Master's thesis in Coastal Ecology

Visual fish assemblage inventory of outer Oslo fjord, with special emphasis on young-of-the-year Atlantic cod (*Gadus morhua*)

Alexandra Marie Daly

SUPERVISOR

Even Moland

University of Agder, 2021

Faculty of Technology and Science

Department of Natural Sciences

University of Agder

Faculty of Technology and Science

Department of Natural Sciences

Gimlemoen 25

4604 Kristiansand

http: / /www.uia.no

Sammendrag

Kystsonen langs ytre Oslofjord er fysisk unik, og huser flere økologisk utarmede men økonomisk viktige arter, blant annet torsk (*Gadus morhua*). Kystnær torsk i Skagerrak deles opp i to "økotyper" som sameksisterer i Ytre Oslofjord. Resterende lokale populasjoner vurderes å være sårbare overfor klimaendringer og andre menneskeskapte påvirkninger. Rekruttering av torsk i Skagerrak overvåkes av "strandnotserien" som .av Havforskningsinstituttet. En sterk årsklasse med torsk rekrutterte til oppvekstområder i Ytre Oslofjord i 2017. Rekruttering av torskeyngel har i betydelig grad vært gjenstand for overvåking langs i Skagerrakkysten. Likevel er det kun i nyere tid at studier har begynt å undersøke rekrutteringsfasen hos torsk og hvilke faktorer som betinger suksess etter bunnslåing, og hvordan dette påvirker populasjonen som helhet. En metode som i økende grad er blitt tatt i bruk for undersøkelse av fiskebestander er agnet stereo-video eller "Baited Remote Underwater Video" (BRUV). Studier som benytter denne visuelle teknikken kan gi nærmere og mindre invasivt innsyn - uten fysisk prøvetaking. Videoopptak sikrer at data kan deles og evalueres objektivt. Målene for dette arbeidet har vært å samle biologisk grunnlinje-data from studieområdet, innhente informasjon om fiskesamfunnet på habitatene i fjorden på tidspunktet for gjennomføring, vurdere om ulike habitattyper er egnet for å huse juvenil torsk, samt å avgjøre hvilke av de tilgjengelige forklaringsvariablene som påvirker forekomst av torskerekrutter, samt juvenile stadier hos andre arter. Et BRUV-feltarbeid ble gjennomført i august 2017, og data er blitt generert ved videoanalyser i programvaren EventMeasure. Basert på analyser av det genererte datasettet var arter i familien Gadidae mest tallrike på stasjonene, og torsk den mest tallrike ut av disse. Tilstedeværelse og antall torsk lot til å ha innvirkning på mangfoldet av fiskearter i området, og funnene indikerer at grunne dyp og sandrike habitat var faktorer som forklarte tilstedeværelse av torsk ved målte stasjoner.

Abstract

The coastal zone along the outer Oslofjord is physically unique and home to many ecologically depleted and economically important species, including Atlantic cod (*Gadus morhua*). Coastal cod in Skagerrak is divided into two ecotypes that co-exist in outer Oslofjord.Remnant local populations are considered sensitive to climate disturbances and anthropogenic influences. Recruitment of cod in Skagerrak is monitored in the annual beach seine sampling. A strong year-class of cod recruited to nursery areas in outer Oslo fjord in 2017. Although recruitment of cod has been monitored extensively in Skagerrak, studies have only recently begun to examine the recruitment phase of Atlantic cod, the factors that confer success or failure post settlement, and how recruitment plays a role in the overall health and survival of the population. One method increasingly used to examine fish populations is Baited Remote Underwater Video (BRUV). Studies using visual sampling techniques enable scientists to take a closer and less invasive look, while video recording of observations means that data might be shared and validated objectively. The aims of this study were to gather baseline biological data, capture information on the fish assemblage present in the fjord habitats at the time of sampling, scope for features of habitat and substrate that is conducive to harboring young-of-the-year Atlantic cod and determine factors with explanatory properties regarding occurrence and co-occurrence of cod recruits and juveniles of other species. A BRUV study was conducted in August 2017 and the data derived from the video in the purpose built software EventMeasure. Based on analyses of the data thus obtained, Gadidae was found to be the most abundant family with cod being the most abundant species in that family. Cod presence and counts (abundance) seemed to have an impact on the fish diversity of the area, and it was found that shallow depths and sandy habitats were the strongest drivers in determining cod presence at sampled stations.

Table of contents

Table of Contents

Foreword

I would like to first thank my supervisor, Even Moland of IMR. Thank you for the support and hospitality on site at the Flødevigveien research station. I would also like to thank Tove Margrethe Gabrielsen at Universitetet i Agder for the guidance in choosing a thesis and support for all her students through this pandemic.

Thank you to my Jedi master Alli Cramer for the constant statistical advice and support, and setting this Padawan on this academic path. Thank you also to my other RStudio Wizard Marte Sodeland of UiA for the guidance and statistical assistance.

Thank you to my dear friends Dorian Norheim and Kane Haakonsen for the unconditional love, support and cooking. I would have been a malnourished and nervous wreck without you.

Thank you to my family who have worked tirelessly to ensure that I could have the opportunity to come to Norway and complete my studies, most especially my grandparents Barbara and Maurice Zepeda and my parents Gregg and Tanya Daly. Thank you to my aunt Cat Haney and cousins Crystal Turgeon, Max Haney and Anthony Hansen for always being the most vocal and encouraging cheerleaders in my corner.

Last but never least, thank you to my wonderful partner McLaughlin and his equally wonderful parents Diana Upton and Richard McLaughlin for the love, support, encouragement and incredible care packages to keep me going.

Kristiansand, Norway, 2 June 2021

Alexandra Marie Daly

1 Introduction

1.1 Coastal habitats in the Skagerrak Sea

The Skagerrak Sea is a diverse and complex area bordering southeast Norway, western Sweden and northern Denmark. Given that it is situated between the saline waters of the North Sea and the brackish waters of the Baltic Sea, this area is a popular study site for countless researchers examining its hydrographical and ecological diversity, which sustain important nursery grounds of many fish species (Munk et al., 2014). The Norwegian coast especially is a favorite among ecologists with many focusing on economically and ecologically important species (Fromentin et al., 2000, Jorde et al., 2007, Johannessen et al., 2011, Rogers et al., 2014, Huserbraaten et al., 2018). These coastal ecosystems are dynamic environments characterized by complex topography and currents (Rogers et al., 2014), leading to an overall hydrography that varies both seasonally and annually, which could explain the communities being dominated by species such as herring, gobies, butterfish, sprat, pipefishes, flatfish and eels (Munk et al., 2014). The shallow coastal zone is known to provide habitats that can sustain high numbers and production levels of fishes, making it a target area for both scientists and fisheries (Wennhage & Pihl, 2002). Many fish species, including the ones named above, can utilize multiple habitats, and are therefore influenced by the complexity and structure of the seascape (Perry et al., 2018).

1.1.1 Hydrology and ecology of Oslofjord

One of the primary features of the Norwegian coast is the Oslofjord, which is equally as diverse and complex as the sea it connects to. The outer part of the fjord itself is unique due to three sources of water mixing and contributing to its diversity in many ways. The coastal

current comes from the Bay of Bothnia, deep Atlantic water surfaces south of the Hvaler archipelago and fjord water from the fjord is influenced by the discharge of Norway's' two largest rivers, the Glomma and Drammenselva. Having a multitude of water sources influences both the physical and biological mechanics of the ecosystems. Knutsen et al. (2004) detailed how continuous water masses facilitate the dispersal of adult individuals and pelagic eggs and larvae by passive drift, and these conditions are quite suitable for many species including crustaceans, fish, birds and marine mammals. A few of the most common species seen in this area are cod (*Gadus morhua*), whiting (*Merlangius merlangus*), haddock (*Melanogrammus aeglefinus*), corkwing wrasse (*Symphodus melops*), and saithe (*Pollachius virens*). The composition of the habitats themselves are also unique, and the highly complex seascapes positively influence the abundance of organisms that can utilize multiple habitats. However, even less complex seascapes are conducive for juvenile assemblages in the summer and still prove to be ecologically useful (Staveley et al., 2016). While these habitats can be a thriving home, they are also very sensitive to trophic cascades and other impacts on the food web. Staveley et al. (2016) found that a decrease in larger predatory fish led to an increase in smaller predator species in shallow-water seagrass areas, causing cascading trophic level effects, and Baden et al. (2012) found that overgrowth by filamentous algae reduces seagrass growth as a result of top-down processes caused by the decline in top predators. These top predators include the saithe, whiting and cod in these types of vegetation-rich habitats (Wennhage & Pihl, 2002). Seagrass and seaweed habitats are diversity hotspots especially in shallow coastal waters, as they provide food and shelter for countless species (Östman et al., 2016).

1.2 The importance and impact of Atlantic cod (*Gadus morhua*)

1.2.1 Ecology and economical influence

Atlantic cod (*Gadus morhua*) is one of the most prominent and economically important species for Norway, given its large offshore spawning populations (Knutsen et al., 2004). Fromentin et al. (2000) detailed this species life history quite extensively in their publication. The cod in this area usually spawn in early March, with the juveniles settling to the ocean floor when they are about 3–5 cm and staying in more shallow and more sheltered areas, with a large portion of the North Sea juveniles settling on the Norwegian Sea shelf (Huserbraaten et al., 2018). They can live as long as 12 years, but very few survive to even five years which may explain why coastal individuals mature much more quickly than other cod populations. The majority of the spawning stock biomass is cod aged two to three years, given that any individuals four years or older have a higher mortality rate. Older juveniles and adult cod (>30 cm length) feed mostly on benthic invertebrates and fish in vegetated areas that have better resources (Freitas et al., 2015).

Many studies have examined the connection between cod habitat selection and physical factors such as temperature. They usually remain in shallow waters during autumn and spring and go into deeper water during summer and winter (Fromentin et al., 2000). Climate change through Anthropogenic Global Warming (AGW) is responsible for increases in these temperature profiles and resulting adjustments from the cod in these areas (~1 degree Celsius) and will be described in more detail in section 1.2.3 of this paper, “Anthropogenic impacts”.

1.2.2 The influence of subpopulations

Coastal cod have been genetically categorized into various sub-populations by many researchers, and these studies have revealed important factors that influence the species as a whole. Jorde et al. (2007) stated that coastal cod have a limited range of 30 kilometers or less, with multiple local populations along the Norwegian Skagerrak coast with a saline gradient as a barrier. Coastal cod are also highly sedentary and older individuals are highly site-attached as coastal topography may be linked to population spatial structure (Rogers et al., 2014), but Knutsen et al. (2004) found that offshore cod may influence coastal cod populations over large distances. Fromentin et al. (2000) found differences in cod between fjords along the Skagerrak coast, and that Skagerrak cod grow faster than northern and western cod, but slower than the ones in the North Sea and with very limited migration. Recent work suggests that coastal cod in Skagerrak and outer Oslofjord are made up by two co-existing but genomically unique 'ecotypes' - with potentially differing and persisting adaptations to local environmental conditions (Knutsen et al., 2018; Barth et al., 2019).

1.2.3 Anthropogenic impacts

Oslofjord is used heavily by humans, both for commercial and recreational purposes. Trawling for northern shrimps (*Pandalus borealis*) and recreational fishing are very common, along with recreational boating including overnight anchoring and mooring. Two national parks are established in the area, Ytre Hvaler and Færder national park. Multiple and consistent human activities eventually lead to changes in the environment both physical and biological, and almost always end up impacting the organisms that live in the used area. Because of this inevitability, it is crucial to determine the extent of the impact.

Overfishing in the seagrass areas that the juvenile cod rely on is a common theme among the studies that examine the impacts humans have on these environments. Baden et al. (2010) stated in their study that the area from southern Finland to western Sweden is affected by overfishing and resulting eutrophication, which Baden et al. elaborated on in their 2012 paper. They investigated how the overgrowth by filamentous algae which reduces seagrass growth can be explained by a top-down cascading effect caused by declines in top predators such as cod. They stated that the depletion of fish stocks has been identified as one of the most serious threats to marine ecosystems worldwide, and that overfishing seems to have an indirect effect on seagrass survival through trophic cascades. Barcelo et al. (2016) focused on the "critical nursery phase", and determined that juvenile communities have shifted several times, and currently are in a warm community period. In the North Sea specifically, small southerly species are increasing while large bodied northerly species are decreasing, and the changes in ranges and biodiversity are linked to rising temperatures. Moksnes et al. (2008) reiterated that seagrass communities are dominated by strong top-down processes controlling the aggregate biomass of mesograzers and macroalga.

The other main stressor on these environments is eutrophication, detailed in these studies and in Östman et al. (2016). The authors of this study stated that eutrophication favors fast-growing algae that quickly takes over seagrasses, and that management agencies need to consider measures to improve stocks of predatory fish to reduce mesopredators. Their suggestion is reinforced by their statement that the top-down effects are at least as important as nutrient effects for the structure of coastal food webs. This action would restore and conserve these habitats and improve their health in the long run.

One anthropogenic impact that is much more difficult to rectify is warming temperatures. Many species are already adapting, and not always positively, to climate change

in addition to already having natural adaptations to normal temperatures. Freitas et al. (2015) studied extensively the relationship between cod behavior and temperature changes. Habitat selection was greatly influenced by ocean temperature as the cod physiology is very sensitive to ambient temperature. Smaller cod appeared less sensitive to summer temperatures, but their growth also slows in the late summer. They stated that cod are a cold-water species found in temperatures ranging from -1.5 to 19°C but prefer a range of 9-16°C for optimal growth. Cod will select vegetation-heavy habitats in shallow areas if the temperature is below 16°C and avoid these areas when it is too warm. In the latter situation, they will search for deeper waters and/or rocky substrates populated by anemones and soft corals (Freitas et al., 2015; Freitas et al., 2021). The cod may also utilize areas such as sandy habitats as surface temperature increases. Future increases in ocean temperature are expected to further influence the spatial behavior of marine fish, potentially affecting individual fitness and population dynamics, and surface temperatures have already risen above their thresholds.

1.2.4 Recruitment

Recruitment is one of the most important stages in an ecological community as it influences the overall longevity of a population, and therefore the habitats for the recruits are equally as important. Understanding and then protecting both the recruits and their conditions is critical to recruitment success (Perry et al., 2018). Cod eggs are found in greater abundance in shallow waters (<15 m) flowing away from the ocean where the currents are gentler (Ciannelli et al., 2010). Young juveniles are planktonic feeders and rely on eelgrass meadows in coastal areas to access food and safety from predators (Freitas et al., 2015). Recruitment in 0-group cod in these waters is determined mainly between late June and mid-August, after the cod have settled, and had a better chance of survival in the presence of large copepods and 0-

group gobies (Johannessen et al., 2011). Coastal populations in the Skagerrak appear largely self-recruiting, as indicated by their partial (i.e., weak) genetic distinctness, but seem partly open to (and most likely receive) recruits from offshore sources in the North Sea. (Jorde et al., 2017).

Barcelo et al. (2016) stated that the juvenile nursery habitats provide a source for the adult stocks but are more sensitive to changes brought on by climate change. Recruitment failure or even total collapses can quickly spell the end of a once-thriving population, and many researchers have investigated both the rates of collapse and the implications for the species' ecosystem. Johannessen et al. (2011) found concurrent recruitment failure in gadoids (including cod, whiting and pollack) and changes in the plankton community along the Norwegian Skagerrak coast after 2002, which could have affected the study area of this project. Recruits can have a strong impact on their immediate area as shown in Moksnes et al. (2008). The authors of that paper determined that the seagrass community is dominated by strong top-down processes, and that overexploitation of gadoid fish (such as cod) may be linked to increased macroalgal blooms and loss of eelgrass in the area. This could prove to be detrimental both to the juveniles who rely on these seagrasses to settle after hatching, since habitat has a significant effect on recruitment (Rogers et al., 2014), and adults of fish assemblages on rocky bottoms that predominantly rely on food found in vegetation (Wennhage and Pihl, 2002). However, the recruitment could also be affected by human activities as demersal fish catches have been dominated by immature fish and cod especially showed changes in size (Svedaang 2003). Elliott et al. (2017) determined that cod were most abundant in shallow, sheltered areas composed of gravel, while haddock and whiting preferred softer substrate like sand and mud. Both cod and whiting were positively related to

the diversity of epibenthic and demersal fauna. In their article, they examined the effects of substratum type on juvenile gadoid abundance.

1.3 Stereo BRUV (Baited Remote Underwater Video)

1.3.1 What is Baited Remote Underwater Video?

Stereo Baited Remote Underwater Video (BRUV) is a useful tool for implementing studies in benthic habitats. The rig itself consists of two GoPro cameras mounted onto a metal frame (base bar) in order to film left and right video. In the middle of the rig, an arm extends outward, and at the end of the arm is a mesh bag containing bait - in this study frozen Atlantic mackerel (*Scomber scombrus*) was used. It is one of the less intrusive methods of investigating an ecosystem and therefore is becoming an increasingly popular method of choice (e.g., Goetze et al., 2021; Taylor et al., 2013). The BRUV's ability to measure several key aspects of the environment has proven to be useful in terms of management on multiple occasions.

1.3.2 Benefits of utilizing Stereo BRUV systems

Another reason that BRUV setups are becoming a new favorite in fieldwork is for its simplicity in collection and analysis. It is not difficult for a researcher to learn how to use and is cost-effective which removes many financial restraints on projects. It is well suited for protected areas due to its unobtrusive presence in the environment and specializes in examining abundance and density. This method does not have the depth and time restraints or bias as in scuba studies, and it is quite easy to train observers in its process both in collection and verification whenever needed (Stobart et al., 2015). Langlois et al. (2010) stated that

stereo video can reduce inter-observer variability, improve definition of the sample area and accuracy of length estimates and can be re-analyzed or validated with greater ease than a visual census. Their study determined that this method is suitable to obtain greater estimates on the biomass of generalist carnivores, which is why it is very suitable to examine the cod in this project. In the long run, BRUV studies are more cost-efficient, time-efficient and more reliable than methods such as diver based underwater visual census surveys.

1.3.3 Video based studies

Over the last several years, there has been an increase in studies that utilize video methods and highlight its usefulness and versatility. Caghlan et al. (2017) used BRUVs to determine that the density of large-bodied target species was higher inside closed fishery areas. Perry et al. (2018) focused on connectivity in coastal habitats with unbaited video systems and found that shallow-water fish communities were similar in adjacent habitats within a seascape, and that all habitats were dominated by juveniles. They discussed how shallow-water habitats are crucial to the health of the coastal seascape and how they contribute to both biodiversity and fishery stocks. Elliot et al. (2017) found depth and wave fetch to be the most influential factors in terms of suitable habitat, and that whiting abundance increased with increasing substratum extent. Their conclusion was that landscape effects should be considered as well, as it has implications for managing demersal fish populations.

2 Aims of the study

The overall aim of the project within which this study was carried out was to gather baseline biological data and create an inventory of fish assemblages in the fjord to assist management agencies in making decisions regarding protocols and propose options to form marine protected areas. With that dataset and goal in mind, and focusing on coastal cod as the study species - there are two subgoals that this study intends to accomplish:

- Scope for features of habitat and substrate that is conducive to harboring young-of-the-year Atlantic cod
- Determine factors with explanatory properties regarding occurrence and co-occurrence of cod recruits and juveniles of other species.

Specifically, the following hypotheses and research questions were tested:

- Does the presence of cod have an impact on the diversity? The null hypothesis (H0) was that the presence of cod has no impact on the diversity of the area, with its alternative hypothesis (HA) that cod presence does influence the diversity.
- Does the presence of cod have an impact on the overall Max N? The null hypothesis (H0) was that the presence of cod has no impact on the MaxN, with its alternative hypothesis that cod presence does influence MaxN.
- Does cod presence lead to different diversities?
- What specific factors predict cod presence in the environment between habitat type, depth and the diversity of the area?
- Which factors influence the overall Max N between cod presence, habitat type and depth?
- Is there a difference in the length measurements between the northern group of

stations and the southern group?

- Is there a difference in the diversities between the northern group of stations and the southern group?
- Is there a difference in the species abundances between the northern group of stations and the southern group?
- Is there a difference in the Max N means between the northern group, southern group and middle station?

3 **Materials and Methods**

3.1 Study sites

This study was conducted in the outer Oslofjord in Færder National Park. Figure 3.1 depicts a map of where the stations sampled in 2017 were located, created from the latitude and longitude data supplied for each location.

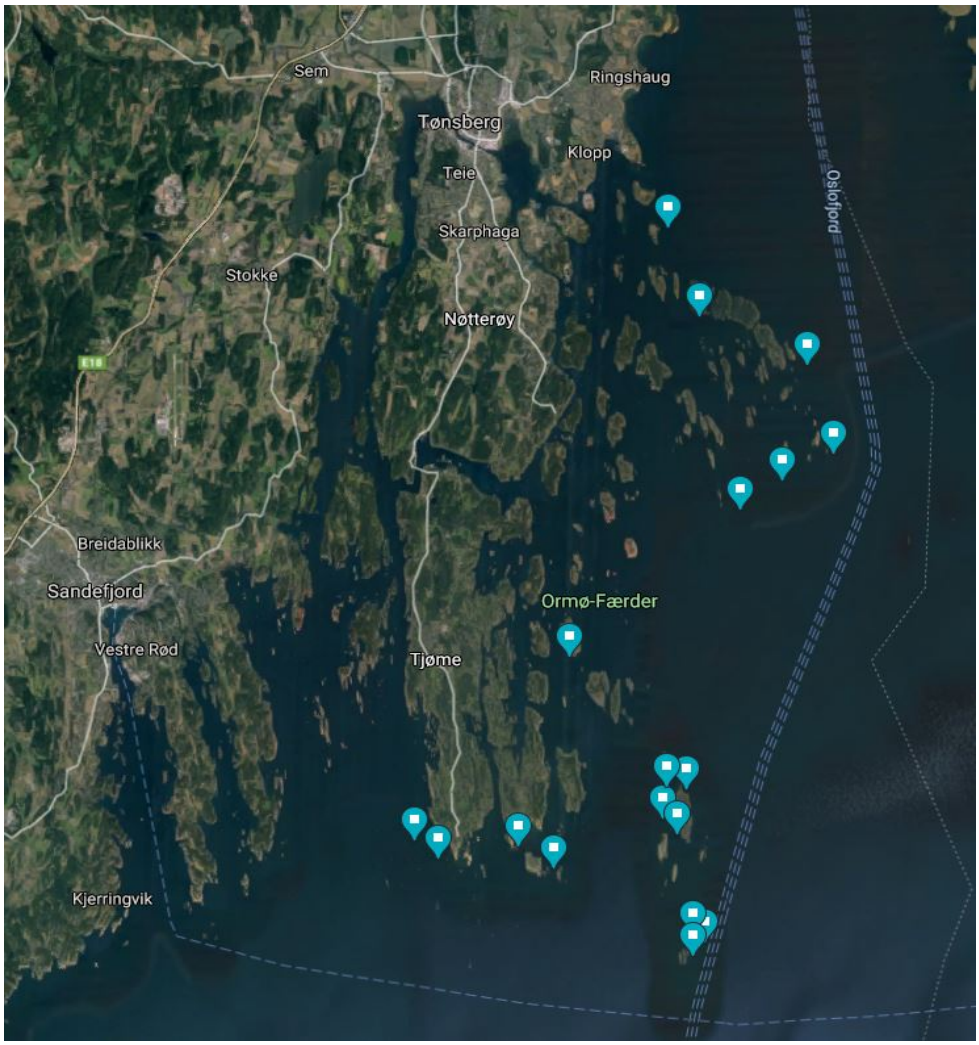

Figure 3.1: Map showing locations of the 18 stations used in this study, compiled from latitude and longitude data during BRUV drops

3.2 Sampling and equipment

3.2.1 Sampling procedure

Stereo video was recorded in August of 2017 during a three-day cruise. Thirty-five stations were visited in FNP, but only 18 had usable footage. Once at a location, the rig was lowered onto the seafloor and recorded one hour of video per station sampled. The depths ranged from 10 to 37 meters, with an average depth of 22 meters. Geographical data such as latitude and longitude were also noted.

3.2.2 Stereo Baited Remote Underwater Video (BRUV)

Figure 3.2 Stereo Baited Remote Underwater Video (BRUV) setup: a metal rig with two GoPro cameras with a bait arm in the center

The video was collected with a Baited Remote Underwater Video (BRUV) rig, which consists of two GoPro-cameras mounted on a metal frame with an arm extending from the center. At the end of the arm, an attachment is secured with bait made of frozen mackerel. Calibration files were written for each camera setup in order to complete the next step of data collection in EventMeasure. These files were created based on the principles of

photogrammetry, which is defined as the science of determining measurements from photos (University of Arizona, 1993). The main principle that the process is based upon is triangulation, using photos from multiple locations and establishing “lines of sight” from the cameras to the object in question. These lines are then mathematically calculated to create 3D points, like how human eyes process depth perception (Geodetic Systems, 2020).

The two main components of photogrammetry are photography (turning a 3D object into a 2D image) and metrology (turning a 2D image back into a 3D image). Measurements are not possible with just one image, hence the need for multiple photographs from multiple sources. These images produce a 2D (x and y) location of the object and are completed with a “z” point to convert it to 3D using a process called resection, which is the process of determining the final position and orientation of the cameras. In order to do this, the cameras’ positions and directions of aim are needed from three different coordinates to produce x, y and z. Even after all this effort, further calibration is needed in order to eliminate any errors (Geodetic Systems, 2020). Once these 3D points have been established, the program can calculate the length of the target object.

3.3 EventMeasure

The next part of the data collection was done through the SeaGIS program “EventMeasure”. All species that were visible in the video frame were identified and individuals were counted. The individual had to be fully in the frame and able to be identified to at least family level in order to be included. Once every organism was counted, each was measured if it was possible to do so. The organism could only be measured if it was fully visible on both cameras, hence the need for the calibration files. For the stations that were densely populated with the same individuals in the area for extended periods of time, the

author of this study chose to select a few key frames with the greatest number of individuals, highest number of species and the most potential to be measured. This was implemented into the protocol for two reasons. Firstly, to be as time efficient as possible so the author could complete the study within the next century. Secondly, the aspect of the data that was of the most interest was the MaxN, which is the maximum number of individuals of one species seen in a frame at any given time.

In addition to species identification and length measurements, the behavior of the organisms was also recorded. The categories available were “passing”, “attracted”, “feeding”, “scavenging”, “chase conspecific” (members of its own species), “chase other” and “guarding bait”. “Passing” behavior was defined as an organism that did not make a visible effort on camera to remain in the area and/or close to the rig and bait and would just swim through the frame. “Attracted” was defined as organisms that actively stayed in the area, especially if they

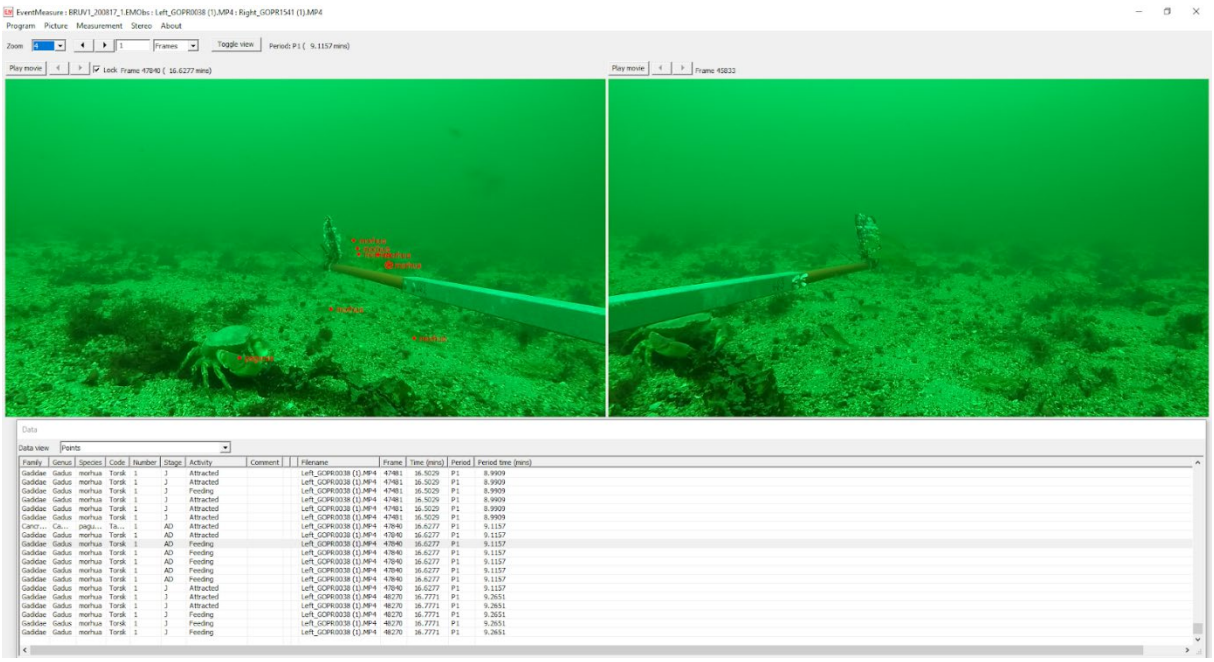

Figure 3.3 A screenshot of the EventMeasure program showing the left and right video clips with an example of how the data is presented

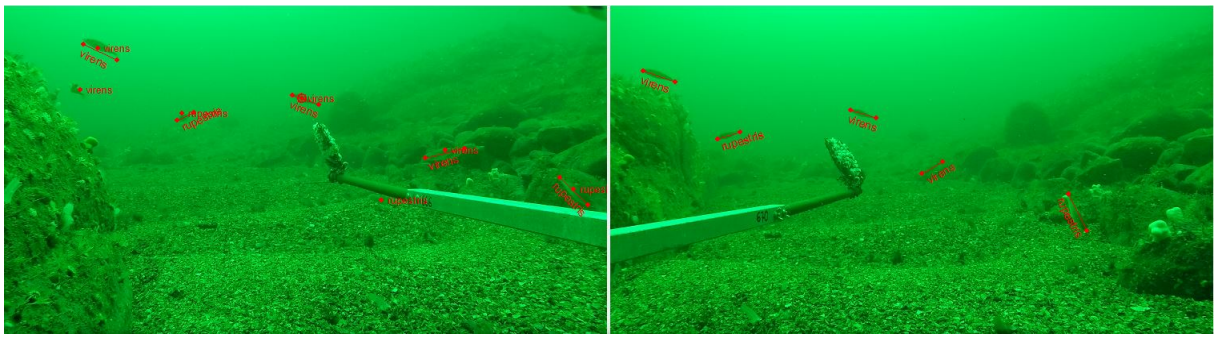

Figure 3.4 Screenshot of EventMeasure showing a school of saithe (*Pollachius virens*) and Labridae

were making significant efforts to swim against strong surges. “Feeding” was defined as an organism eating the bait, and only the bait. The author determined that “scavenging” was to be defined as an encounter (for example) where a piece of bait would be falling to the seafloor in the middle of a feeding frenzy with several organisms, and a separate individual would quickly descend upon the falling piece to take its meal “to go” and prevent its food from being taken. Scavenging also applied to any organisms who were feeding on anything other than the bait, i.e., crabs feasting on algae to the side of the frame. “Chase conspecific” and “chase other” are self-explanatory, and “guarding bait” was defined as an organism acting aggressively to defend the food for itself.

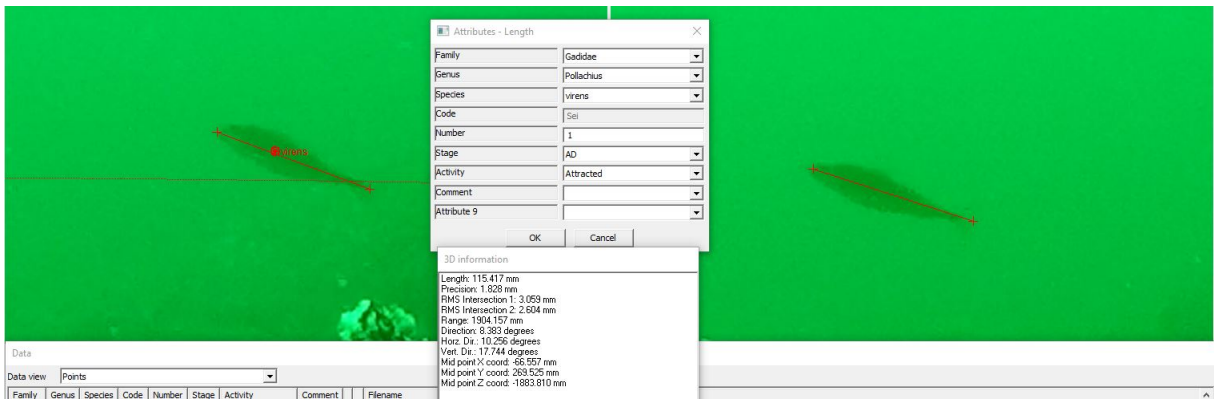

Figure 3.5 Screenshot of EventMeasure showing how the data is inputted with all of the available categories

Figure 3.6 Screenshot from EventMeasure showing an adult edible crab (*Cancer pagarus*) and a juvenile Atlantic cod (*Gadus morhua*) fondly nicknamed “the bully cod”
for being extremely territorial over the bait

Figure 3.7 Screenshot from EventMeasure showing two juvenile Atlantic cod (*Gadus morhua*)

Figure 3.8 Screenshot from EventMeasure showing an adult edible crab (*Cancer paragus*, left), juvenile Atlantic cod (*Gadus morhua,* far right) and a greater weever (*Trachinus draco*, right underneath bait arm)

3.4 Statistical analysis

3.4.1 Data treatment

The first step in the data analyses was to filter out data points that exceeded the Z-limit (horizontal distance from cameras) set by the author. While analyzing the videos with EventMeasure, the author of this paper determined that any organism beyond a Z-measurement of two meters was to be excluded from further analysis, as these organisms were extremely difficult to distinguish in lower quality videos. In the interest of scientific integrity, only organisms that could be positively identified to at least family level were included in the next stage. Factors such as behavior and habitat type were considered to fulfill that minimum requirement.

In BRUV studies, a proxy for abundance termed “MaxN” is used. MaxN is defined as the maximum number of individuals for a given species counted within the field of view at the same time (Harvey et al., 2013). The authors of this manuscript describe how it avoids double counts of individuals, and how the measures of cumulative MaxN are useful in the study of fish behavior and the influence of that behavior on abundance, which is key to this study given that one of the aims is to determine how cod affect other species. MaxN is calculated by using the number of fish within the focal distance of the camera and summing those counts (Campbell et al., 2015).

The data needed manual cleaning and “tidying” in Microsoft Excel. Fields that were not relevant to the analysis (date, tape reader, time, etc.) were removed and the dataset was divided into two subsets. One dataset contained observations that included a species identification, while the other was the observations that could only be identified at the family level with genus and species fields removed. This was done in order to make the dataset more workable in RStudio (R core team, 2021).

3.4.2 Preliminary exploration

The next step in the analyses was to check for normality in R for both the length data and the MaxN data. The length data was right-skewed and transformed using the logarithmic function, and the MaxN data did not change very much after transformation. After the work in R, the data was examined in Microsoft Excel to determine the most predominant family and species. The most predominant family was Gadidae (cod, haddock, whiting, pollock), with Atlantic cod being the most prevalent species out of the most predominant family (see Figure 4.1 in the *Results* section).

3.4.3 Data wrangling and tests in R

Once the dataset had been cleaned up in Excel, the data was "wrangled" and reformatted in R. A column was added in order to reflect the number of a particular species in every frame, since the species were counted individually, and frames appeared repeatedly. A presence/absence factor and diversity index column were also added to the dataset. The data were reformatted using the "reshape2" package in order to make it compatible with the "vegan" package to run the Shannon-Wiener diversity index calculations. The diversity index dataset was then merged with the adjusted count dataset, and that dataset was tested using a chi-square test of independence. Cod presence was used as the predictor variable for the chi-square test to see whether it had any impact on the presence of other species. Linear models were then run to determine predicting factors in cod presence. One model examined diversity in terms of cod presence/absence and cod counts, using the following model structure:

Equation 1: lm(formula = newdata$IndexNum ~ newdata$codpa + newdata$codcount, data = newdata)

where the response variable was the diversity of the area (newdata$IndexNum), and the predicting variables being cod presence (newdata$codpa) and cod abundance (newdata$codcount).

Second, the author ran a generalized linear model determining factors that predicted cod presence using the following model structure:

Equation 2: glm(formula = codpa ~ habitat + Depth + IndexNum, family = binomial, pl.data = completedat)

where the response variable is cod presence (codpa), and the predicting variables are habitat type, depth and diversity index (IndexNum).

A two-sample t-test was used to determine if there were any statistically significant differences between the northern and southern stations for length measurements, diversity and abundance. There were six northern stations and eleven southern stations, with the final station being in the center of these two areas near Ormø-Færder and therefore was treated as an outlier and not included in the t-tests.

For the Max N dataset, a chi square test was conducted in order to determine whether cod presence had any impact on the Max N. A linear model was then constructed to examine which factors had the strongest influence on Max N, the factors being cod presence, habitat and depth, using the following model structure:

Equation 3: lm(formula = MaxN ~ codpa + habitat + Depth, data = maxn)

Where the response variable is MaxN and the predicting variables are cod presence, habitat type and depth.

To examine variance in the means between stations, the stations were sorted into the northern and southern stations as used previously, along with the center outlier station as the

third group. This was done because the outlier station had the highest Max N results overall, and the author wished to investigate whether that had any significance.

A one-way ANOVA test was conducted, following a Tukey test to determine exactly what differences existed. There are three assumptions needed to run this test: each sample originated from a normally distributed population, each sample was independent of all other samples, and that the variance in each group is equal. The data were first examined with Q-Q plots and the Shapiro-Wilkes test to assess normality. The Shapiro-Wilkes test suggested that the data may not be normally distributed with a p-value <0.05, but the Q-Q plots showed that a large majority of the data fell along the regression line with some outliers. Even after a logarithmic transformation, the data did not improve very much. While this technically does not follow the assumption, a one-way ANOVA can withstand violations of the normality assumption if the sample size is large enough (Blanca et al., 2017). Given that the size was n= 96, the author concluded that breaking this assumption would not severely affect the results of the test.

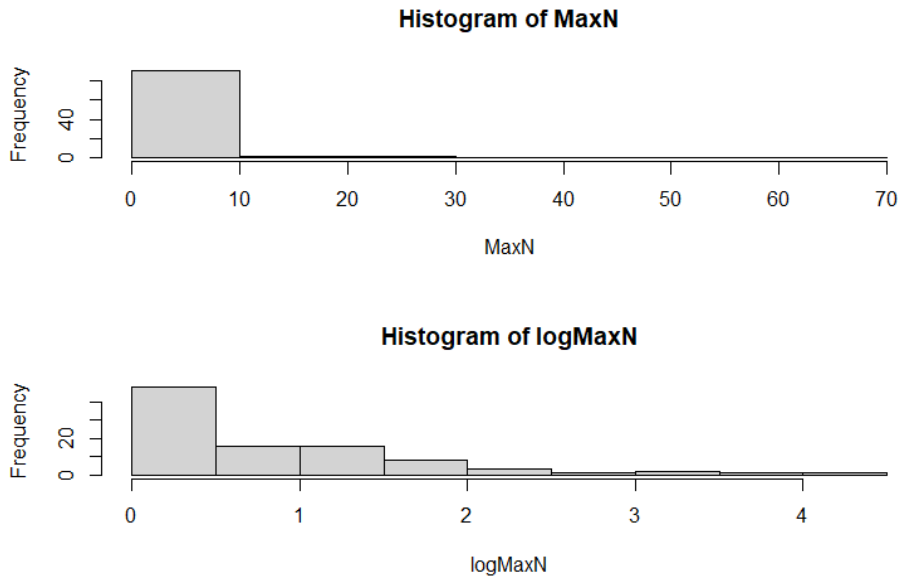

Figure 3.9 Max N data histogram raw and transformed

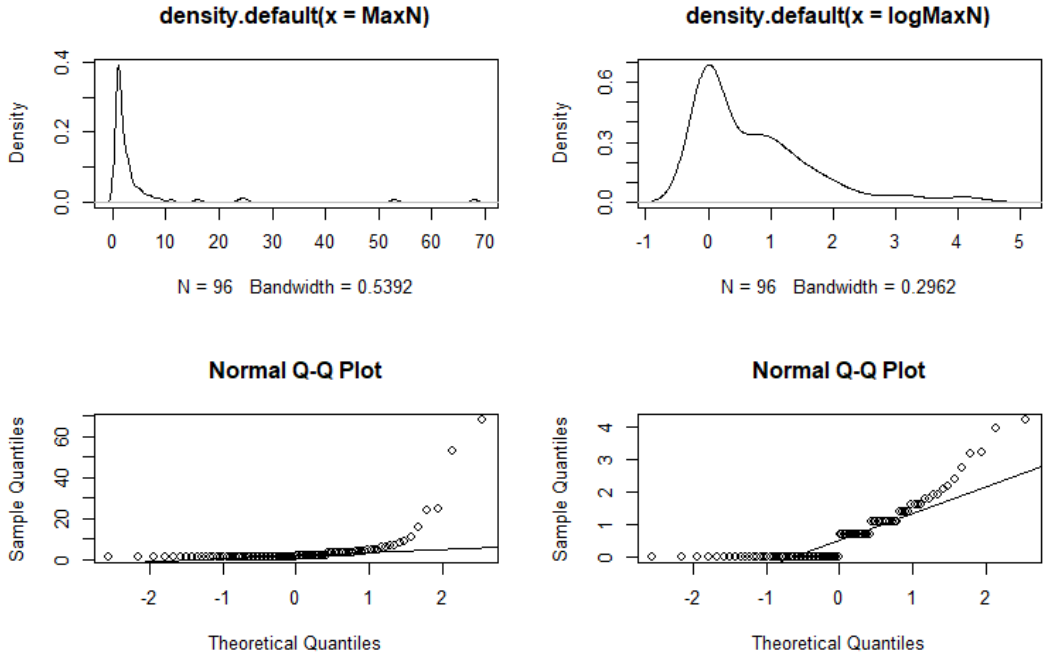

Figure 3.10 Density plots for Max N both raw and transformed

The second assumption of independent samples unfortunately has no formal test to evaluate the strength of the assumption. The test assumes that the observations were collected randomly and independent of all other observations. The author assumed that the assumption had been met given that the design of the data collection was randomized. It is impossible to determine precisely whether the observations were completely independent given that animals move around, so the author relied on the randomized procedure to state that the assumption holds true.

The final assumption is that the variances between the groups are equal, which was examined with boxplots and Levene's test since it is less sensitive to slightly abnormal distributions (Garson, 2012). The p-value of Levene's test was less than 0.05 suggesting that there may be significant differences between the variance, however the box plot showed the difference between north and south stations were minimal. The middle station had greater variance, which can be explained by the fact that the middle station had a MaxN of x=69 which would skew the results.

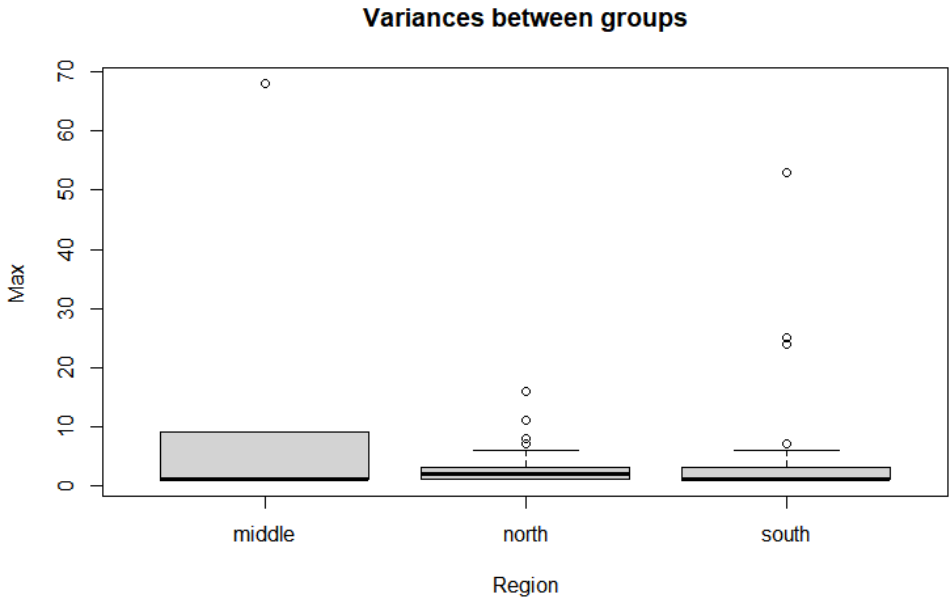

Figure 3.11 Boxplot showing variances in the means of the samples from the north, middle and south groups

The overall conclusion was that the assumptions were not met perfectly, but the violations were not severe enough to completely derail the one-way ANOVA test. The test is sturdy against mild to moderate violations (Garson, 2012), and it is up to the researcher whether to continue forward or use an alternative method such as the Kruskal-Wallis test (the non-parametric partner to ANOVA) to be conservative. The author chose to proceed with the ANOVA test, noting that the results should be interpreted with care. The structure was as follows:

Equation 4: aov(formula = maxn$MaxN ~ maxn$Region, data = maxn)

4 Results

4.1 Abundance

The first step in the analysis was to determine the most abundant family in the dataset, and the most abundant species within that family to determine whether we were on the right track examining cod. The most abundant family was Gadidae, followed by Labridae, Portunidae, Pleuronectidae and Cancridae as reflected in Figure 4.1, “Chart showing abundance by family across all stations”. Four other families (Rajidae, Triglidae, Trachinidae and Soleidae) made up the remaining 0.3%. Atlantic cod (*Gadus morhua)* was the most abundant species out of the Gadidae family, with saithe (*Pollachius virens*) being the next most abundant as shown in Figure 4.2.

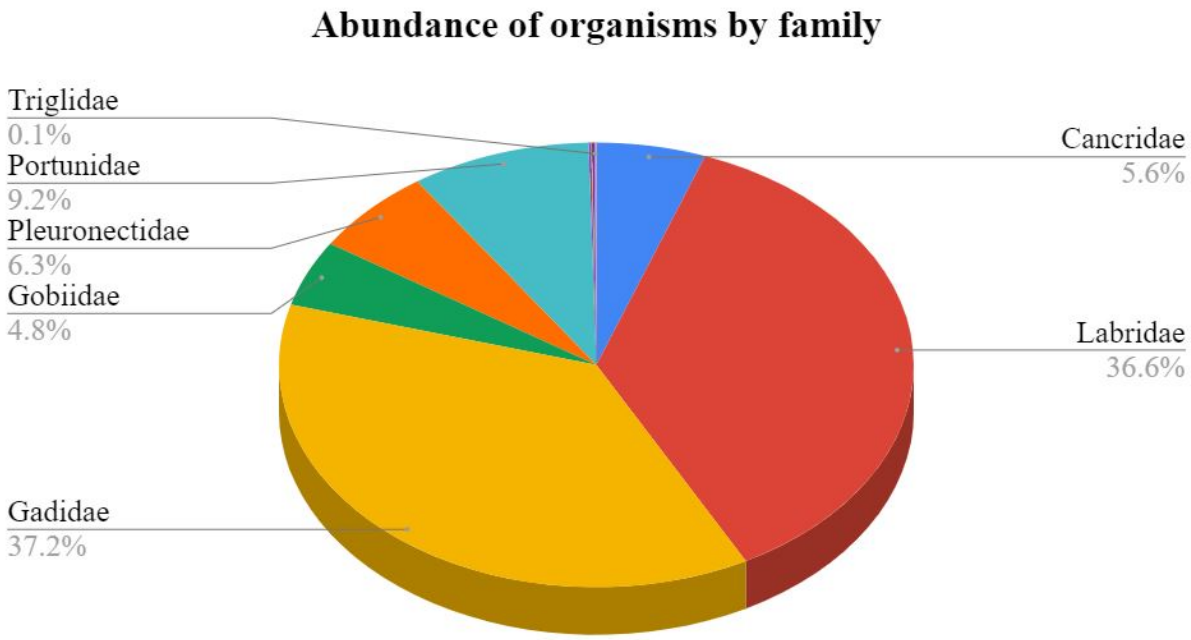

Figure 4.1: Chart showing abundance by family
from all stations

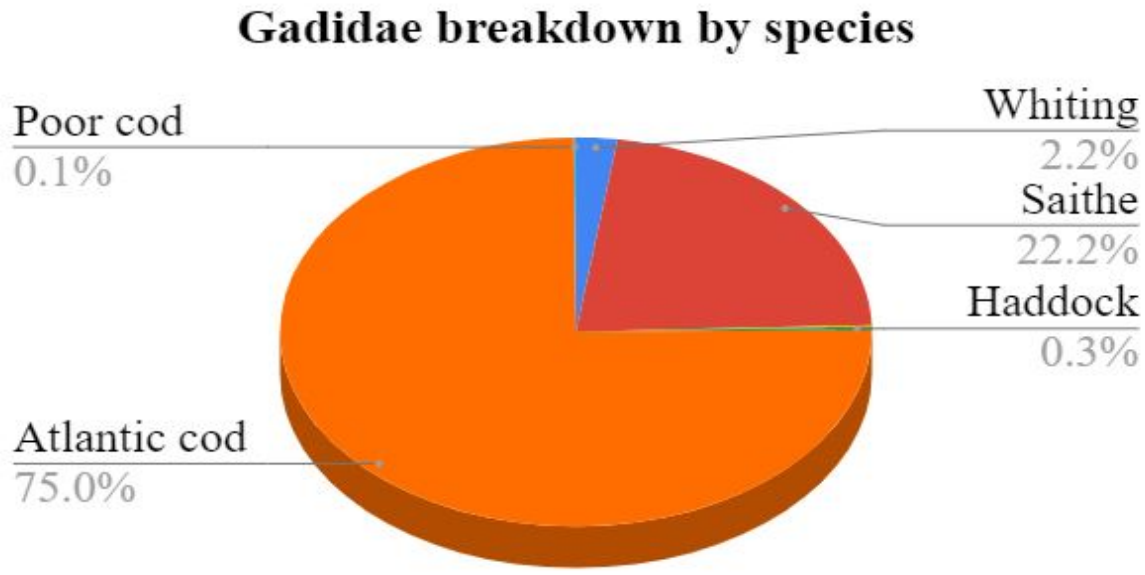

Figure 4.2 Chart showing the breakdown of the most predominant family
Gadidae by species

4.2 Length measurements distributions

Because the focus of this study is on juveniles of species in order to assess recruitment, the length distributions for all species and species individually were calculated using density plots in R. The code is attached in Appendix 1. The mean for length across all species was 118 mm as shown in Figure 4.3, and the mean for each species is shown on the density plots below. Table 4.1 shows all species measurements with minimum, maximum, mean and standard deviations.

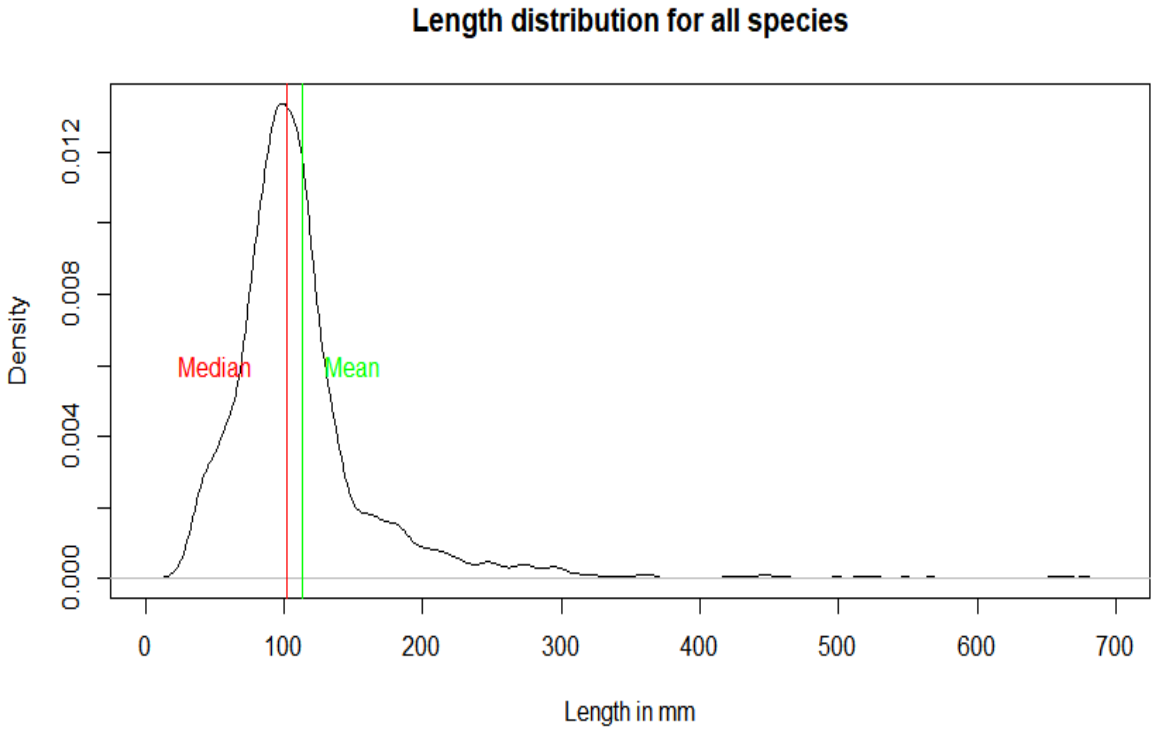

Figure 4.3: Density plot showing the length measurement distribution for all species across all stations

Table 4.1: Measurements (in mm) for all species with size range (minimum and maximum), mean and standard deviation. The precision measurements in EventMeasure varied greatly, but only realistic measurements were included in the dataset.

Species (fish)	Species (crustacean)	Minimum	Maximum	Mean	Standard Deviation
Ballan wrasse		95	218	148	52
Goldsinny wrasse		42	180	160	40
Cuckoo wrasse		115	515	213	87
Scale-rayed wrasse		90	137	111	16
Greater weever		299	313	306	10
Spotted ray		424	426	425	1
Corkwing wrasse		178	207	193	20
Whiting		73	95	110	18
Atlantic cod		34	498	104	41
Saithe		64	567	128	49
Haddock		78	102	90	17
Grey gurnard		206	248	227	29
Atlantic halibut		44	462	198	63

Species (fish)	Species (crustacean)	Minimum	Maximum	Mean	Standard Deviation
Pollack		111	111	111	0
Common dab		162	199	180	26
Sand goby		77	112	95	25
European flounder		179	292	232	40
Poor cod		130	130	130	0
Pomatoschistus sp.		35	337	101	47
	European green crab	34	213	120	65
	Edible crab	32	271	150	112

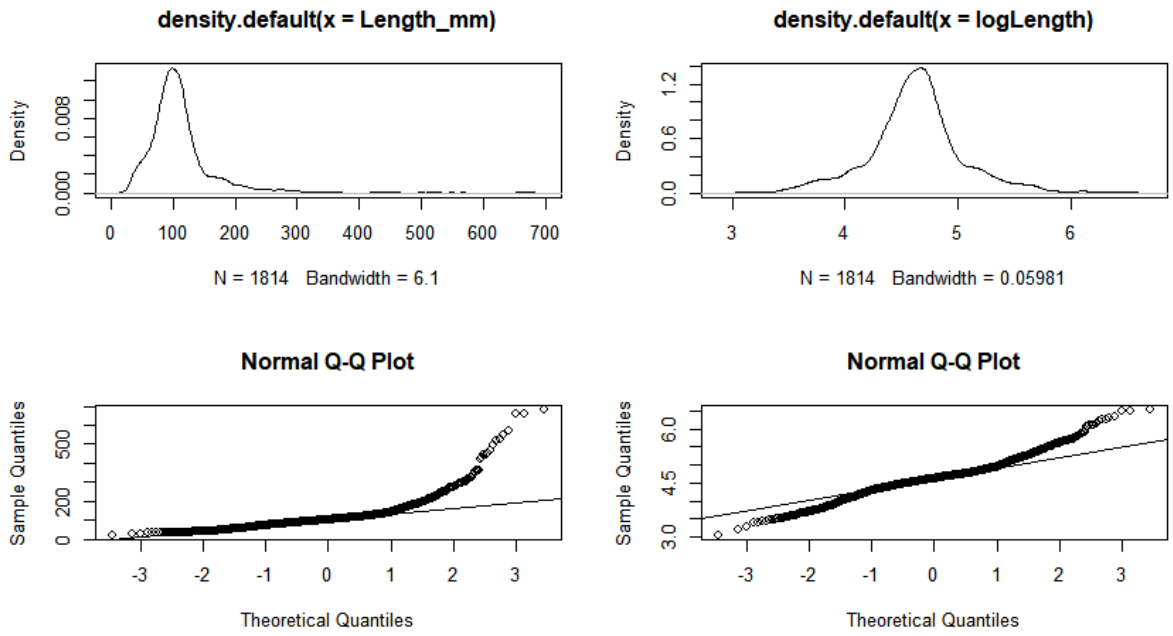

Figure 4.4: Q-Q plots for the length data set raw (left) and transformed with the logarithmic function (right)

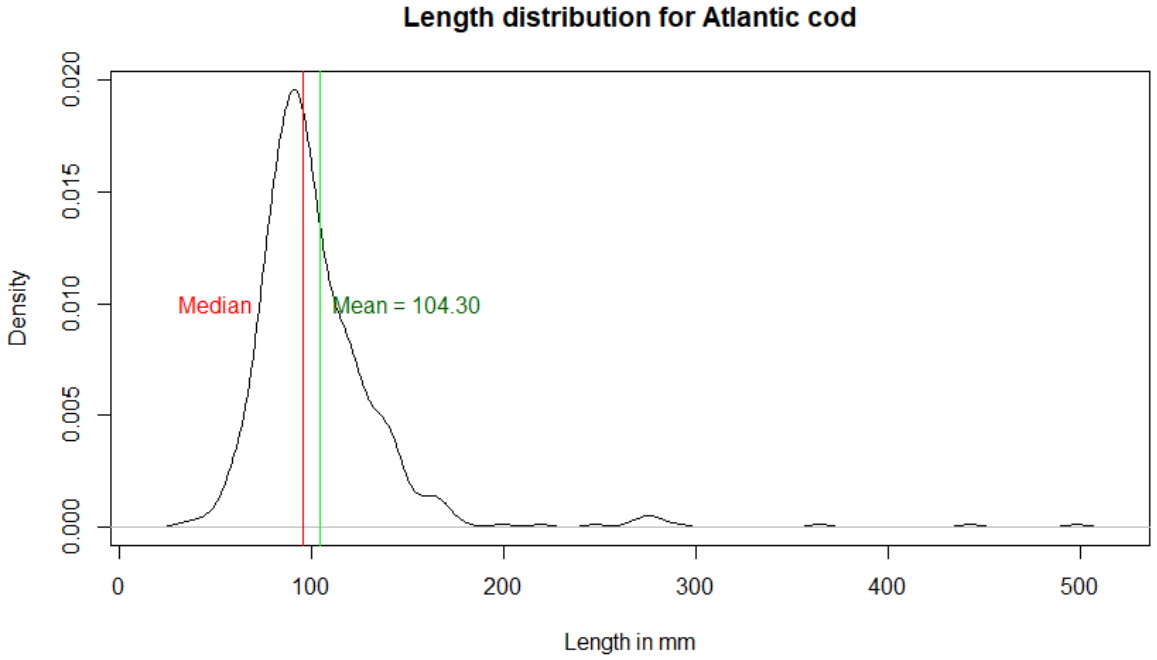

Figure 4.5: Length distribution of Atlantic cod (*Gadus morhua*) with a mean of 104 mm.

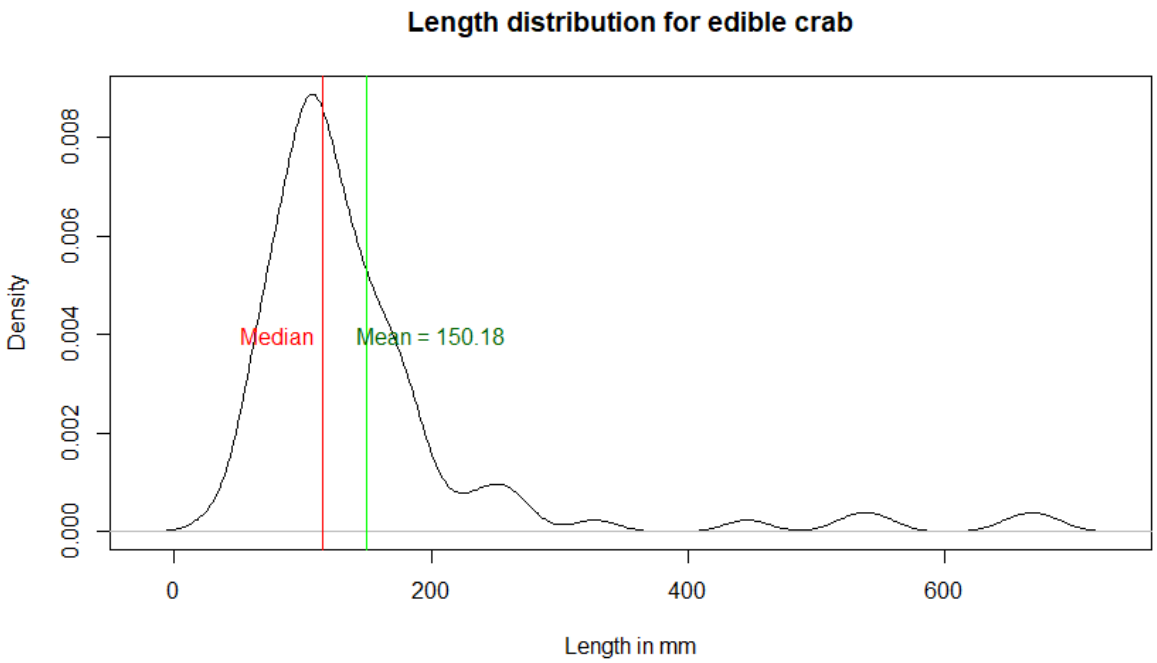

Figure 4.6: Length distribution for edible crab (*Cancer pagurus*) with a mean of 150 mm.

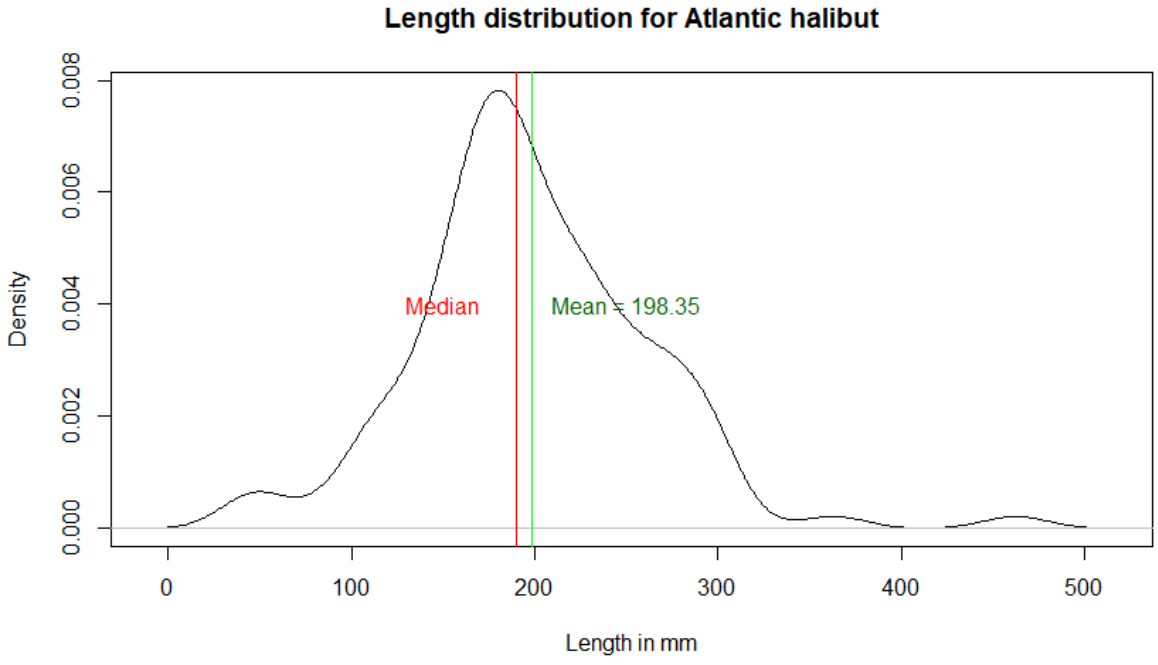

Figure 4.7: Length distribution for Atlantic halibut (*Hippoglossus hippoglossus*) with a mean of 198 mm.

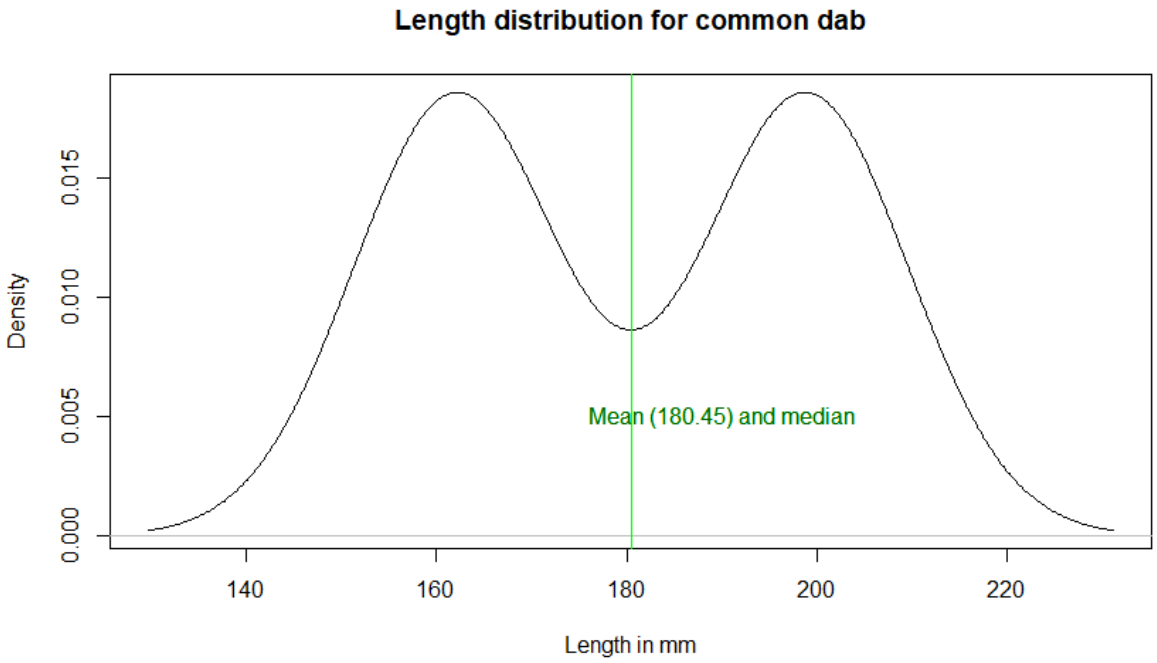

Figure 4.8: Length distribution for common dab (*Limnada limnada*) with a mean of 180 mm.

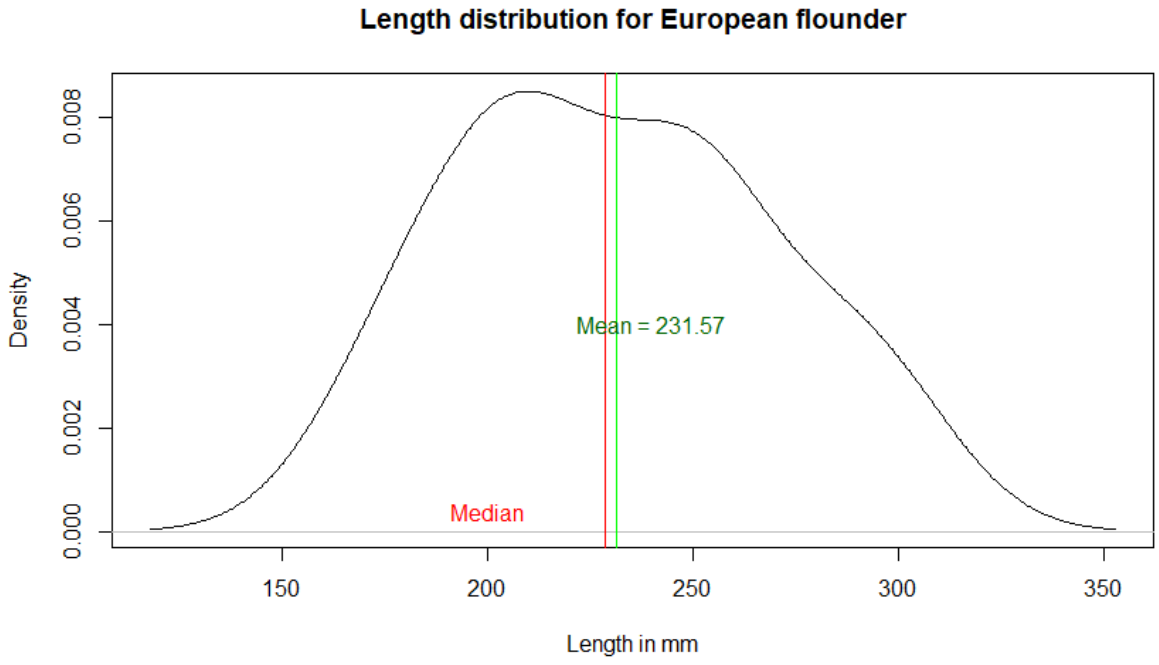

Figure 4.9: Length distribution for European flounder (*Platichthys flesus*) with a mean of 231 mm.

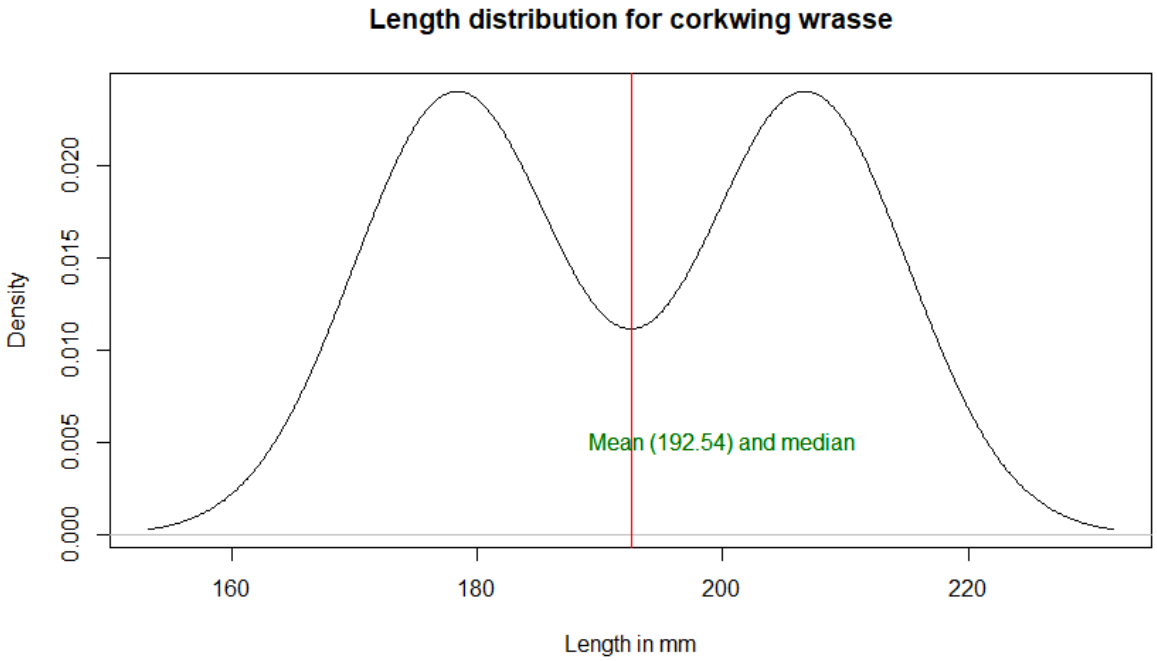

Figure 4.10 Length distribution for corkwing wrasse (*Symphodus melops*) with a mean of 192 mm.

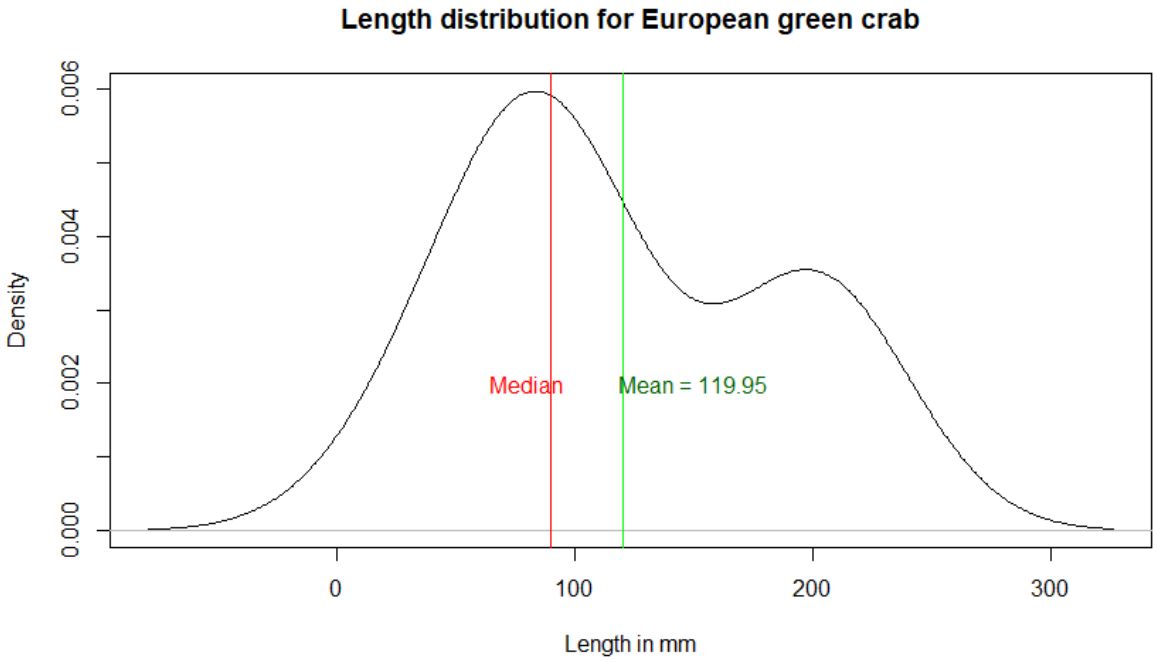

Figure 4.11 Length distribution for European green crab (*Carcinus maenas*) with a mean length of 119 mm.

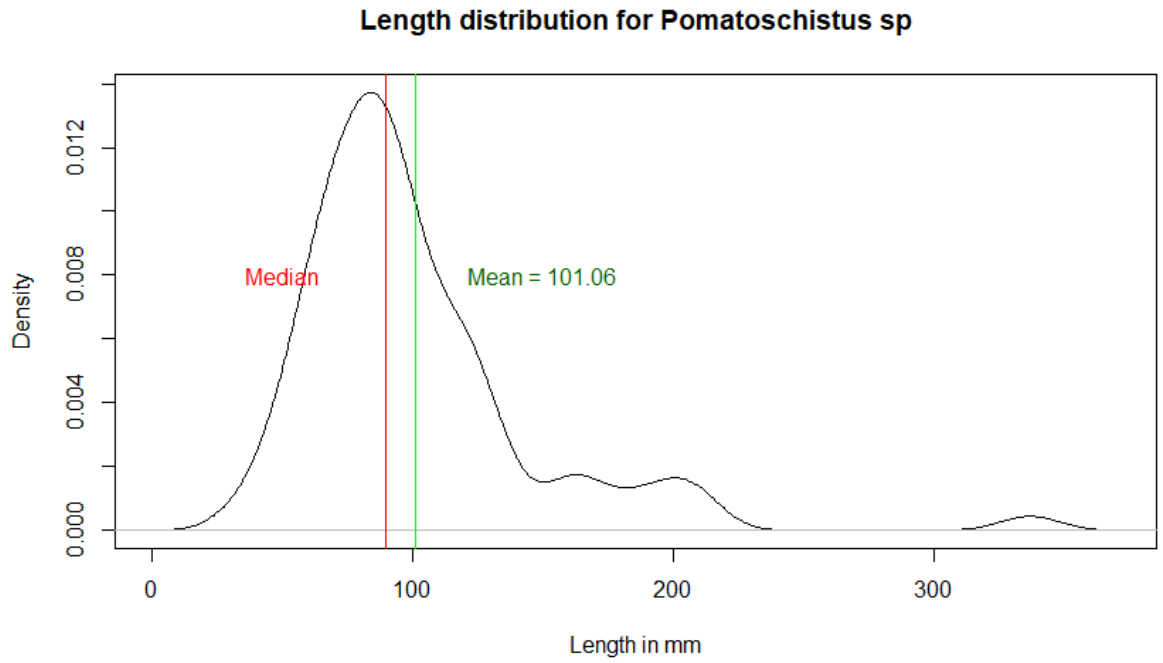

Figure 4.12 Length distribution for Pomatoschistus sp with a mean length of 101 mm.

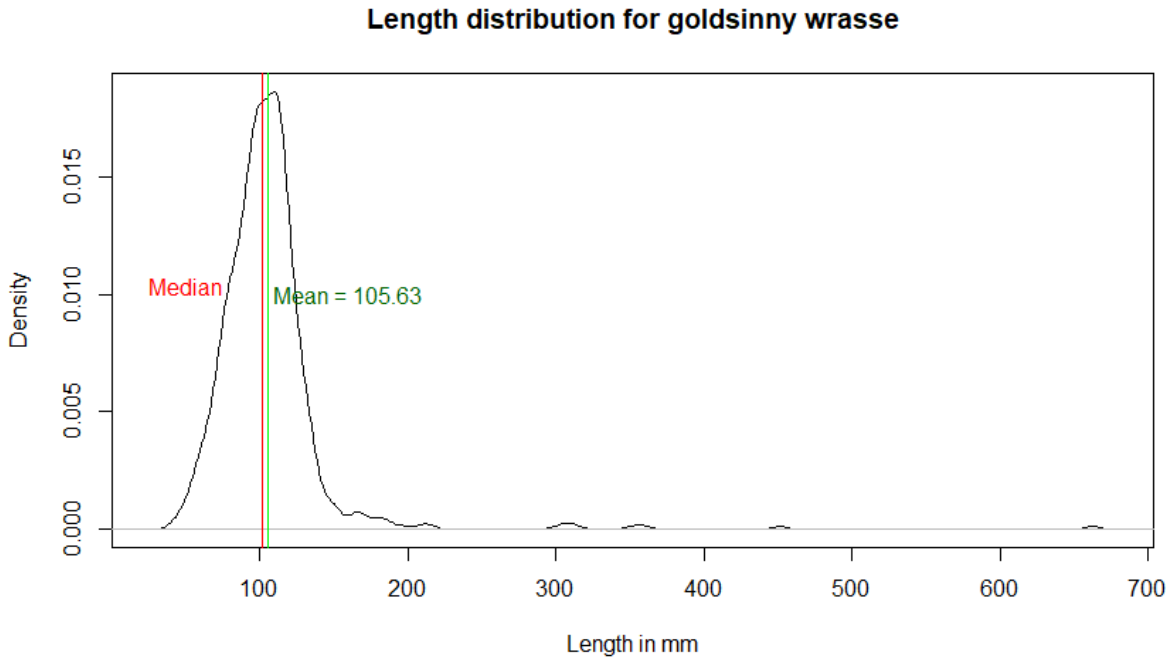

Figure 4.13 Length distribution for goldsinny wrasse (*Ctenolabrus rupestris*) with a mean length of 105 mm.

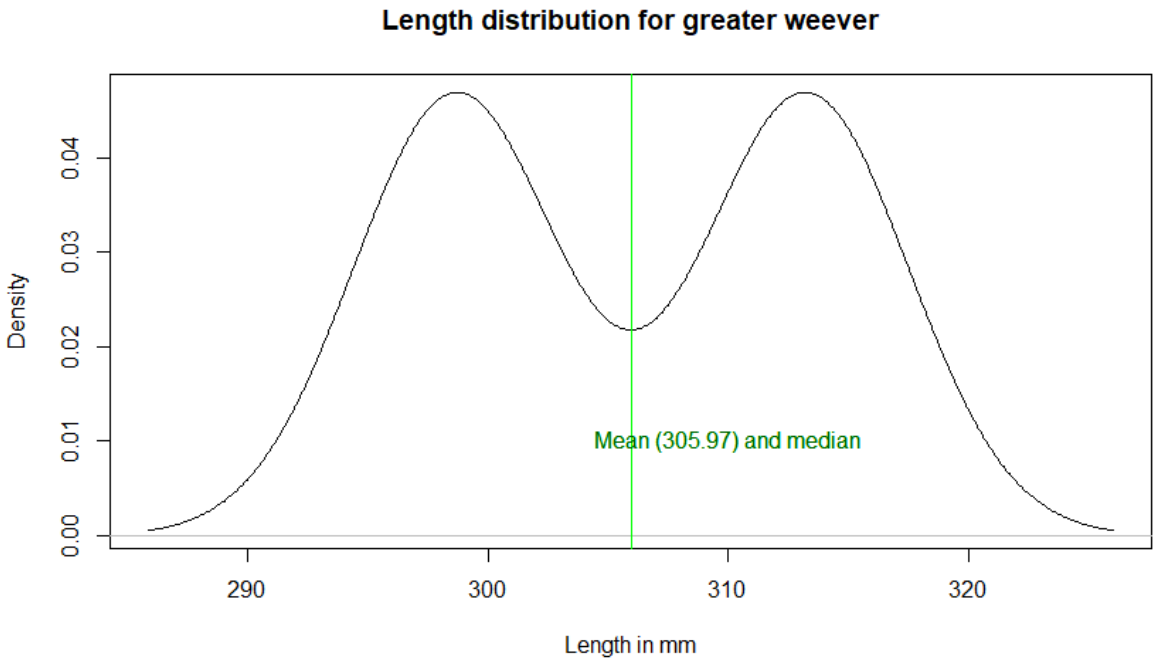

Figure 4.14 Length distribution for greater weever (*Trachinus draco*) with a mean length of 305 mm.

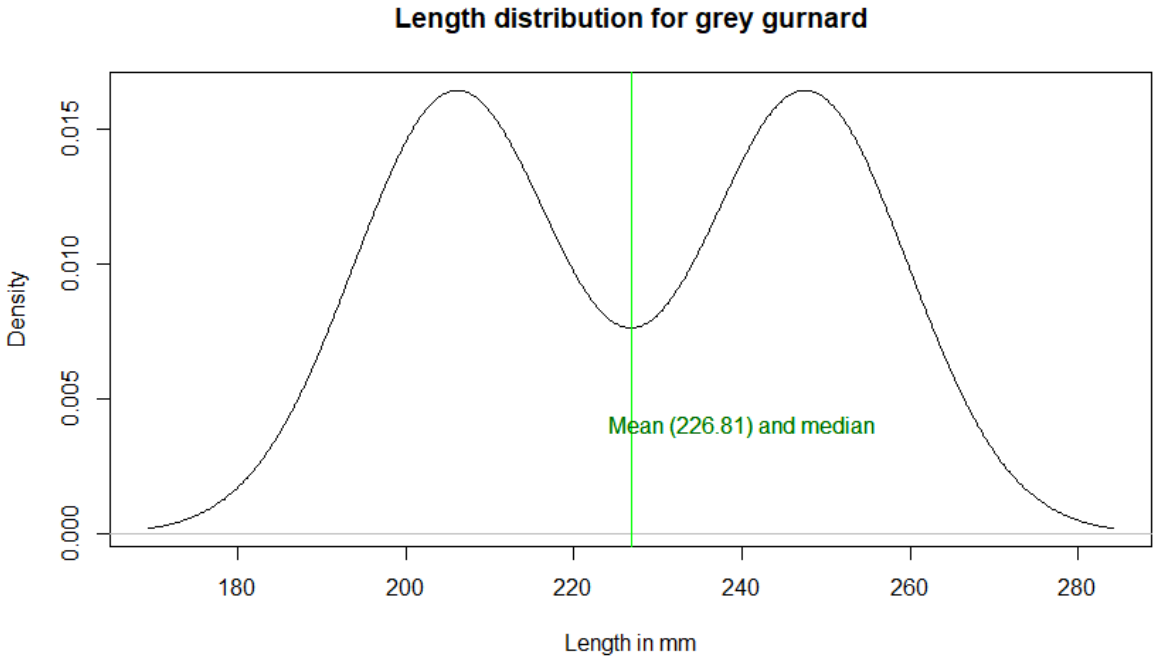

Figure 4.15 Length distribution for grey gurnard (*Eutrigla gurnardus*) with a mean length of 226 mm.

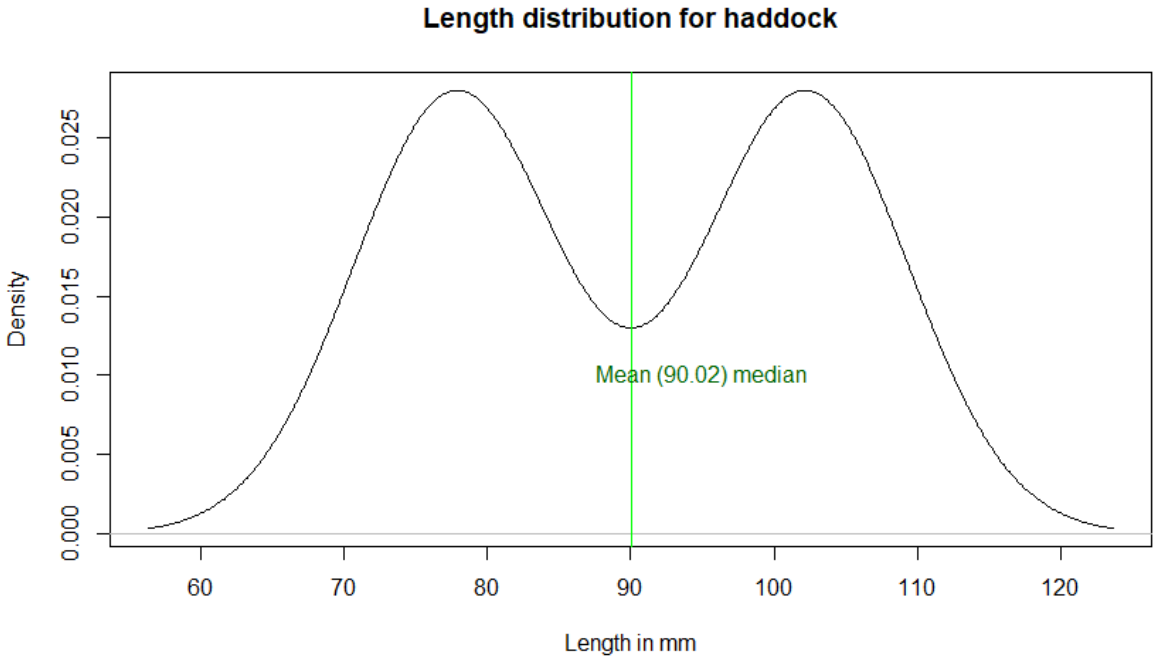

Figure 4.16 Length distribution for haddock (*Melanogrammus aeglefinus*) with a mean length of 90 mm.

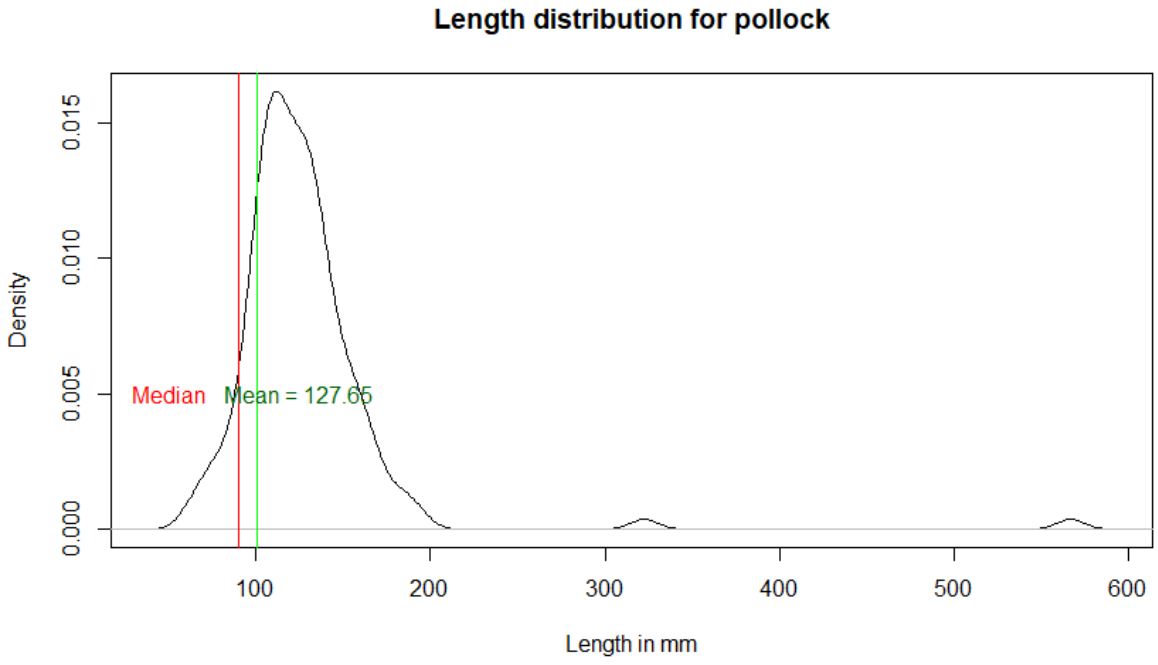

Figure 4.17 Length distribution for pollock, also known as saithe (*Pollachius virens*)
with a mean length of 127 mm.

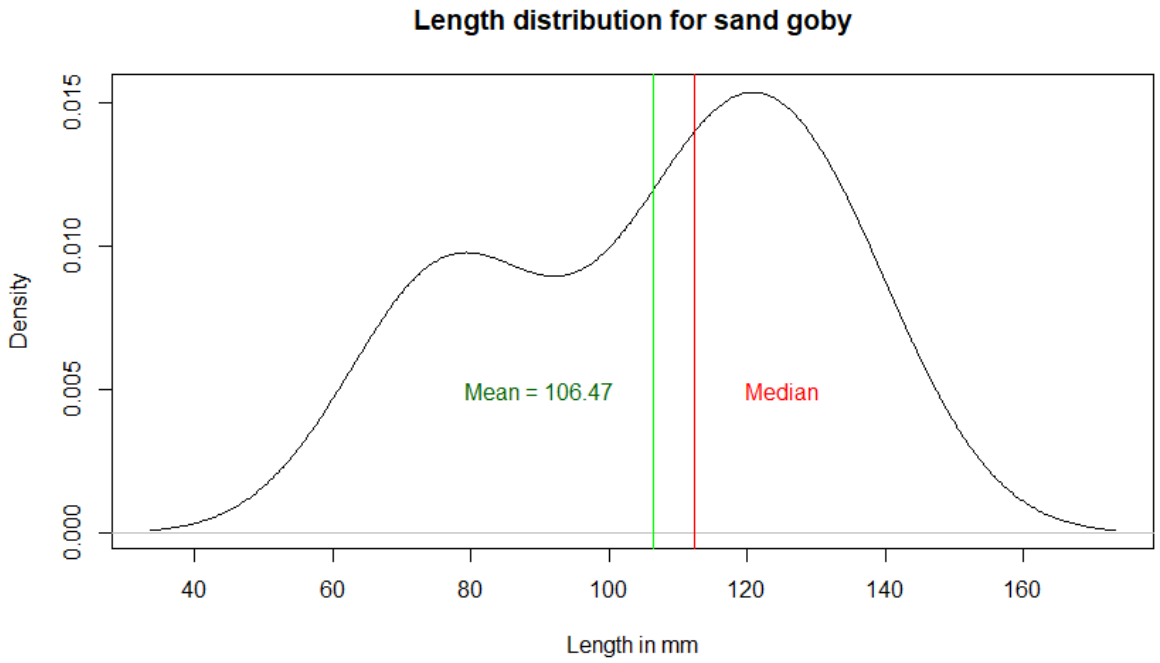

Figure 4.18 Length distribution for sand goby (*Pomatoschistus minutus*) with a mean length of 106 mm.

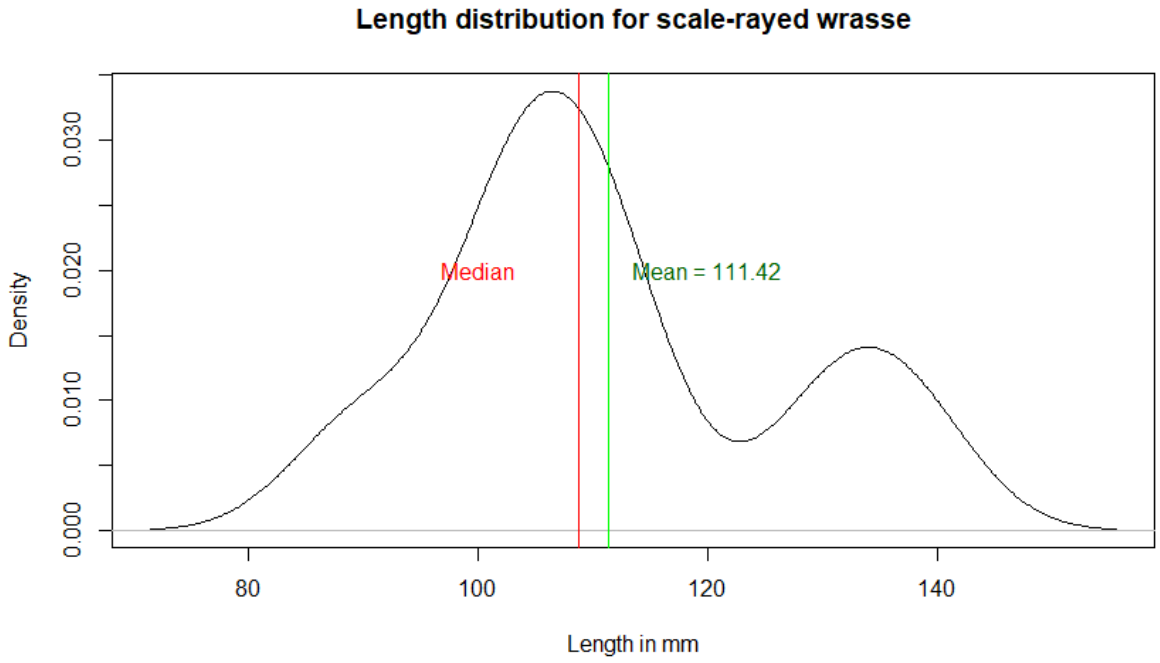

Figure 4.19 Length distribution for scale-rayed wrasse (*Acantholabrus palloni*) with a mean length of 111 mm.

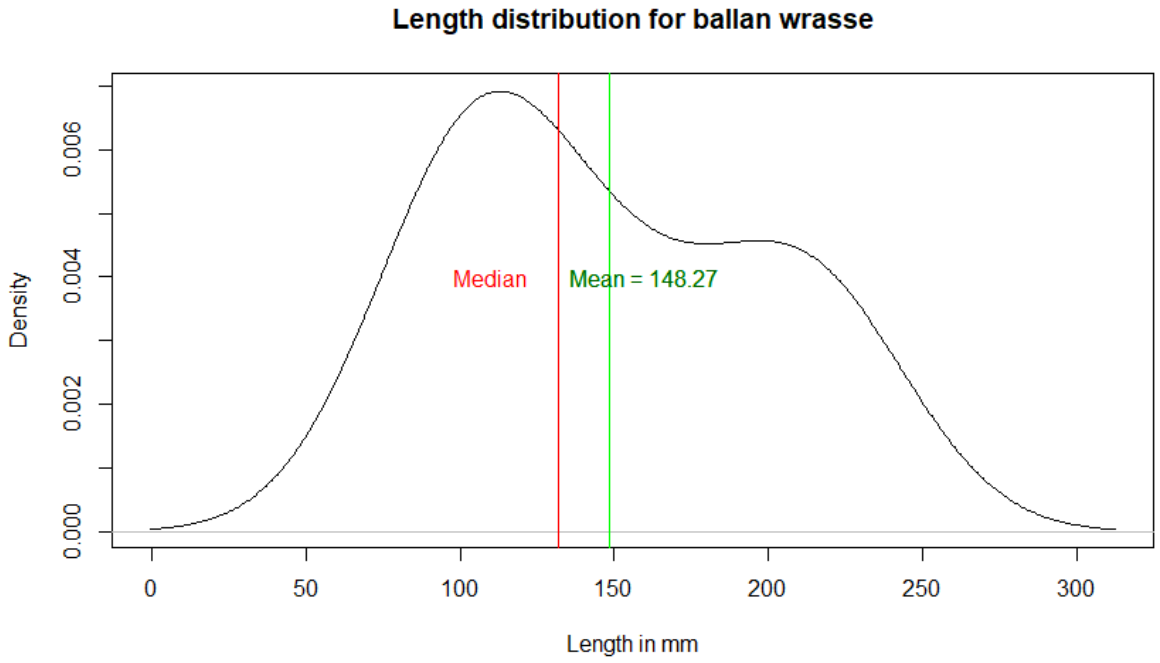

Figure 4.20 Length distribution for ballan wrasse (*Labris bergylta*) with a mean length of 148 mm.

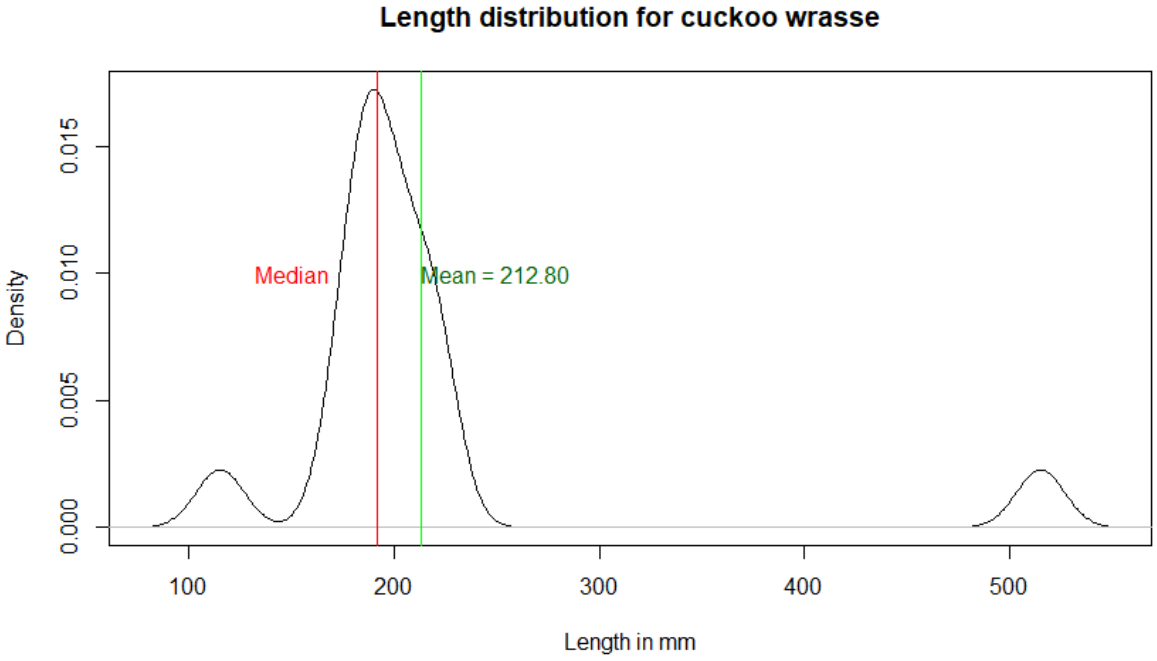

Figure 4.21 Length distribution for cuckoo wrasse (*Labrus mixtus*) with a mean length of 212 mm.

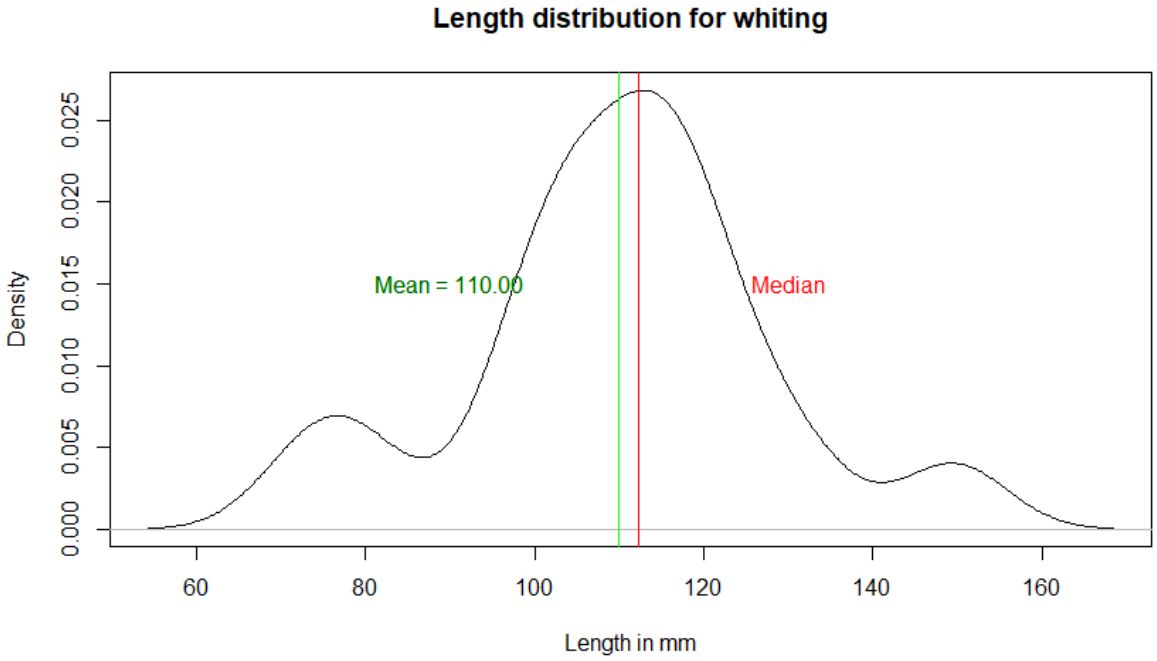

Figure 4.22 Length distribution for whiting (*Merlangius merlangus*) with a mean length of 110 mm.

4.3 Explanatory properties of cod presence

The logistic regression linear model aimed to predict suitable features for juvenile cod between habitat type, depth and the diversity of an area. The habitat type factor returned a prediction that there was a higher probability of finding cod on sand, but more variance on rocky habitats as shown in Figure 4.23. Figure 4.24 shows that there was a higher probability of cod presence at shallower depths, marked by a notable and rapid change at 20 meters. Figure 4.25 shows that there is a weak relationship between the diversity of an area and cod presence.

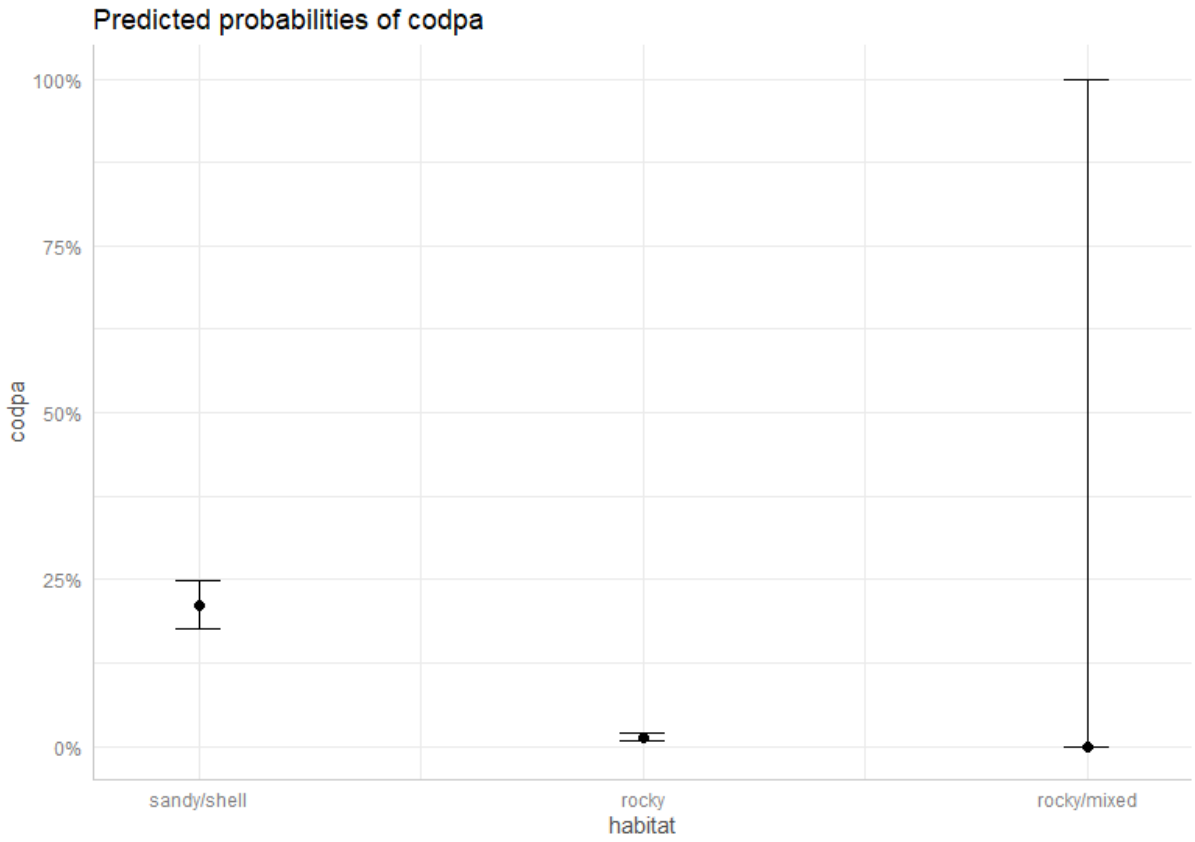

Figure 4.23: Logistic regression linear model output showing predictions of cod presence based on habitat type.

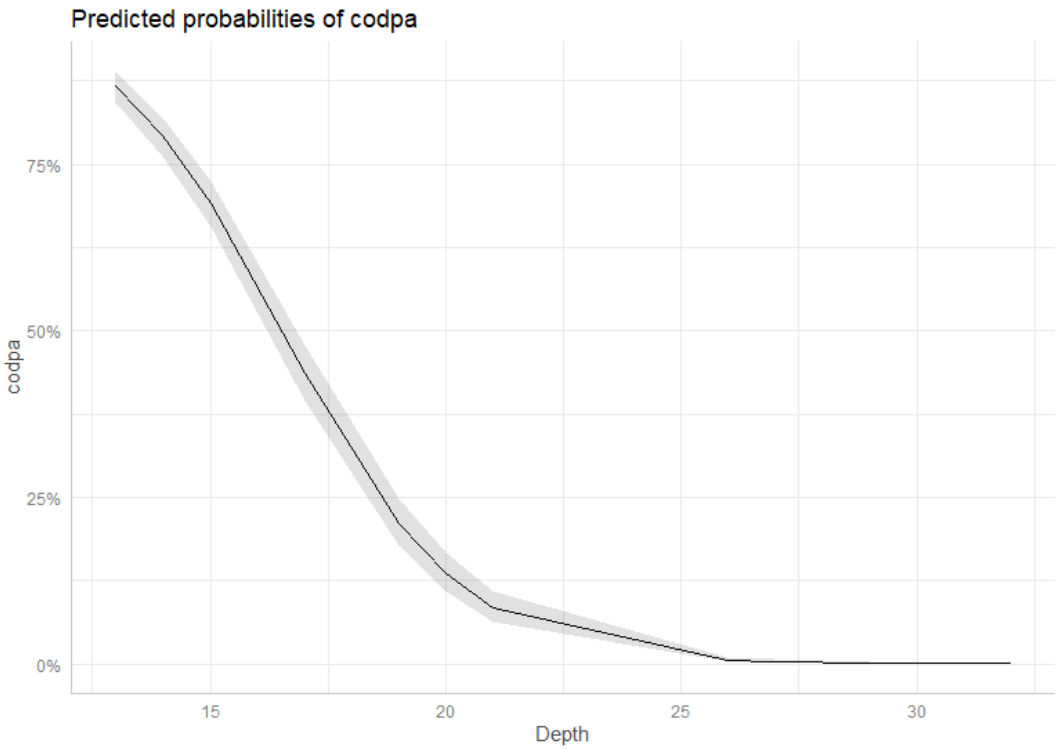

Figure 4.24: Logistic regression linear model output showing predictions for cod presence based on depth.

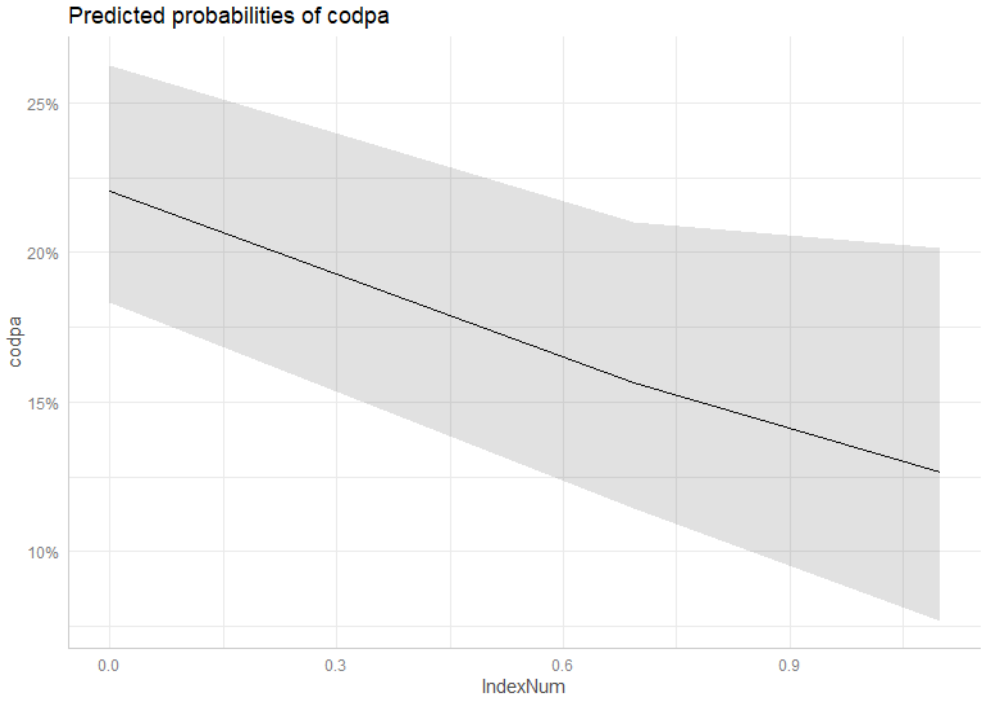

Figure 4.25 Logistic regression linear model output showing the probability of cod presence based on the diversity of the area.

4.4 Body size, diversity and habitat specificity

The chi-square test examining the relationship between cod presence and the diversity index had the null hypothesis that the presence of cod has no impact on the presence of other species, while the alternative hypothesis was that cod presence does influence the presence of other species. There was a significant relationship between cod presence and diversity, X^2 (2, N = 1645) = 8.57, p = 0.014, so the alternative hypothesis was accepted.

The linear regression model (Equation 1) aimed to determine if cod presence leads to different diversities, with the diversity index being predicted by the presence of cod and cod counts. Although the model had a low adjusted R^2 (0.017), it suggested a negative effect of cod abundance on species diversity (estimate: -0.026, $p < 0.001$).

A generalized linear model determining factors that predicted cod presence was run, with the predicting factors being habitat type, depth and diversity index (Equation 2), to determine the most influential factors affecting cod presence as reflected in Figures 4.23-4.25. Model selection based on Akaike's information criterion (AIC) was carried out to validate the most parsimonious model (Glen, 2018). Four variations of this model were constructed: a "global model" with all the factors included, one excluding diversity, one excluding depth and one excluding habitat types in order to validate parsimony by means of AIC scores. The results of these models are shown in Table 4.2 below. Model selection suggested that the global model containing all the factors had more support, improved by more than 2 AIC steps compared to the model with the second lowest AIC score. The habitat type "sandy/shell" had a positive effect on cod presence (estimate: 2.97061, $p < 0.0001$). Depth had a negative effect on cod presence (estimate: -0.53319, $p < 0.0001$), indicating that cod would be present with higher MaxN counts in the shallower stations sampled. Diversity also showed a negative effect on cod presence (estimate: -0.60749, $p < 0.05$), suggesting that

fewer species other than cod were registered at stations in which cod were present.

Table 4.2: Comparison of different variations of the model that aims to determine the influential factors affecting cod presence.

Model	**AIC score**
Global (all)	1245.4
Habitat type + depth	1248.4
Habitat type + diversity index	2283.6
Depth + diversity index	1517.6

Three t-tests were conducted to examine whether there was a difference in length measurements, abundance and diversity in the northern and southern groups. The results are presented in Table 4.3.

Table 4.3: Results of the three t-tests examining differences between the northern group of stations and the southern group.

Factor of interest	**Degrees of freedom**	***T* statistic**	**P-value**
Length measurements	1571	0.63083	0.5282
Abundance	630	0.38902	0.6974
Diversity	722	-2.9703	0.003073

There were no significant differences between north and south in terms of length measurements (means of 120.24 and 117.66 mm, respectively) and abundance (means of 0.10

and 0.10, respectively), but there was a significant difference in diversity (means of 1.72 and 2.29, respectively) with higher diversity in the southern group.

4.5 MaxN

The chi-square test examining the relationship between cod presence and the Max N had the null hypothesis that the presence of cod has no impact on the Max N, while the alternative hypothesis was that cod presence does influence the overall Max N. There was a significant relationship between cod presence and Max N, X^2 (14, N = 96) = 23.749, $p < 0.05$.

The linear regression model (Equation 3) aimed to determine which factors had the strongest influence on Max N, with Max N being predicted by cod presence, habitat and depth. Like Equation 1, the adjusted R^2 value was low (0.060) but suggested that a habitat that was "rocky/mixed" was the most significant factor with a positive effect on the MaxN (estimate 9.94, $p < 0.05$).

A one-way ANOVA test (Equation 4) was used to determine if there was any difference between the northern and southern stations, along with the center outlier station. The results are shown in Table 4.4.

Table 4.4: One-way ANOVA table examining differences in means between three groups.

	Sum of squares	df	Mean Square	F	p-value
maxn$Region	590	2	295.09	3.65	0.03
Residuals	7519	93	80.85		

The results showed a significant difference in the means with a p-value < 0.05, so a Tukey test was then conducted to identify the specific differences. The test produced the

following output shown in Table 4.5. The results show that while there is no significant difference between the northern and southern stations, there was a significant difference between the northern group and middle station, and the middle station and the southern group.

Table 4.5: Output from Tukey HSD test to show specific differences in the ANOVA test.

Grouping	Mean difference	Lower interval	Upper interval	P-value
north-middle	-10.63	-20.04	-1.22	0.02
south-middle	-9.62	-18.85	-0.38	0.04
south-north	1.02	-3.55	5.59	0.86

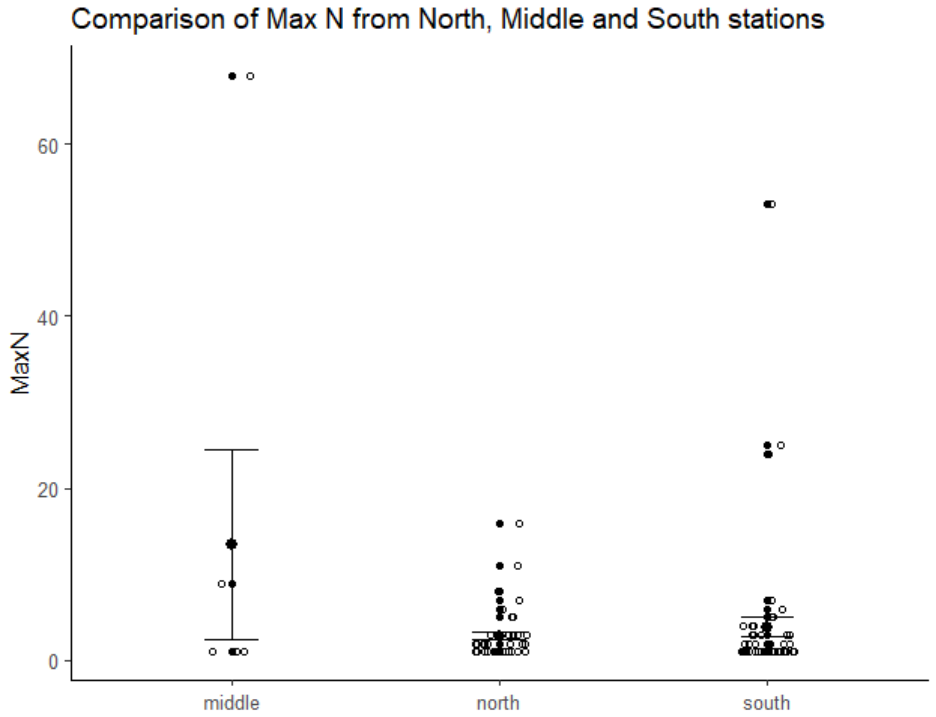

Figure 4.27 ANOVA plot comparing the Max N means from the northern group of stations, southern group of stations and middle outlier station

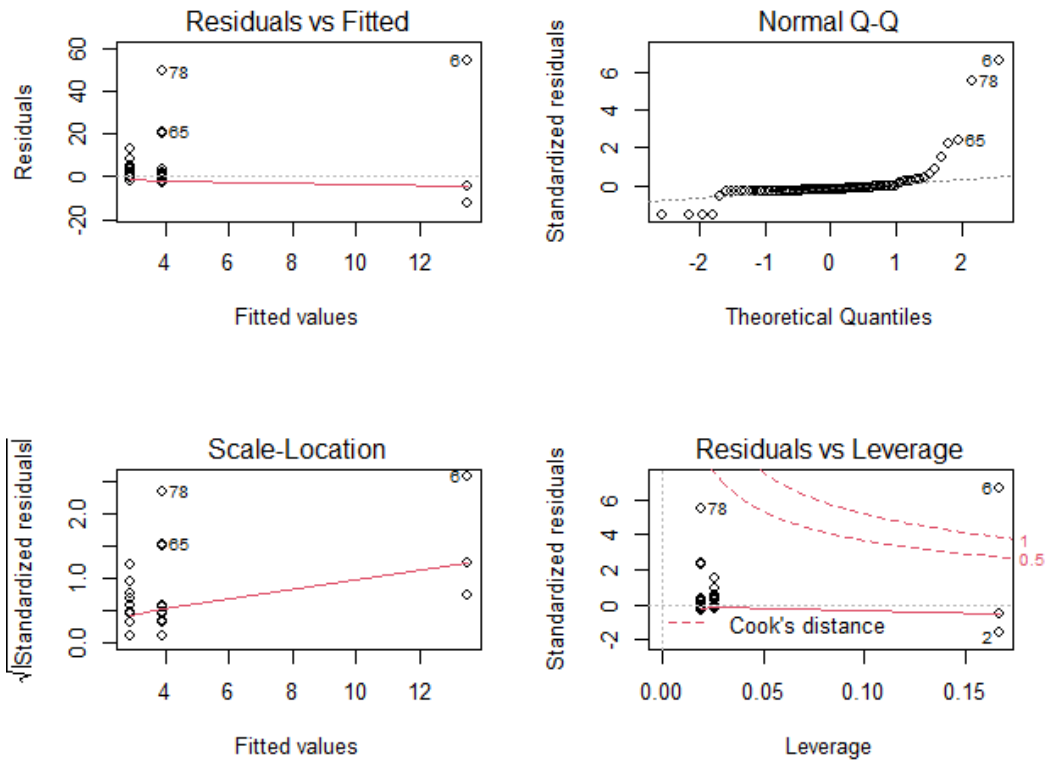

Figure 4.28 Q-Q plots for the ANOVA test comparing the Max N results for the northern group of stations, southern group of stations and middle outlier station

5 Discussion

5.1 Disclaimers and sources of error

There are several points regarding this study that the author wishes to discuss before proceeding to the remainder of this chapter. Firstly, this dataset does not focus on specific individuals so much as the population and ecosystem as a whole. Double counts were not a matter of concern since the primary interest was to determine MaxN. Secondly, there are multiple potential sources of error in terms of the data collection portion of this study,

specifically in terms of EventMeasure. While an extremely useful and interesting program, it did not seem to appreciate the author's time and effort, with multiple crashes a day and deleting hours or even days' worth of work. The author recovered the data to the best of her ability but felt it necessary to include in this section in the interest of scientific integrity.

Another major obstacle was in the calculations of length measurements in the EventMeasure program. Due to unknown causes, the guidelines provided by the calibration files were unable to sync in several stations and resulted in error messages for multiple measurements. In order to combat this, the author measured as many individuals as possible from a few well-populated frames in order to gather a subset for the measurements. The lengths of the organisms were not the highest priority for this study, so it did not doom the project, but it could have impacted the author's ability to assess the juveniles.

The quality of the video at several stations was heavily affected by the surge and current stirring up debris and substrate, which made species identification very difficult in those areas. The author attempted to at least identify the family level in order to produce data, but some could not be identified past that taxonomic level.

In terms of the plots presented in the Results section, density plots were presented rather than histograms because several species only had one or two points and therefore did not make a clear plot. Three species were not graphed given that they only had one data point and did not provide useful information via a plot: poor cod (*Trisopterus minutus*), Atlantic pollock (*Pollachius pollachius*) and spotted ray (*Raja montagui*). The species with very few points are the ones presented with a bimodal distribution, with the mean and median in the same location. The previously mentioned calibration issues most likely contributed to some puzzling results as shown in Figure 4.12 depicting the length distribution of goldsinny wrasse. This species is not capable of reaching lengths over 180 mm (FishBase, n.d.), so the odd

outliers reaching points as high as 700 mm are due to errors in the program itself. This error most likely caused the abnormal results for Figure 4.16 showing the length distribution for pollock (also known as saithe) as well.

5.2 Species size distributions and abundance

The size distributions suggested a fish species assemblage dominated by relatively small sized species and juveniles of larger species. The size distribution for cod confirmed that young-of-the-year juveniles were the most abundant age-group/ year class present in the study area in August 2017, although a few larger older cod were encountered. Even species such as whiting and pollock, which also can grow to be quite large, were overall quite small which only strengthens the findings of a juvenile-dominated community captured in areas that might be considered as nursery areas for these species. Very few species reflected an average size expected of adults of that species, such as sand goby and the greater weever.

The most abundant family was Gadidae, and the most abundant species was Atlantic cod (*Gadus morhua*). This is not surprising as 2017 (the year this data were collected) was an excellent year for cod recruits as documented in the IMR beach seine sampling monitoring carried out annually in September (IMR, unpublished data), hence it being the focus of this study. When considering the generalist diet in the cod's ecology and the fact they will eat just about anything, the next most abundant family being Labridae makes sense. Most labrids are too large to be an easy meal, and their size especially in comparison to juvenile cod provides them with a certain "safety net" so to speak.

Overall, the organisms in this study were on the smaller end with a mean of just 118.08 mm. Cod had a mean length of 104.30 mm, and nearly all the observations were juveniles. Only 18 observations were adults out of the 534 observations for cod altogether. This reinforces the finding that 2017 was an excellent year for cod recruitment and supported

by the literature that states that the juveniles have settled by mid-August. The data were collected mere weeks after this established timeline, and therefore the results are logical.

There were multiple t-tests conducted in order to compare the northern and southern stations for length, abundance and diversity. While the mean lengths between the areas was not very different at all (120.23 mm ad 117.66 mm respectively), nor the diversity (p-value >0.05), the abundances in these two areas were significantly different (p-value < 0.05 with means of 1.72 and 2.29 respectively). This is interesting because the literature states that juveniles tend to settle away from the open ocean in more sheltered areas due to gentler currents, while this data suggests otherwise. One possible explanation is that the southern stations are in Færder nasjonalpark, and the fact that it is a marine park may provide the organisms in and near the area better resources and stronger protection, although no specific regulations are in place to limit fishing in this part of the park. Alternatively, the view that sheltered or less exposed areas are better nursery habitat is biased by the sampling methods most commonly used in the past and warrants further study.

The ANOVA test examining the MaxN, and the following Tukey HSD test provided some very interesting results. Like the t-tests for the length data, the northern and southern groups had no significant differences but both groups were different from the middle outlier station. This could be because the middle station had the highest MaxN (x=69), but the fact that it did have the highest MaxN warrants further investigation to determine exactly why that is.

5.3 Impact of cod presence on other species

The chi-square test of independence on the length data showed that there is a relationship between the presence of cod and the diversity of the area. The linear model aimed to answer if cod presence led to different diversities to further investigate the previous question, with both cod presence and cod counts as predicting variables for the diversity index. The output returned that while the presence of cod was not significant, the actual number of cod in the area is what influences the diversity. The next question to answer was "what determines cod presence between habitat, depth and the diversity?" which was examined with the logistic regression model. The most significant factors were sandy habitats and depth, followed by diversity.

These findings are highly informative given that nearly all observations were juveniles as mentioned in 5.2. They are also quite logical when considering the literature review presented in the introduction (Freitas et al., 2015; Perry et al., 2018; Freitas et al., 2021). The literature described how cod, especially juveniles, utilize different depths and sandy habitats during warmer temperatures. It also stated that juveniles are less sensitive to warmer temperatures than the adults, and this data follows that statement given that the logistic regression model predicted finding the cod in shallower waters less than 20 meters, and that nearly all observations were juveniles. This also follows the finding that the growth of juvenile cod is slower in the summer given that the finding in this study was that most of the community was very small juveniles who were nowhere near what is considered a mature size.

5.4 Management implications

These results have significant implications for management agencies regarding protocols that involve juveniles of economically and ecologically important species. Agencies have been increasingly criticized in recent years for not incorporating this crucial life stage in their protocols, especially agencies that oversee large and busy fisheries. Although this work is limited, with few stations and one sampling season only, the findings suggest that shallow habitat throughout the park (north to south) is nursery habitat for cod and other species. Without protection for juveniles who eventually grow up to become part of the stock, eventually there will be no stock when it collapses for any number of reasons. The stock could become inbred and have a genetic collapse if too many mature individuals are removed from the population. Many populations have been maturing at younger ages and smaller sizes, which can lead to juveniles with decreased fitness and other complications. Studies have shown that in temperate waters, older and large fish produce studier offspring that grow three times faster than the offspring of younger fish, and survive starvation 2.5 times longer (Berkeley et al., 2004).

The findings discussed in section 5.2 and the literature regarding genetically distinguishing sub-populations could potentially be explained by the source-sink dynamics of the area. This is an ecological model to explain how the same species can have different populations in two separate habitats, with one being of high quality that encourages growth (the source) and one of low quality (the sink) that has a faster rate of decrease than the increase in population (Palumbi, 2004). While the “sink” habitat is unsustainable within itself, individuals that migrate to it from the overflowing “source” habitat can improve the sustainability Figure 4.19 Length distribution for whiting (*Merlangius merlangus*) with a mean length of 110 mm.

. With this theory in mind and the literature stating how offshore cod can affect other populations (implying some level of migration), a population genetics study could determine where the juveniles are occurring from (using the spawning site as the source) and where they could possibly go (the nursery habitat as the sink). Both habitats would need to be protected, possibly by a network of protected areas in order to ensure the juveniles' survivability during the crucial life stage of recruitment. The population genetics study would also provide key information about the home range and connectivity of the sub-populations that would narrow down where and how this network should be implemented. This project could be built upon the work of studies such as Barth et al. (2019) and Huserbraaten et al. (2018) with a focus on not only the recruits, but the actual eggs and larvae themselves to produce a more fine-tuned picture of their movement.

5.5 Importance of baseline monitoring studies

Baseline monitoring studies are key to any topic under investigation in an environment. They provide the foundation for future studies to build upon, allowing them to pick apart specific areas and questions in order to expand that area of knowledge. However, the following studies should be mindful of the shortcomings of baseline studies and focus on filling in the gaps. An example is this study itself with the presence-absence data. Barcelo et al. (2016) stated that this type of data has the possibility of overemphasizing the role of rare occurrences, which may have happened with species in this study such as the greater weever and spotted ray.

6 Conclusion

6.1 Conclusion

This work demonstrated that BRUVs may be used in temperate and relatively turbid conditions such as those experienced during the collection of video recording utilized herein. Moreover, BRUVs are highly useful in assessing young-of-the-year juveniles and their association with other species in the coastal fish assemblage, and with different habitat categories. This information is crucial to management agencies who are in the process of establishing and maintaining viable fisheries and marine protected areas. This work also highlights which topics to investigate further, and which scientific holes need to be filled.

6.2 Future work

Future studies should focus heavily on juveniles of all economically and ecologically important species, including Atlantic cod. Recruitment in these species has only recently been considered more frequently in terms of management strategies, and it is the author's belief that stronger protection of juveniles is key to both the fishery's success and the health of the surrounding environment. Specifically, their habitat preferences should be towards the top of the priority list since adequate wildlife management requires knowledge of animal habitat requirements and how environmental variables influence their habitat selection (Freitas et al., 2015; 2021). In terms of proposing regions for marine protected areas, the physical composition of the environment and the waters themselves should be considered, given that factors such as salinity can change significantly depending on the specific location in relation to the fjord. This is strengthened by the fact that the studies mentioned in the introduction highlighted how unique the hydrology of this area is. All these factors have an impact on the species and the overall structure and composition of the community. This was reflected in this study with the ANOVA comparing the MaxN means.

Management agencies should also investigate the possibility of using the spillover strategy in order to maintain a sustainable fishing stock and encourage a thriving ecosystem. This strategy has proven extremely effective in many cases and could prove so in this region given how sedentary cod populations tend to be. Huserbraaten et al. (2018) found that the spillover from the North Sea cod into Norwegian nursery habitats could be beneficial to the stock. When setting up the protected area for the spillover strategy, it is important to note that while marine reserves are not usually considered for species such as Atlantic cod, even a small area could boost conservation efforts as it would protect sedentary adults (Rogers et al., 2014). The tricky part of setting up this reserve would be the placement along the coast but would keep the populations healthy and viable for years to come.

References / bibliography

Barceló, C., Ciannelli, L., Olsen, E. M., Johannessen, T., & Knutsen, H. (2015). Eight decades of sampling reveal a contemporary novel fish assemblage in coastal nursery habitats. *Global Change Biology*, *22*(3), 1155–1167. doi: 10.1111/gcb.13047

Barth, J. M., Villegas-Ríos, D., Freitas, C., Moland, E., Star, B., André, C., Knutsen, H., Bradbury, I., Dierking, J., Petereit, C., Righton, D., Metcalfe, J., Jakobsen, K. S., Olsen, E. M., and Jentoft, S. (2019). Disentangling structural genomic and behavioural barriers in a sea of connectivity. Molecular Ecology 28, 1394–1411

Berkeley, S. A., Hixon, M. A., Larson, R. J., & Love, M. S. (2004). Fisheries Sustainability via Protection of Age Structure and Spatial Distribution of Fish Populations. *Fisheries*, *29*(8), 23–32. https://doi.org/10.1577/1548-8446(2004)29[23:fsvpoa]2.0.co;2

Blanca, M. J., Alarcón, R., Arnau, J., Bono, R., & Bendayan, R. (2017). Non-normal data: Is ANOVA still a valid option?. *Psicothema*, *29*(4), 552-557.

Campbell, M. D., Pollack, A. G., Gledhill, C. T., Switzer, T. S., & DeVries, D. A. (2015). Comparison of relative abundance indices calculated from two methods of generating video count data. *Fisheries Research*, *170*, 125–133. https://doi.org/10.1016/j.fishres.2015.05.011

Ciannelli, L., Knutsen, H., Olsen, E. M., Espeland, S. H., Asplin, L., Jelmert, A., . . . Stenseth, N. C. (2010). Small-scale genetic structure in a marine population in relation to water circulation and egg characteristics. *Ecology, 91*(10), 2918-2930. doi:10.1890/09-1548.1

Elliott, S. A. M., Sabatino, A. D., Heath, M. R., Turrell, W. R., & Bailey, D. M. (2017). Landscape effects on demersal fish revealed by field observations and predictive seabed modelling. *Plos One*, *12*(12). doi: 10.1371/journal.pone.0189011

Ctenolabrus rupestris summary page. FishBase. (n.d.). https://www.fishbase.de/summary/Ctenolabrus-rupestris.html.

Freitas, C., Olsen, E. M., Knutsen, H., Albretsen, J., & Moland, E. (2015). Temperature-associated habitat selection in a cold-water marine fish. *Journal of Animal Ecology, 85*(3), 628-637. doi:10.1111/1365-2656.12458

Freitas, C., Villegas-Ríos, D., Moland, E., and Olsen, E.M. (2021). Sea temperature effects on depth use and habitat selection in a marine fish community. Journal of Animal Ecology 00, 1–14. https://doi.org/10.1111/1365-2656.13497

Fromentin, J. (2000). Biological processes and environmental factors regulating the dynamics of the Norwegian Skagerrak cod populations since 1919. *ICES Journal of Marine Science*, *57*(2), 330–338. doi: 10.1006/jmsc.1999.0638

Garson, G. D. (2012). *Testing Statistical Assumptions "Blue Book Series"* (2012 Edition). North Carolina State University, School of Public and International Affairs.

The Basics of Photogrammetry. Geodetic Systems, Inc. (n.d.). https://www.geodetic.com/v-stars/what-is-photogrammetry/.

Glen, S. (2018, June 5). *Parsimonious Model: Definition, Ways to Compare Models*. Statistics How To. https://www.statisticshowto.com/parsimonious-model/.

Goetze, J. S., Wilson, S., Radford, B., Fisher, R., Langlois, T. J., Monk, J., Knott, N. A., Malcolm, H., Currey-Randall, L. M., Ierodiaconou, D., Harasti, D., Barrett, N., Babcock, R. C., Bosch, N. E., Brock, D., Claudet, J., Clough, J., Fairclough, D. V., Heupel, M. R., . . . Harvey, E. S. (2021). Increased connectivity and depth improve the effectiveness of marine reserves. *Global Change Biology*. Published. https://doi.org/10.1111/gcb.15635

Harvey, E. S., McLean, D., Frusher, S., E., H. M. D., Newman, S. J., & Williams, A. (2013). *The use of BRUVs as a tool for assessing marine fisheries and ecosystems: a review of the hurdles and potential: 2011 national workshop*. The University of Western Australia.

Huserbråten, M., Moland, E., & Albretsen, J. (2018). Cod at drift in the North Sea. *Progress in Oceanography*, *167*, 116–124. doi: 10.1016/j.pocean.2018.07.005

Johannessen, T., Dahl, E., Falkenhaug, T., & Naustvoll, L. J. (2011). Concurrent recruitment failure in gadoids and changes in the plankton community along the Norwegian Skagerrak coast after 2002. *ICES Journal of Marine Science*, *69*(5), 795–801. doi: 10.1093/icesjms/fsr194

Jorde, P., Knutsen, H., Espeland, S., & Stenseth, N. (2007). Spatial scale of genetic structuring in coastal cod Gadus morhua and geographic extent of local populations. *Marine Ecology Progress Series, 343*, 229-237.

Knutsen, H., André, C., Jorde, P. E., Skogen, M. D., Thuróczy, E., & Stenseth, N. C. (2004). Transport of North Sea cod larvae into the Skagerrak coastal populations. *Proceedings of the Royal Society of London. Series B: Biological Sciences, 271*(1546), 1337-1344. doi:10.1098/rspb.2004.2721

Knutsen, H., Jorde, P. E., Hutchings, J. A., Hemmer-Hansen, J., Grønkjær, P., Jørgensen, K. E. M., André, A., Sodeland, M., Albretsen, J., and Olsen, E. M. (2018). Stable coexistence of genetically divergent Atlantic cod ecotypes at multiple spatial scales. Evolutionary Applications 11, 1527–153.

Langlois, T., Harvey, E., Fitzpatrick, B., Meeuwig, J., Shedrawi, G., & Watson, D. (2010). Cost-efficient sampling of fish assemblages: comparison of baited video stations and diver video transects. *Aquatic Biology*, *9*(2), 155–168. doi: 10.3354/ab00235

Moksnes, P.-O., Gullström, M., Tryman, K., & Baden, S. (2008). Trophic cascades in a temperate seagrass community. *Oikos*, *117*(5), 763–777. doi: 10.1111/j.0030-1299.2008.16521.x

Munk, P., Cardinale, M., Casini, M., & Rudolphi, A. (2014). The community structure of over-wintering larval and small juvenile fish in a large estuary. *Estuarine, Coastal and Shelf Science, 139*, 27-39. doi:10.1016/j.ecss.2013.12.024

Palumbi, S. R. (2004). MARINE RESERVES AND OCEAN NEIGHBORHOODS: The Spatial Scale of Marine Populations and Their Management. *Annual Review of Environment and Resources*, *29*(1), 31–68. https://doi.org/10.1146/annurev.energy.29.062403.102254

Perry, D., Staveley, T. A. B., & Gullström, M. (2018). Habitat Connectivity of Fish in Temperate Shallow-Water Seascapes. *Frontiers in Marine Science*, *4*. doi: 10.3389/fmars.2017.00440

R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Rogers, L., Olsen, E., Knutsen, H., & Stenseth, N. (2014). Habitat effects on population connectivity in a coastal seascape. *Marine Ecology Progress Series*, *511*, 153–163. doi: 10.3354/meps10944

Staveley, T. A., Perry, D., Lindborg, R., & Gullström, M. (2016). Seascape structure and complexity influence temperate seagrass fish assemblage composition. *Ecography, 40*(8), 936-946. doi:10.1111/ecog.02745

Stobart, B., Díaz, D., Álvarez, F., Alonso, C., Mallol, S., & Goñi, R. (2015). Performance of Baited Underwater Video: Does It Underestimate Abundance at High Population Densities? *Plos One*, *10*(5). doi: 10.1371/journal.pone.0127559

Taylor, M. D., Baker, J., & Suthers, I. M. (2013). Tidal currents, sampling effort and baited remote underwater video (BRUV) surveys: Are we drawing the right conclusions? *Fisheries Research*, *140*, 96–104. https://doi.org/10.1016/j.fishres.2012.12.013

University of Arizona. (1993, March 31). Chapter 10 Principles of Photogrammetry. https://www.lpl.arizona.edu/hamilton/sites/lpl.arizona.edu.hamilton/files/courses/ptys551/Principles_of_Photogrammetry.pdf.

Wennhage, H. & Pihl, L. (2002). Fish feeding guilds in shallow rocky and soft bottom areas on the Swedish west coast. *Journal of Fish Biology, 0*. doi:10.1016/s0022-1112(02)92078-7

Östman, Ö, Eklöf, J., Eriksson, B. K., Olsson, J., Moksnes, P., & Bergström, U. (2016). Top-down control is as important as nutrient enrichment for eutrophication effects in North Atlantic coastal ecosystems. *Journal of Applied Ecology, 53*(4), 1138-1147. doi:10.1111/1365-2664.12654

Overview of tables and figures

a mean of 198 mm.

- Figure 4.8 Length distribution for common dab (*Limnada limnada*) with a mean length of 180 mm.
- Figure 4.9: Length distribution for European flounder (*Platichthys flesus*) with a mean of 231 mm.
- Figure 4.10: Length distribution for corkwing wrasse (*Symphodus melops*)
- Figure 4.11 Length distribution for European green crab (*Carcinus maenas*) with a mean length of 119 mm.
- Figure 4.12 Length distribution for Pomatoschistus sp with a mean length of 101 mm.
- Figure 4.13 Length distribution for goldsinny wrasse (*Ctenolabrus rupestris*) with a mean length of 105 mm.
- Figure 4.14 Length distribution for greater weever (*Trachinus draco*) with a mean length of 305 mm.
- Figure 4.15 Length distribution for grey gurnard (*Eutrigla gurnardus*) with a mean length of 226 mm.
- Figure 4.16 Length distribution for haddock (*Melanogrammus aeglefinus*) with a mean length of 90 mm.
- Figure 4.17 Length distribution for pollock, also known as saithe (*Pollachius virens*)
- with a mean length of 127 mm.
- Figure 4.18 Length distribution for sand goby (*Pomatoschistus minutus*)
- with a mean length of 106 mm.
- Figure 4.19 Length distribution for scale-rayed wrasse (*Acantholabrus palloni*) with a mean length of 111 mm.
- Figure 4.20 Length distribution for ballan wrasse (*Labris bergylta*) with a mean length

Appendices

Appendix 1: R code

Length measurement work: density plots

```
species_length=read.csv(file.choose(), header=T)
attach(dat)
head(dat)
##Length distribution for all species
plot(Length.mm, main="Length raw data")
abline(h=mean(Length.mm,), col="red")
abline(h=mean(Length.mm,)-sd(Length.mm,), col="blue")
abline(h=mean(Length.mm,)+sd(Length.mm,), col="blue")
plot(density(Length.mm), main="Length density raw")
abline(v=mean(Length.mm,), col="red")
abline(v=mean(Length.mm,)-sd(Length.mm,), col="blue")
abline(v=mean(Length.mm,)+sd(Length.mm,), col="blue")
library(ggplot2)
d<-density(Length.mm)
qplot(d$x,d$y)
p<-ggplot()
p<-p+geom_density(data=dat, aes(x=Length.mm,))
p


### Assessing normality
## qqplot
qqnorm(Length.mm, main="Length measurements raw")
```

```
qqline(Length.mm)
# make a qqplot for the norm vector
qqnorm(norm, main="Random data from the Standard normal")
qqline(norm)
##############  data transformation
par(mfrow=c(2,2))
logLength=log(Length.mm)
plot(density(Length.mm))
plot(density(logLength))
qqnorm(Length.mm)
qqline(Length.mm)
qqnorm(logLength)
qqline(logLength)

#Histograms for length distributions by species
par(mfrow=c(1,1))
ballan=species_length[species_length$Species=="bergylta",]
hist(ballan$Length_mm)
abline(v=mean(ballan$Length_mm), col="red", lwd=3)
abline(v=median(ballan$Length_mm), col="blue", lwd=3)
mean(ballan$Length_mm)
text(160,2.0,label="Mean = 148.27", col="red")
text(110,2.0,label="Median", col="blue")
plot(density(ballan$Length_mm), main="Length distribution for ballan wrasse", xlab="Length in
mm", ylab="Density")
abline(v=mean(ballan$Length_mm), col="green")
```

```
abline(v=median(ballan$Length_mm), col="red")
text(160,0.004,label="Mean = 148.27", col="dark green")
text(110,0.004,label="Median", col="red")

goldsinny=species_length[species_length$Species=="rupestris",]
hist(goldsinny$Length_mm)
abline(v=mean(goldsinny$Length_mm), col="red", lwd=3)
abline(v=median(goldsinny$Length_mm), col="blue", lwd=3)
text(150,200,label="Mean", col="red")
text(60,200,label="Median", col="blue")
mean(goldsinny$Length_mm)
plot(density(goldsinny$Length_mm), main="Length distribution for goldsinny wrasse", xlab="Length
in mm", ylab="Density")
abline(v=mean(goldsinny$Length_mm), col="green")
abline(v=median(goldsinny$Length_mm), col="red")
text(160,0.010,label="Mean = 105.63", col="dark green")
text(50,0.0104,label="Median", col="red")

haddock=species_length[species_length$Species=="aeglefinus",]
mean(haddock$Length_mm)
plot(density(haddock$Length_mm), main="Length distribution for haddock", xlab="Length in mm",
ylab="Density")
abline(v=mean(haddock$Length_mm), col="green")
text(95,0.010,label="Mean (90.02) median", col="dark green")

weever=species_length[species_length$Species=="draco",]
```

```
mean(weever$Length_mm)
plot(density(weever$Length_mm), main="Length distribution for greater weever", xlab="Length in mm", ylab="Density")
abline(v=mean(weever$Length_mm), col="green")
text(310,0.01,label="Mean (305.97) and median", col="dark green")

flounder=species_length[species_length$Species=="flesus",]
hist(flounder$Length_mm)
abline(v=mean(flounder$Length_mm), col="red", lwd=3)
abline(v=median(flounder$Length_mm), col="blue", lwd=3)
text(240,1.5,label="Mean", col="red")
text(210,1.5,label="Median", col="blue")
mean(flounder$Length_mm)
plot(density(flounder$Length_mm), main="Length distribution for European flounder", xlab="Length in mm", ylab="Density")
abline(v=mean(flounder$Length_mm), col="green")
abline(v=median(flounder$Length_mm), col="red")
text(240,0.004,label="Mean = 231.57", col="dark green")
text(200,0.0004,label="Median", col="red")

gurnard=species_length[species_length$Species=="gurnardus",]
mean(gurnard$Length_mm)
plot(density(gurnard$Length_mm), main="Length distribution for grey gurnard", xlab="Length in mm", ylab="Density")
abline(v=mean(gurnard$Length_mm), col="green")
text(240,0.004,label="Mean (226.81) and median", col="dark green")
```

```
halibut=species_length[species_length$Species=="hippoglossus",]
hist(halibut$Length_mm)
abline(v=mean(halibut$Length_mm), col="red", lwd=3)
abline(v=median(halibut$Length_mm), col="blue", lwd=3)
text(210,20,label="Mean", col="red")
text(180,20,label="Median", col="blue")
mean(halibut$Length_mm)
plot(density(halibut$Length_mm), main="Length distribution for Atlantic halibut", xlab="Length in mm", ylab="Density")
abline(v=mean(halibut$Length_mm), col="green")
abline(v=median(halibut$Length_mm), col="red")
text(250, 0.004,label="Mean = 198.35", col="dark green")
text(150,0.004,label="Median", col="red")

dab=species_length[species_length$Species=="limanda",]
mean(dab$Length_mm)
plot(density(dab$Length_mm), main="Length distribution for common dab", xlab="Length in mm", ylab="Density")
abline(v=mean(dab$Length_mm), col="green")
text(190, 0.005,label="Mean (180.45) and median", col=" dark green")

greencrab=species_length[species_length$Species=="maenas",]
hist(greencrab$Length_mm)
abline(v=mean(greencrab$Length_mm), col="red", lwd=3)
abline(v=median(greencrab$Length_mm), col="blue", lwd=3)
```

```
text(140,2,label="Mean", col="red")
text(60,2,label="Median", col="blue")
mean(greencrab$Length_mm)
plot(density(greencrab$Length_mm), main="Length distribution for European green crab",
xlab="Length in mm", ylab="Density")
abline(v=mean(greencrab$Length_mm), col="green")
abline(v=median(greencrab$Length_mm), col="red")
text(150, 0.002,label="Mean = 119.95", col="dark green")
text(80, 0.002,label="Median", col="red")

corkwing=species_length[species_length$Species=="melops",]
mean(corkwing$Length_mm)
plot(density(corkwing$Length_mm), main="Length distribution for corkwing wrasse", xlab="Length
in mm", ylab="Density")
abline(v=mean(corkwing$Length_mm), col="green")
abline(v=median(corkwing$Length_mm), col="red")
text(200, 0.005,label="Mean (192.54) and median", col="dark green")

whiting=species_length[species_length$Species=="merlangus",]
hist(whiting$Length_mm)
abline(v=mean(whiting$Length_mm), col="red", lwd=3)
abline(v=median(whiting$Length_mm), col="blue", lwd=3)
text(90,6,label="Mean", col="red")
text(130,6, label="Median", col="blue")
mean(whiting$Length_mm)
```

```
plot(density(whiting$Length_mm), main="Length distribution for whiting", xlab="Length in mm",
ylab="Density")
abline(v=mean(whiting$Length_mm), col="green")
abline(v=median(whiting$Length_mm), col="red")
text(90,0.015,label="Mean = 110.00", col="dark green")
text(130,0.015,label="Median", col="red")

sandgoby=species_length[species_length$Species=="minutus",]
mean(sandgoby$Length_mm)
plot(density(sandgoby$Length_mm), main="Length distribution for sand goby", xlab="Length in
mm", ylab="Density")
abline(v=mean(sandgoby$Length_mm), col="green")
abline(v=median(sandgoby$Length_mm), col="red")
text(90,0.005,label="Mean = 106.47", col="dark green")
text(125,0.005,label="Median", col="red")

cuckoo=species_length[species_length$Species=="mixtus",]
hist(cuckoo$Length_mm)
abline(v=mean(cuckoo$Length_mm), col="red", lwd=3)
abline(v=median(cuckoo$Length_mm), col="blue", lwd=3)
text(250,4,label="Mean", col="red")
text(150,4, label="Median", col="blue")
mean(cuckoo$Length_mm)
plot(density(cuckoo$Length_mm), main="Length distribution for cuckoo wrasse", xlab="Length in
mm", ylab="Density")
abline(v=mean(cuckoo$Length_mm), col="green")
```

```
abline(v=median(cuckoo$Length_mm), col="red")
text(250,0.010,label="Mean = 212.80", col="dark green")
text(150,0.010,label="Median", col="red")

cod=species_length[species_length$Species=="morhua",]
hist(cod$Length_mm)
abline(v=mean(cod$Length_mm), col="red", lwd=3)
abline(v=median(cod$Length_mm), col="blue", lwd=3)
text(130,150,label="Mean", col="red")
text(50,150, label="Median", col="blue")
mean(cod$Length_mm)
plot(density(cod$Length_mm), main="Length distribution for Atlantic cod", xlab="Length in mm",
ylab="Density")
abline(v=mean(cod$Length_mm), col="green")
abline(v=median(cod$Length_mm), col="red")
text(150,0.010,label="Mean = 104.30", col="dark green")
text(50,0.010,label="Median", col="red")

ediblecrab=species_length[species_length$Species=="pagurus",]
hist(ediblecrab$Length_mm)
abline(v=mean(ediblecrab$Length_mm), col="red", lwd=3)
abline(v=median(ediblecrab$Length_mm), col="blue", lwd=3)
text(200,40,label="Mean", col="red")
text(50,40, label="Median", col="blue")
mean(ediblecrab$Length_mm)
```

```
plot(density(ediblecrab$Length_mm), main="Length distribution for edible crab", xlab="Length in mm", ylab="Density")
abline(v=mean(ediblecrab$Length_mm), col="green")
abline(v=median(ediblecrab$Length_mm), col="red")
text(200,0.004,label="Mean = 150.18", col="dark green")
text(80,0.004,label="Median", col="red")

scaleray=species_length[species_length$Species=="palloni",]
hist(scaleray$Length_mm)
abline(v=mean(scaleray$Length_mm), col="red", lwd=3)
abline(v=median(scaleray$Length_mm), col="blue", lwd=3)
text(115,1.5,label="Mean", col="red")
text(105,1.5, label="Median", col="blue")
mean(scaleray$Length_mm)
plot(density(scaleray$Length_mm), main="Length distribution for scale-rayed wrasse", xlab="Length in mm", ylab="Density")
abline(v=mean(scaleray$Length_mm), col="green")
abline(v=median(scaleray$Length_mm), col="red")
text(120,0.020,label="Mean = 111.42", col="dark green")
text(100,0.020,label="Median", col="red")

scaleray=species_length[species_length$Species=="palloni",]
hist(scaleray$Length_mm)
abline(v=mean(scaleray$Length_mm), col="red", lwd=3)
abline(v=median(scaleray$Length_mm), col="blue", lwd=3)
text(115,1.5,label="Mean", col="red")
```

```
text(105,1.5, label="Median", col="blue")
plot(density(scaleray$Length_mm), main="Length distribution for scale-rayed wrasse", xlab="Length in mm", ylab="Density")
abline(v=mean(scaleray$Length_mm), col="green")
abline(v=median(scaleray$Length_mm), col="red")
text(120,0.020,label="Mean", col="green")
text(100,0.020,label="Median", col="red")

gobysp=species_length[species_length$Genus=="Pomatoschistus",]
hist(gobysp$Length_mm)
abline(v=mean(gobysp$Length_mm), col="red", lwd=3)
abline(v=median(gobysp$Length_mm), col="blue", lwd=3)
text(125,30,label="Mean", col="red")
text(50,30, label="Median", col="blue")
mean(gobysp$Length_mm)
plot(density(gobysp$Length_mm), main="Length distribution for Pomatoschistus sp", xlab="Length in mm", ylab="Density")
abline(v=mean(gobysp$Length_mm), col="green")
abline(v=median(gobysp$Length_mm), col="red")
text(150,0.008,label="Mean = 101.06", col="dark green")
text(50,0.008,label="Median", col="red")

pollock=species_length[species_length$Species=="virens",]
hist(pollock$Length_mm)
abline(v=mean(pollock$Length_mm), col="red", lwd=3)
abline(v=median(pollock$Length_mm), col="blue", lwd=3)
```

```
text(150,40,label="Mean", col="red")
text(50,40, label="Median", col="blue")
mean(pollock$Length_mm)
plot(density(pollock$Length_mm), main="Length distribution for pollock", xlab="Length in mm",
ylab="Density")
abline(v=mean(gobysp$Length_mm), col="green")
abline(v=median(gobysp$Length_mm), col="red")
text(125,0.005,label="Mean = 127.65", col="dark green")
text(50,0.005,label="Median", col="red")
```

Data wrangling/ “tidying up”

```
dat=read.csv(file.choose(), header=T)
install.packages("vegan")
library("vegan") ##To run diversity function
install.packages("dplyr")
library("dplyr") ##To tidy up data
head(dat)
str(dat)      #Shows structure of data
#Add count column for cod
dat$codcount <-  #for each frame sum the number column if the species is cod
Countdat <- dat %>%
  group_by(Frame, Code) %>%
  summarize(NumSp = sum(Number))
head(Countdat)
install.packages("reshape2")
library("reshape2")
wrangle1 <- acast(Countdat, Frame + NumSp ~ Code)
wrangle1[is.na(wrangle1)] = 0
```

```
head(wrangle1)
```

Diversity calculations and merging datasets

```
diversity.vector <- diversity(wrangle1, "shannon")
#setting up a little dataset to put the frame numbers back in
Index <- diversity.vector
IndexNum <- as.numeric(Index)
Frame <- rownames(wrangle1)
diversity.dataset <- read.csv(file.choose()) ##EXPORTED DIV. VECTOR IN ORDER TO FIX
FRAMES IN EXCEL TO MATCH COUNTDAT
head(diversity.dataset)
#Smush data sets together
newdata <- merge(Countdat, diversity.dataset, all.x=TRUE)
newdata[is.na(newdata)] = 0
head(newdata)
#Add pres/abs for cod
newdata$codpa <- ifelse(newdata$Code == "Torsk", 1, 0)
#Add count column for cod
newdata$codcount <-  ifelse(newdata$Code == "Torsk", newdata$NumSp, 0)
```

Testing

```
#CHI SQUARED
#H0: The presence of cod has no impact on the presence of other species
#H1: Cod presence does influence the abundance of other species
test <- chisq.test(table(newdata$codpa, newdata$IndexNum))
test
#Does cod presence lead to different diversities?
# diversity index predicted by cod presence cod counts
```

```
divmod <-  lm(newdata$IndexNum ~ newdata$codpa + newdata$codcount, data = newdata)
summary(divmod)
#Logistic regression to determine factors in cod presence
install.packages(lme4)
library(lme4)
#Putting the big dataset with the original to get habitat and depth information in the same dataset
completedat <- merge(newdata, dat, all.x=TRUE)
attach(completedat)
#Removing irrelevant columns
completedat = subset(completedat, select = -c(Number, Date, Region, Precision_mm))
#Logsitic regression
codpres1 <- glm(codpa ~ habitat + Depth + IndexNum, data = completedat,  family = binomial)
codpres2 <- glm(codpa ~ habitat + Depth, data = completedat,  family = binomial)
codpres3 <- glm(codpa ~ habitat + IndexNum, data = completedat,  family = binomial)
codpres4 <- glm(codpa ~ Depth + IndexNum, data = completedat,  family = binomial)
summary(codpres1)
summary(codpres2)
summary(codpres3)
summary(codpres4)
library(rms)
install.packages("devtools")
library("devtools")
install.packages("ggiraphExtra")
library("ggiraphExtra")
install.packages("ggeffects")
library("ggeffects")
```

```
divplot <- ggpredict(codpres,se=TRUE,interactive=TRUE,digits=3)
plot(divplot)
##############################################################################
```

Independent 2 sample t test, comparing north and south stations, length measurements

```
#innersites = Northern stations closer to mouth of fjord
innersites <- read.csv(file.choose())
#outersites = Southern stations closer to open ocean
outersites <- read.csv(file.choose())
innerlength <- innersites$Length_mm
outerlength <- outersites$Length_mm
res <- t.test(innerlength, outerlength, var.equal = TRUE)
res
##########################################################################
```

Independent 2 sample t test north vs south, diversity

```
install.packages("dplyr")
library("dplyr")
install.packages("vegan")
library("vegan")
#Outer sites first: Add count column for cod
outersites$codcount <-  #for each frame sum the number column if the species is cod
#Add count column for all species
outercount <- outersites %>%
  group_by(Frame, Code) %>%
  summarize(NumSp = sum(Number))
head(outercount)
install.packages("reshape2")
```

```
library("reshape2")
outerwrangle <- acast(outercount, Frame + NumSp ~ Code)
outerwrangle[is.na(outerwrangle)] = 0
head(outerwrangle)
outerdiv <- diversity(outerwrangle, "shannon")
#Inner sites
innersites$codcount <-  #for each frame sum the number column if the species is cod
 #Add count column for all species
innercount <- innersites %>%
  group_by(Frame, Code) %>%
  summarize(NumSp = sum(Number))
head(innercount)
innerwrangle <- acast(innercount, Frame + NumSp ~ Code)
innerwrangle[is.na(innerwrangle)] = 0
head(innerwrangle)
innerdiv <- diversity(innerwrangle, "shannon")
#Comparing north and south diversities
divttest <- t.test(innerdiv, outerdiv, var.equal = TRUE)
divttest
##########################################################################
```

Independent 2 sample test north vs south, abundance

```
abundttest <- t.test(innercount$NumSp, outercount$NumSp, var.equal = TRUE)
abundttest
```

Max N work

```
maxn=read.csv(file.choose(), header=T)
```

```
attach(maxn)
head(maxn)
par(mfrow=c(1,1))
##Histogram
plot(MaxN, main="MaxN")
abline(h=mean(MaxN), col="red")
text(10,9,label="Mean", col="red")
###Density plot
plot(density(MaxN), main="MaxN density")
abline(v=mean(MaxN), col="red")
text(10,0.2,label="Mean", col="red")
library(ggplot2)
d<-density(MaxN)
qplot(d$x,d$y)
p<-ggplot()
p<-p+geom_density(data=dat, aes(x=MaxN))
p

### Assessing normality
## qqplot
qqnorm(MaxN, main="Max N")
qqline(MaxN)
##############  data transformation: Log
par(mfrow=c(2,2))
logMaxN=log(MaxN)
plot(density(MaxN))
```

```
plot(density(logMaxN))

qqnorm(MaxN)

qqline(MaxN)

qqnorm(logMaxN)

qqline(logMaxN)

###################### The square root transform

par(mfrow=c(2,2))

sqrtMaxN=sqrt(MaxN)

plot(density(MaxN))

plot(density(sqrtMaxN))

qqnorm(MaxN)

qqline(MaxN)

qqnorm(sqrtMaxN)

qqline(sqrtMaxN)

par(mfrow=c(2,1))

hist(MaxN)

hist(logMaxN)

#Does cod presence affect Max N?

#Add pres/abs for cod

maxn$codpa <- ifelse(maxn$Code == "Torsk", 1, 0)

#CHI SQUARED

#H0: The presence of cod has no impact on the MaxN

#H1: Cod presence does influence MaxN

test <- chisq.test(table(maxn$codpa, MaxN))

test

#Linear model predicting what influences MaxN between cod presence, habitat and depth
```

```
library(lme4)
maxnmod <-  lm(MaxN ~ codpa + habitat + Depth, data = maxn)
summary(maxnmod)
#One Way Anova to compare means of north, south and middle stations
install.packages(c("ggplot2", "ggpubr", "tidyverse", "broom"))
library(ggplot2)
library(ggpubr)
library(tidyverse)
library(broom)
summary(maxn)
##Data was divided into three “regions”: North, South and Middle
one.way <- aov(maxn$MaxN ~ maxn$Region, data = maxn)
summary(one.way)
par(mfrow=c(2,2))
plot(one.way)
par(mfrow=c(1,1))
#Tukey test to examine specific differences
tukey.maxN<-TukeyHSD(one.way)
tukey.maxN
#Plotting
mean.maxn.data <- maxn %>%
  group_by(Region, MaxN) %>%
  summarise(
    MaxN = mean(MaxN)
  )
one.way.plot <- ggplot(maxn, aes(x = Region, y = MaxN)) +
```

```
  geom_point(cex = 1.5, pch = 1.0,position = position_jitter(w = 0.1, h = 0))
one.way.plot
one.way.plot <- one.way.plot +
  stat_summary(fun.data = 'mean_se', geom = 'errorbar', width = 0.2) +
  stat_summary(fun.data = 'mean_se', geom = 'pointrange') +
  geom_point(data=mean.maxn.data, aes(x=Region, y=MaxN))
one.way.plot
one.way.plot <- one.way.plot +
  theme_classic2() +
  labs(title = "Comparison of Max N from North, Middle and South stations")
One.way.plot
```

Linear model to examine how the diversity index is affected if cod was removed from the index

```
nocod <- read.csv(file.choose())
head(nocod)
#Add count column for all species
NoCodCount <- nocod %>%
  group_by(Frame, Code) %>%
  summarise(NumSp = sum(Number))
head(NoCodCount)
install.packages("reshape2")
library("reshape2")
NoCodWrangle <- acast(NoCodCount, Frame + NumSp ~ Code)
NoCodWrangle[is.na(NoCodWrangle)] = 0
head(NoCodWrangle)
diversity.nocod <- diversity(NoCodWrangle, "shannon")
```

```
#setting up a little dataset to put the frame numbers back in

NoCodIndex <- diversity.nocod

NoCodIndex <- as.numeric(NoCodIndex)

Frame <- rownames(NoCodWrangle)

str(NoCodWrangle)

write.csv(diversity.nocod, "C:\\Users\\Ian McLaughlin\\Documents\\ALEX THESIS\\Data analysis\\R

work\\NoCodDiv.csv")

diversity.codless<- read.csv(file.choose())

head(diversity.codless)

#Smush data sets together

NoCodData <- merge(NoCodCount, diversity.codless, all.x=TRUE)

NoCodData[is.na(NoCodData)] = 0

head(NoCodData)

completedat <- merge(completedat, NoCodData, all.x=TRUE)

completedat[is.na(completedat)] = 0

head(completedat)

#Linear model determining diversity with no cod

nocoddivmod <-  lm(NoCodIndex ~ codcount + codpa, data = completedat)

summary(nocoddivmod)

library("ggplot2")

ggplot(nocoddivmod)

ggplotRegression(lm(NoCodIndex ~ codcount + codpa, data = completedat))+
```