep-Sea: Resolving Terrain Attribute Dependencies', *Marine Geodesy,* 35(4), pp. 343-361.

Rengstorf, A. M., Mohn, C., Brown, C., Wisz, M. S. and Grehan, A. J. (2014) 'Predicting the distribution of deep-sea vulnerable marine ecosystems using high-resolution data: Considerations and novel approaches', *Deep Sea Research Part I: Oceanographic Research Papers,* 93, pp. 72-82.

Rengstorf, A. M., Yesson, C., Brown, C. and Grehan, A. J. (2013) 'High-resolution habitat suitability modelling can improve conservation of vulnerable marine ecosystems in the deep sea', *Journal of Biogeography,* 40(9), pp. 1702-1714.

Rex, M. A. and Etter, R. J. (2010) *Deep-sea biodiversity: pattern and scale.* Harvard University Press.

Rex, M. A., Stuart, C. T., Hessler, R. R., Allen, J. A., Sanders, H. L. and Wilson, G. D. (1993) 'Global-scale latitudinal patterns of species diversity in the deep-sea benthos', *Nature,* 365(6447), pp. 636.

Rice, A. L., Thurston, M. H. and New, A. L. (1990) 'Dense aggregations of a hexactinellid sponge, Pheronema carpenteri, in the Porcupine Seabight (northeast Atlantic Ocean), and possible causes', *Progress in Ocenaography,* 24, pp. 179-196.

Ripley, B., Venables, W. and Ripley, M. B. (2016) 'Package ‘nnet’', *R package version*, pp. 7-3.

Riqaud, V., Michel, J. L., Ferguson, J., Laframboise, J., Crees, T., Leon, P., Opderbecke, J. and Chardard, Y. 'First Steps In Ifremer's Autonomous Underwater Vehicle Program-A 3000m Depth Operational Survey AUV For Environmental Monitoring'. *The Fourteenth International Offshore and Polar Engineering Conference*: International Society of Offshore and Polar Engineers.

Robert, K., Huvenne, V. A., Georgiopoulou, A., Jones, D. O., Marsh, L., Carter, G. D. and Chaumillon, L. (2017) 'New approaches to high-resolution mapping of marine vertical structures', *Scientific reports,* 7(1), pp. 9005.

Robert, K., Jones, D. O., Roberts, J. M. and Huvenne, V. A. (2016) 'Improving predictive mapping of deep-water habitats: Considering multiple model outputs and ensemble techniques', *Deep Sea Research Part I: Oceanographic Research Papers,* 113, pp. 80-89.

Robert, K., Jones, D. O. B. and Huvenne, V. A. I. (2014) 'Megafaunal distribution and biodiversity in a heterogeneous landscape: the iceberg-scoured Rockall Bank, NE Atlantic', *Marine Ecology Progress Series,* 501, pp. 67-88.

Roberts, J. M., Harvey, S. M., Lamont, P. A., Gage, J. D. and Humphery, J. D. (2000) 'Seabed photography, environmental assessment and evidence for deep-water trawling on the continental margin west of the Hebrides', *Hydrobiologia,* 441(1-3), pp. 173-183.

Roberts, J. M., Wheeler, A. J. and Freiwald, A. (2006) 'Reefs of the deep: The biology and geology of cold-water coral ecosystems', *Science,* 312(5773), pp. 543-547.

Robinson, N. M., Nelson, W. A., Costello, M. J., Sutherland, J. E. and Lundquist, C. J. (2017) 'A Systematic Review of Marine-Based Species Distribution Models (SDMs) with Recommendations for Best Practice', *Frontiers in Marine Science,* 4(421).

Rogers, A. D. (2015) 'Environmental change in the deep ocean', *Annual Review of Environment and Resources,* 40, pp. 1-38.

Rogers, A. D., Brierley, A., Croot, P., Cunha, M. R., Danovaro, R., Devey, C., Hoel, A. H., Ruhl, H. A., Sarradin, P. M. and Trevisanut, S. (2015) 'Delving Deeper: Critical challenges for 21st century deep-sea research', *Position Paper,* 22, pp. 224.

Rohlf, F. J. and Sokal, R. R. (1967) 'Taxonomic structure from randomly and systematically scanned biological images', *Systematic Zoology,* 16(3), pp. 246-260.

Roig Marí, C. (2016) *Deep learning architectures for Computer Vision.* Universitat Politècnica de Catalunya.

Rolke, M. and Lenz, J. (1984) 'Size structure analysis of zooplankton samples by means of an automated image analyzing system', *Journal of Plankton Research,* 6(4), pp. 637-645.

Romero-Ramirez, A., Grémare, A., Bernard, G., Pascal, L., Maire, O. and Duchêne, J. C. (2016) 'Development and validation of a video analysis software for marine benthic applications', *Journal of Marine Systems,* 162(Supplement C), pp. 4-17.

Rooper, C. N., Sigler, M. F., Goddard, P., Malecha, P., Towler, R., Williams, K., Wilborn, R. and Zimmermann, M. (2016) 'Validation and improvement of species distribution models for structure-forming invertebrates in the eastern Bering Sea with an independent survey', *Marine Ecology Progress Series,* 551, pp. 117-130.

Roper, D. T., Phillips, A. B., Harris, C. A., Salavasidis, G., Pebody, M., Templeton, R., Amma, S. V. S., Smart, M. and McPhail, S. 'Autosub long range 1500: An ultra-endurance AUV with 6000 Km range'. *OCEANS 2017 - Aberdeen*, 19-22 June 2017, 1-5.

Ross, L. K., Ross, R. E., Stewart, H. A. and Howell, K. L. (2015) 'The Influence of Data Resolution on Predicted Distribution and Estimates of Extent of Current Protection of Three ‘Listed’Deep-Sea Habitats', *PLOS ONE,* 10(10), pp. e0140061.

Ross, R. E. and Howell, K. L. (2013) 'Use of predictive habitat modelling to assess the distribution and extent of the current protection of 'listed' deep-sea habitats', *Diversity and Distributions,* 19(4), pp. 433-445.

Rowden, A. A., Anderson, O. F., Georgian, S. E., Bowden, D. A., Clark, M. R., Pallentin, A. and Miller, A. (2017) 'High-Resolution Habitat Suitability Models for the Conservation and Management of Vulnerable Marine Ecosystems on the Louisville Seamount Chain, South Pacific Ocean', *Frontiers in Marine Science,* 4(335).

Rowden, A. A., Schnabel, K. E., Schlacher, T. A., Macpherson, E., Ahyong, S. T. and de Forges, B. R. (2010) 'Squat lobster assemblages on seamounts differ from some, but not all, deep-sea habitats of comparable depth', *Marine Ecology-an Evolutionary Perspective,* 31, pp. 63-83.

RStudio Team, R. (2015) 'RStudio: integrated development for R', *RStudio, Inc., Boston, MA URL http://www. rstudio. com,* 42, pp. 14.

Samek, W., Wiegand, T. and Müller, K.-R. (2017) 'Explainable artificial intelligence: Understanding, visualizing and interpreting deep learning models', *arXiv preprint arXiv:1708.08296*.

Sanders, H. L. (1968) 'Marine benthic diversity: a comparative study', *The American Naturalist,* 102(925), pp. 243-282.

Sardà, F. and Aguzzi, J. (2012) 'A review of burrow counting as an alternative to other typical methods of assessment of Norway lobster populations', *Reviews in Fish Biology and Fisheries,* 22(2), pp. 409-422.

Schmid, M. S., Aubry, C., Grigor, J. and Fortier, L. (2016) 'The LOKI underwater imaging system and an automatic identification model for the detection of zooplankton taxa in the Arctic Ocean', *Methods in Oceanography,* 15, pp. 129-160.

Schneider, S., Taylor, G. W. and Kremer, S. C. (2018) 'Deep Learning Object Detection Methods for Ecological Camera Trap Data', *arXiv preprint arXiv:1803.10842*.

Schoening, T., Bergmann, M., Ontrup, J., Taylor, J., Dannheim, J., Gutt, J., Purser, A. and Nattkemper, T. W. (2012) 'Semi-automated image analysis for the assessment of megafaunal densities at the Arctic deep-sea observatory HAUSGARTEN', *PloS one,* 7(6), pp. e38179.

Schoening, T., Durden, J., Preuss, I., Albu, A. B., Purser, A., De Smet, B., Dominguez-Carrió, C., Yesson, C., de Jonge, D. and Lindsay, D. (2017) 'Report on the Marine Imaging Workshop 2017', *Research Ideas and Outcomes,* 3, pp. e13820.

Schoening, T., Köser, K. and Greinert, J. (2018) 'An acquisition, curation and management workflow for sustainable, terabyte-scale marine image analysis', *Scientific Data,* 5, pp. 180181.

SCOR (1994) 'Suggested criteria for describing deep-sea benthic communities; the final report of SCOR Working Group 76', *Progress in Oceanography,* 34(2-3), pp. 81-100.

Seymour, A. C., Dale, J., Hammill, M., Halpin, P. N. and Johnston, D. W. (2017) 'Automated detection and enumeration of marine wildlife using unmanned aircraft systems (UAS) and thermal imagery', *Scientific Reports,* 7, pp. 45127.

Shafait, F., Harvey, E. S., Shortis, M. R., Mian, A., Ravanbakhsh, M., Seager, J. W., Culverhouse, P. F., Cline, D. E. and Edgington, D. R. (2017) 'Towards automating underwater measurement of fish length: a comparison of semi-automatic and manual stereo–video measurements', *ICES Journal of Marine Science,* 74(6), pp. 1690-1701.

Sherman, A. D. and Smith, K. L. (2009) 'Deep-sea benthic boundary layer communities and food supply: A long-term monitoring strategy', *Deep Sea Research Part II: Topical Studies in Oceanography,* 56(19), pp. 1754-1762.

Sibuet, M., Lambert, C. E., Chesselet, R. and Laubier, L. (1989) 'Density of the major size groups of benthic fauna and trophic input in deep basins of the Atlantic Ocean', *Journal of Marine Research,* 47(4), pp. 851-867.

Siddiqui, S. A., Salman, A., Malik, M. I., Shafait, F., Mian, A., Shortis, M. R., Harvey, E. S. and Handling editor: Howard, B. (2018) 'Automatic fish species classification in underwater videos: exploiting pre-trained deep neural network models to compensate for limited labelled data', *ICES Journal of Marine Science,* 75(1), pp. 374-389.

Simon-Lledó, E., Bett, B. J., Huvenne, V. A., Köser, K., Schoening, T., Greinert, J. and Jones, D. O. (2019) 'Biological effects 26 years after simulated deep-sea mining', *Scientific reports,* 9(1), pp. 1-13.

Simon-Lledó, E., Bett, B. J., Huvenne, V. A. I., Schoening, T., Benoist, N. M. A., Jeffreys, R. M., Durden, J. M. and Jones, D. O. B. (2018) 'Megafaunal variation in the abyssal landscape of the Clarion Clipperton Zone', *Progress in Oceanography,* 170, pp. 119-133.

Singh, H., Armstrong, R., Gilbes, F., Eustice, R., Roman, C., Pizarro, O. and Torres, J. (2004) 'Imaging coral I: Imaging coral habitats with the SeaBED AUV', *Subsurface Sensing Technologies and Applications,* 5(1), pp. 25-42.

Skiljan, I. (2012) 'IrfanView', https://www.irfanview.com/.

Smale, D. A., Kendrick, G. A., Harvey, E. S., Langlois, T. J., Hovey, R. K., Van Niel, K. P., Waddington, K. I., Bellchambers, L. M., Pember, M. B. and Babcock, R. C. (2012) 'Regional-scale benthic monitoring for ecosystem-based fisheries management (EBFM) using an autonomous underwater vehicle (AUV)', *ICES Journal of Marine Science,* 69(6), pp. 1108-1118.

Snelgrove, P. V. (1998) 'The biodiversity of macrofaunal organisms in marine sediments', *Biodiversity and Conservation,* 7(9), pp. 1123-1132.

Snelgrove, P. V. (2016a) 'An Ocean of Discovery: Biodiversity Beyond the Census of Marine Life', *Planta medica,* 82(09/10), pp. 790-799.

Snelgrove, P. V. R. (2016b) 'An Ocean of Discovery: Biodiversity Beyond the Census of Marine Life', *Planta medica*.

Soetaert, K. and Heip, C. (1990) 'Sample-size dependence of diversity indices and the determination of sufficient sample size in a high-diversity deep-sea environment', *Marine Ecology Progress Series*, pp. 305-307.

Solan, M., Germano, J. D., Rhoads, D. C., Smith, C., Michaud, E., Parry, D., Wenzhöfer, F., Kennedy, B., Henriques, C., Battle, E., Carey, D., Iocco, L., Valente, R., Watson, J. and Rosenberg, R. (2003) 'Towards a greater understanding of pattern, scale and process in marine benthic systems: a picture is worth a thousand worms', *Journal of Experimental Marine Biology and Ecology,* 285-286(Supplement C), pp. 313-338.

Sorensen, S., Treible, W., Hsu, L., Wang, X., Mahoney, A. R., Zitterbart, D. P. and Kambhamettu, C. 'Deep Learning for Polar Bear Detection'. *Scandinavian Conference on Image Analysis*: Springer, 457-467.

Sun, C., Shrivastava, A., Singh, S. and Gupta, A. 'Revisiting Unreasonable Effectiveness of Data in Deep Learning Era'. *2017 IEEE International Conference on Computer Vision (ICCV)*, 22-29 Oct. 2017, 843-852.

Sunday, J. M., Bates, A. E. and Dulvy, N. K. (2012) 'Thermal tolerance and the global redistribution of animals', *Nature Climate Change,* 2(9), pp. 686-690.

Sweetman, A. K., Thurber, A. R., Smith, C. R., Levin, L. A., Mora, C., Wei, C.-L., Gooday, A. J., Jones, D. O., Rex, M. and Yasuhara, M. (2017) 'Major impacts of climate change on deep-sea benthic ecosystems', *Elem Sci Anth,* 5, pp. 4.

Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J. and Wojna, Z. 'Rethinking the inception architecture for computer vision'. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2818-2826.

Tarnita, C. E., Bonachela, J. A., Sheffer, E., Guyton, J. A., Coverdale, T. C., Long, R. A. and Pringle, R. M. (2017) 'A theoretical foundation for multi-scale regular vegetation patterns', *Nature,* 541(7637), pp. 398-401.

Taylor, L., Woiwod, I. and Perry, J. (1978) 'The density-dependence of spatial behaviour and the rarity of randomness', *The Journal of Animal Ecology,* 47, pp. 383-406.

Taylor, M. L. and Roterman, C. N. (2017) 'Invertebrate population genetics across Earth's largest habitat: The deep-sea floor', *Molecular Ecology,* 26(19), pp. 4872-4896.

Team, R. C. 2014. R: A Language and Environment for Statistical Computing. Vienna; 2014.

Tendal, O. S. (1972) 'A monograph of the Xenophyophoria (Rhizopodea, Protozoa)', *Galathea Report,* 12, pp. 7-103.

Thornton, B., Takahashi, T., Sato, T., Sakka, T., Tamura, A., Matsumoto, A., Nozaki, T., Ohki, T. and Ohki, K. (2015) 'Development of a deep-sea laser-induced breakdown spectrometer for in situ multi-element chemical analysis', *Deep Sea Research Part I: Oceanographic Research Papers,* 95, pp. 20-36.

Tolimieri, N., Clarke, M. E., Singh, H. and Goldfinger, C. (2008) 'Evaluating the SeaBED AUV for monitoring groundfish in untrawlable habitat', *Marine habitat mapping technology for Alaska*, pp. 129-141.

Tong, R. J., Purser, A., Guinan, J. and Unnithan, V. (2013) 'Modeling the habitat suitability for deep-water gorgonian corals based on terrain variables', *Ecological Informatics,* 13, pp. 123-132.

Turner, P. J., Campbell, L. M. and Van Dover, C. L. (2017) 'Stakeholder perspectives on the importance of rare-species research for deep-sea environmental management', *Deep Sea Research Part I: Oceanographic Research Papers,* 125(Supplement C), pp. 129-134.

Underwood, A. J. (1994) 'On Beyond BACI: Sampling Designs that Might Reliably Detect Environmental Disturbances', *Ecological Applications,* 4(1), pp. 3-15.

Van Dover, C., Ardron, J., Escobar, E., Gianni, M., Gjerde, K., Jaeckel, A., Jones, D., Levin, L., Niner, H. and Pendleton, L. (2017) 'Biodiversity loss from deep-sea mining', *Nature Geoscience,* 10(7), pp. 464-465.

Van Dover, C., Aronson, J., Pendleton, L., Smith, S., Arnaud-Haond, S., Moreno-Mateos, D., Barbier, E., Billett, D., Bowers, K. and Danovaro, R. (2014) 'Ecological restoration in the deep sea: Desiderata', *Marine Policy,* 44, pp. 98-106.

Van Dover, C. L. (2011) 'Mining seafloor massive sulphides and biodiversity: what is at risk?', *Ices Journal of Marine Science,* 68(2), pp. 341-348.

Vanreusel, A., Hilario, A., Ribeiro, P. A., Menot, L. and Arbizu, P. M. (2016) 'Threatened by mining, polymetallic nodules are required to preserve abyssal epifauna', *Scientific reports,* 6, pp. 26808.

Vierod, A. D., Guinotte, J. M. and Davies, A. J. (2014) 'Predicting the distribution of vulnerable marine ecosystems in the deep sea using presence-background models', *Deep Sea Research Part II: Topical Studies in Oceanography,* 99, pp. 6-18.

Villero, D., Pla, M., Camps, D., Ruiz-Olmo, J. and Brotons, L. (2016) 'Integrating species distribution modelling into decision-making to inform conservation actions', *Biodiversity and Conservation,* 26(2), pp. 251-271.

Watling, L., Guinotte, J., Clark, M. R. and Smith, C. R. (2013) 'A proposed biogeography of the deep ocean floor', *Progress in Oceanography,* 111, pp. 91-112.

Webb, T. J., Berghe, E. V. and O'Dor, R. (2010) 'Biodiversity's big wet secret: the global distribution of marine biological records reveals chronic under-exploration of the deep pelagic ocean', *PLoS One,* 5(8), pp. e10223.

Wei, C.-L., Rowe, G. T., Escobar-Briones, E., Boetius, A., Soltwedel, T., Caley, M. J., Soliman, Y., Huettmann, F., Qu, F. and Yu, Z. (2010a) 'Global patterns and predictions of seafloor biomass using random forests', *PLoS One,* 5(12), pp. e15323.

Wei, C.-L., Rowe, G. T., Hubbard, G. F., Scheltema, A. H., Wilson, G. D., Petrescu, I., Foster, J. M., Wicksten, M. K., Chen, M. and Davenport, R. (2010b) 'Bathymetric zonation of deep-sea macrofauna in relation to export of surface phytoplankton production', *Marine Ecology Progress Series,* 399, pp. 1-14.

Weinstein, B. G. (2018) 'A computer vision for animal ecology', *Journal of Animal Ecology,* 87(3), pp. 533-545.

Wenger, S. J. and Olden, J. D. (2012) 'Assessing transferability of ecological models: an underappreciated aspect of statistical validation', *Methods in Ecology and Evolution,* 3(2), pp. 260-267.

Wheeler, B. and Torchiano, M. (2010) 'lmPerm: Permutation tests for linear models', *R package version,* 1(1.2).

White, C. F., Lin, Y., Clark, C. M. and Lowe, C. G. (2016) 'Human vs robot: Comparing the viability and utility of autonomous underwater vehicles for the acoustic telemetry tracking of marine organisms', *Journal of Experimental Marine Biology and Ecology,* 485, pp. 112-118.

Wickham, H. (2017) 'Tidyverse: Easily install and load'tidyverse'packages', *R package version,* 1(1).

Wiklund, H., Taylor, J. D., Dahlgren, T. G., Todt, C., Ikebe, C., Rabone, M. and Glover, A. G. (2017) 'Abyssal fauna of the UK-1 polymetallic nodule exploration area, Clarion-Clipperton Zone, central Pacific Ocean: Mollusca', *ZooKeys,* 707, pp. 1.

Williams, A., Althaus, F. and Schlacher, T. A. (2015) 'Towed camera imagery and benthic sled catches provide different views of seamount benthic diversity', *Limnology and Oceanography: Methods,* 13(2), pp. 62-73.

Williams, S. B., Pizarro, O., Jakuba, M. V., Johnson, C., Barrett, N., Babcock, R., Kendrick, G., Steinberg, P., Heyward, A. and Doherty, P. (2012a) 'Autonomous Underwater Vehicle monitoring of benthic reference sites', *IEEE Robotics and Automation Magazine,* 19, pp. 73-84.

Williams, S. B., Pizarro, O., Steinberg, D. M., Friedman, A. and Bryson, M. (2016) 'Reflections on a decade of autonomous underwater vehicles operations for marine survey at the Australian Centre for Field Robotics', *Annual Reviews in Control,* 42, pp. 158-165.

Williams, S. B., Pizarro, O. R., Jakuba, M. V., Johnson, C. R., Barrett, N. S., Babcock, R. C., Kendrick, G. A., Steinberg, P. D., Heyward, A. J. and Doherty, P. J. (2012b) 'Monitoring of benthic reference sites: using an autonomous underwater vehicle', *IEEE Robotics & Automation Magazine,* 19(1), pp. 73-84.

Wisz, M. S., Hijmans, R. J., Li, J., Peterson, A. T., Graham, C. H., Guisan, A. and Distribut, N. P. S. (2008) 'Effects of sample size on the performance of species distribution models', *Diversity and Distributions,* 14(5), pp. 763-773.

Wölfl, A.-C., Snaith, H., Amirebrahimi, S., Devey, C. W., Dorschel, B., Ferrini, V., Huvenne, V. A. I., Jakobsson, M., Jencks, J., Johnston, G., Lamarche, G., Mayer, L., Millar, D., Pedersen, T. H., Picard, K., Reitz, A., Schmitt, T., Visbeck, M., Weatherall, P. and Wigley, R. (2019) 'Seafloor Mapping – The Challenge of a Truly Global Ocean Bathymetry', *Frontiers in Marine Science,* 6(283).

Woodall, L. C., Andradi-Brown, D. A., Brierley, A. S., Clark, M. R., Connelly, D., Hall, R. A., Howell, K. L., Huvenne, V. A., Linse, K. and Ross, R. E. (2018) 'A multidisciplinary approach for generating globally consistent data on mesophotic, deep-pelagic, and bathyal biological communities', *Oceanography,* 31(3), pp. 76-89.

Woodall, L. C., Sanchez-Vidal, A., Canals, M., Paterson, G. L., Coppock, R., Sleight, V., Calafat, A., Rogers, A. D., Narayanaswamy, B. E. and Thompson, R. C. (2014) 'The deep sea is a major sink for microplastic debris', *Royal Society Open Science,* 1(4), pp. 140317.

Woolley, S. N. C., Tittensor, D. P., Dunstan, P. K., Guillera-Arroita, G., Lahoz-Monfort, J. J., Wintle, B. A., Worm, B. and O'Hara, T. D. (2016) 'Deep-sea diversity patterns are shaped by energy availability', *Nature,* 533(7603), pp. 393-410.

Wright, D., Lundblad, E., Larkin, E., Rinehart, R., Murphy, J., Cary-Kothera, L. and Draganov, K. (2005) 'ArcGIS benthic terrain modeler', *Oregon State University, Corvallis, OR, USA*.

Wynn, R., Bett, B., Evans, A., Griffiths, G., Huvenne, V., Jones, A., Palmer, M., Dove, D., Howe, J. and Boyd, T. (2012) 'Investigating the feasibility of utilizing AUV and Glider technology for mapping and monitoring of the UK MPA network', *Final report for Defra project MB0118*.

Wynn, R. B., Huvenne, V. A., Le Bas, T. P., Murton, B. J., Connelly, D. P., Bett, B. J., Ruhl, H. A., Morris, K. J., Peakall, J. and Parsons, D. R. (2014) 'Autonomous Underwater Vehicles (AUVs): Their past, present and future contributions to the advancement of marine geoscience', *Marine Geology,* 352, pp. 451-468.

Yasuhara, M. and Danovaro, R. (2016) 'Temperature impacts on deep-sea biodiversity', *Biological Reviews,* 91(2), pp. 275-287.

Yates, K. L., Bouchet, P. J., Caley, M. J., Mengersen, K., Randin, C. F., Parnell, S., Fielding, A. H., Bamford, A. J., Ban, S., Barbosa, A. M., Dormann, C. F., Elith, J., Embling, C. B., Ervin, G. N., Fisher, R., Gould, S., Graf, R. F., Gregr, E. J., Halpin, P. N., Heikkinen, R. K., Heinänen, S., Jones, A. R., Krishnakumar, P. K., Lauria, V., Lozano-Montes, H., Mannocci, L., Mellin, C., Mesgaran, M. B., Moreno-Amat, E., Mormede, S., Novaczek, E., Oppel, S., Ortuño Crespo, G., Peterson, A. T., Rapacciuolo, G., Roberts, J. J., Ross, R. E., Scales, K. L., Schoeman, D., Snelgrove, P., Sundblad, G., Thuiller, W., Torres, L. G., Verbruggen, H., Wang, L., Wenger, S., Whittingham, M. J., Zharikov, Y., Zurell, D. and Sequeira, A. M. M. (2018) 'Outstanding Challenges in the Transferability of Ecological Models', *Trends in Ecology & Evolution,* 33(10), pp. 790-802.

Yesson, C., Taylor, M. L., Tittensor, D. P., Davies, A. J., Guinotte, J., Baco, A., Black, J., Hall-Spencer, J. M. and Rogers, A. D. (2012) 'Global habitat suitability of cold-water octocorals', *Journal of Biogeography,* 39(7), pp. 1278-1292.

Zeppilli, D., Pusceddu, A., Trincardi, F. and Danovaro, R. (2016) 'Seafloor heterogeneity influences the biodiversity–ecosystem functioning relationships in the deep sea', *Scientific reports,* 6, pp. 26352.

Zurowietz, M., Langenkämper, D., Hosking, B., Ruhl, H. A. and Nattkemper, T. W. (2018) 'MAIA—A machine learning assisted image annotation method for environmental monitoring and exploration', *PloS one,* 13(11), pp. e0207498.

Publications

Chapter 5 is published as:

https://www.int-res.com/abstracts/meps/v615/p15-30/