



Effects of towing- and environmental factors on expected trawl catches of cod and redfish

Davíð Freyr Hlynsson



Faculty of Industrial Engineering,
Mechanical Engineering and Computer Science
University of Iceland
2016

EFFECTS OF TOWING- AND ENVIRONMENTAL FACTORS ON EXPECTED TRAWL CATCHES OF COD AND REDFISH

Davíð Freyr Hlynsson

30 ECTS thesis submitted in partial fulfillment of a
Magister Scientiarum degree in Industrial Engineering

Advisors

Tómas Philip Rúnarsson

Birgir Hrafnkelsson

Haraldur A. Einarsson

Faculty Representative

Ólafur Pétur Pálsson

Faculty of Industrial Engineering,
Mechanical Engineering and Computer Science
School of Engineering and Natural Sciences
University of Iceland
Reykjavik, November 2016

Effects of towing- and environmental factors on expected trawl catches of cod and redfish

Effects of towing- and environmental factors on expected trawl catches

30 ECTS thesis submitted in partial fulfillment of a M.Sc. degree in Industrial Engineering

Copyright © 2016 Davíð Freyr Hlynsson

All rights reserved

Faculty of Industrial Engineering,
Mechanical Engineering and Computer Science
School of Engineering and Natural Sciences
University of Iceland
Hjarðarhagi 2
107, Reykjavík, Reykjavík
Iceland

Telephone: 525 4000

Bibliographic information:

Davíð Freyr Hlynsson, 2016, Effects of towing- and environmental factors on expected trawl catches of cod and redfish, M.Sc. thesis, Faculty of Industrial Engineering, Mechanical Engineering and Computer Science, University of Iceland.

Printing: Háskólaprent, Fálkagata 2, 107 Reykjavík
Reykjavík, Iceland, November 2016

Abstract

In this thesis a catch data from the Icelandic groundfish surveys are analysed. The data collection provides further opportunities to investigate the relation between catch performance and the numerous environmental variables, towing variables and other catch related variables. The species that will be investigated are cod and redfish. The investigation is also limited to two areas inside the Icelandic continental shelf, and these two areas are used to see if the results will stay consistent, independent of areas. Box plots are used to visualize the relationship between the catch rate for both cod and redfish, and the chosen explanatory variables. Based on these plots, polynomials were applied to describe the relationship between the variables and the catch rate which will be used to build a linear model. A linear regression model was generated in order to investigate the combined effects of each factor in terms of the catch performance. The model is fitted to different parts of the data where each part represents a specific species and area. The factors that show repeatedly significant effects on the catch performance are location of the tow, depth, surface- and bottom temperature, vessel effects, and towing time. Other factors, such as towing speed, towing length, weather- and sea conditions, and luminosity only explain a small proportion of the variation in the catch data. The model which gave the best fit was for cod in area 1, where the total variance explanation of the response was 35.8%. The analysis has shown that the catch data for both cod and redfish from the bottom trawl survey depend on other factors in addition to those that were investigated and vary in terms of different areas and species.

Útdráttur

Í þessari ritgerð verða aflagögn úr stofnmælingaleiðangrum Hafrannsóknarstofnunar rannsökuð. Gögnin veita tækifæri til að skoða áhrif tog- og umhverfisþátta á aflamagn við togveiðar. Þær tegundir sem verða rannsakaðar eru þorskur og karfi en einnig var notast við tvö afmörkuð svæði til að greina muninn á áhrifum hvers þáttar á veiðina, milli svæða. Notast er við margliður af ýmsum stigum við gerð á línulegu líkani, til að lýsa sambandi hvernar breytu og aflans fyrir bæði þorsk og karfa, en þær voru ákvarðaðar eftir að hafa skoðað sambandið með myndrænum hætti. Beitt er línulegri aðhvarfsgreiningu til að fá betri sýn á áhrif hvers þáttar og sameiginleg áhrif þeirra á aflann. Mismunandi hlutar af gögnunum eru prófaðir á líkanið, þar sem hver hluti stendur fyrir ákveðið svæði og fisktegund. Þeir þættir sem sýna ítrekað martæk áhrif á veiðina eru staðsetning togsins, dýpið, botn hitastig sjávar, yfirborðs hitastig sjávar, togtími og hvaða skip er fyrir valinu. Toghraði, tog lengd, veður- og sjóskilyrði voru einnig notaðir til að spá fyrir um aflann en útskýrðu aðeins lítið hlutfall af heildar breytileikanum í aflagögnunum. Líkanið sem útskýrði mest af heildar breytileikanum í afla gögnunum var fyrir þorsk inn á svæði 1, eða um 35.8%. Rannsóknin hefur sýnt fram á að aflagögnin fyrir bæði þorsk og karfa eru háð öðrum þáttum auk þeirra sem voru rannsakaðir og eru niðurstöðurnar breytilegar eftir svæðum og tegundum.

Contents

| | |
|---|-------------|
| List of Figures | ix |
| List of Tables | xi |
| Acknowledgments | xiii |
| 1. Introduction | 1 |
| 1.1. Motivation and objective | 1 |
| 1.2. Contribution | 2 |
| 1.3. Overview | 3 |
| 2. Background | 5 |
| 2.1. Previous studies | 5 |
| 2.2. Factors that are considered to have an effect on the catch performance | 6 |
| 2.3. General description of the species | 8 |
| 2.3.1. Cod (<i>Gadus morhua</i>) | 8 |
| 2.3.2. Redfish (<i>Sebastes marinus</i>) | 8 |
| 2.4. The Icelandic groundfish survey | 9 |
| 2.5. Summary | 10 |
| 3. Collected Data | 11 |
| 3.1. Data description | 11 |
| 3.2. Data preprocessing | 12 |
| 3.3. Summary | 14 |
| 4. Methods | 15 |
| 4.1. Linear regression | 15 |
| 4.1.1. ANOVA for regression | 16 |
| 4.1.2. Polynomial regression | 17 |
| 4.2. Summary | 18 |
| 5. Statistical summary | 19 |
| 5.1. Box plots analysis | 19 |
| 5.2. Surface- and bottom temperature | 20 |
| 5.3. Depth of the tow | 21 |
| 5.4. Weather and sea condition | 22 |

Contents

| | |
|---|-----------|
| 5.5. Time of the tow | 23 |
| 5.6. Barometric pressure | 24 |
| 5.7. Vessel effects | 25 |
| 5.8. Towing length and towing time | 26 |
| 5.9. The degree of latitude and longitude | 27 |
| 5.10. Summary | 29 |
| 6. Linear regression and analysis | 31 |
| 6.1. Areas analysed separately | 32 |
| 6.2. Area 1 and area 2 combined | 37 |
| 6.3. Summary | 39 |
| 7. Conclusion and Discussion | 41 |
| 7.1. Future work | 42 |
| Bibliography | 43 |
| A. R-script | 45 |
| B. Additional figures and tables | 57 |

List of Figures

| | |
|--|----|
| 2.1. Causal loop diagram of the factors that are considered to have an effect on the catch performance | 7 |
| 2.2. All areas in the Icelandic groundfish survey | 9 |
| 3.1. The fishing areas used for the investigation | 12 |
| 3.2. A normal probability plot and a histogram of the log transformed catch data for both redfish and cod in area 1. | 13 |
| 3.3. A normal probability plot and a histogram of the log transformed catch data for both redfish and cod in area 2. | 14 |
| 5.1. Box plots of the log number of catch per tow against surface- and bottom temperature for both redfish and cod. | 21 |
| 5.2. Box plots of the log number of catch per tow against depth of the tow for both redfish and cod. | 22 |
| 5.3. Box plots of the log number of catch per tow against weather- and sea conditions for both redfish and cod. | 23 |
| 5.4. Box plots of the log number of catch per tow against luminosity for both redfish and cod. | 24 |
| 5.5. Box plots of the log number of catch per tow against barometric pressure of the tow for both redfish and cod. | 25 |
| 5.6. Box plots of the log number of catch per tow against vessel for both redfish and cod. | 26 |
| 5.7. Box plots of the log number of catch per tow against towing length and towing time for both redfish and cod. | 27 |

LIST OF FIGURES

- 5.8. Box plots of the log number of catch per tow against the degree of latitude and longitude for both redfish and cod. The red and brown boxes represent area 1 and area 2 respectively. 28

List of Tables

| | |
|--|----|
| 4.1. Simple ANOVA table for multiple linear regression | 16 |
| 6.1. Total variation in percentages that each factor explains when it is the only factor in the model | 32 |
| 6.2. Analysis of variance for cod in area 1 | 33 |
| 6.3. Analysis of variance for cod in area 2 | 34 |
| 6.4. Analysis of variance for redfish in area 1 | 35 |
| 6.5. Analysis of variance for redfish in area 2 | 36 |
| 6.6. Analysis of variance for cod when area 1 and area 2 are combined . . | 38 |
| 6.7. Analysis of variance for redfish when area 1 and area 2 are combined | 38 |
| B.1. Analysis of variance for cod in area 1 | 57 |
| B.2. Analysis of variance for cod in area 2 | 57 |
| B.3. Analysis of variance for redfish in area 1 | 58 |
| B.4. Analysis of variance for redfish in area 2 | 58 |

Acknowledgments

I would like to thank my supervisors Tómas Philip Rúnarsson, Haraldur A. Einarsson and Birgir Hrafnkelsson, and all of those who supported me during this study. I would also like to thank the Marine Research Institute for making this project possible. I am especially thankful to my family for their support throughout this work.

1. Introduction

The fishing industry has been one of the most important industries in Iceland from the beginning of the 20th century and one of our main sources of income. This makes the industry very valuable for the Icelandic nation from an economical and cultural point of view. We are therefore constantly trying to find new ways to improve the efficiency of the industry. There have been rumours that certain environmental- and towing factors have a significant impact on the catch performance but it can be difficult to prove this because there are so many factors that are involved. It is unlikely to encounter the exact same situation more than once which makes it hard to evaluate the significant level of each environmental- and towing factor. However, various related researches have been carried out that are both informative and insightful.

1.1. Motivation and objective

The Marine Research institute (MRI) is a government institute under the auspices of the ministry of fisheries. MRI conducts various marine-related research and provides the ministry with scientific advice based on its research on marine resources and the environment. The MRI in Iceland has conducted an annual groundfish survey in the Icelandic continental shelf every March since 1985. The purpose of this survey is to gather data which is then used to evaluate the size, condition, and spread of several demersal species. The data that is gathered contains many different variables, such as bottom- and surface temperature, weather conditions, wind speed, towing speed *etc.*, for many types of species and in specific areas around Iceland. The data collection provides further opportunities to investigate the relation between catch performance and these numerous measurements of environmental variables, towing variables and others in order to discover whether it is more likely to catch fish at certain locations, under certain conditions. This kind of analysis of the Icelandic bottom fish is the main subject of this thesis. The species that will be investigated are cod and redfish, which were chosen because they provided a greater number of observations and had a smaller number of missing values inside the dataset when compared to analyses done on other species. Investigating two species also provides opportunities to study the differences between those species. The investigation is

1. Introduction

limited to two areas inside the Icelandic continental shelf and these two areas are used to see if the results will stay consistent, independent of areas [Sólmundsson et al., 2016].

Simple statistical plots called box plots will be used to investigate the relationship between the variables conducted in the survey and the catch rate. Based on these plots, polynomials were applied to describe the relationship between the variables and the catch rate. The statistical summary obtained provides useful information when making a catch forecast model using linear regression, which will be one of the main objectives of this thesis. Such forecasts can potentially be used by vessel captains in the future to help select fishing grounds or with finding the optimal conditions to catch certain demersal species.

The main motivation for this work comes from a project called Optigear. The final aim of the Optigear project is building a software which provides useful information to vessel captains when they are selecting fishing grounds or during fishing operations. Vessel captains often make decisions depending on previous experiences and the latest information from other fishermen, but this software would assist them in making more accurate decisions that result in increased average catch per tow. Decisions that might only have a slight effect on the day-to-day performance can have a great effect on the catch performance over a longer period of time. This would also improve the efficiency of fishing gear as this would likely decrease the amount of tow per tour when there is a certain amount of fishing quota. The results of this thesis will provide useful information for the Optigear project and assist them in the development of their software.

In a master thesis published in June 2002 a statistical analysis of cod catch data from the Icelandic groundfish survey was done by Jenný Brynjarsdóttir. The catch data for cod from the Icelandic groundfish surveys are analysed in that thesis, using mainly generalized linear methods where the objective was to find environmental variables that affect the expected cod catch. The objective of that master thesis is similar to this one and the conclusions from Jenný's thesis will be used for comparison. Jenný also did a similar research with Gunnar Stefánsson when they published an article in 2004 where a summary of her master thesis was made [Brynjarsdóttir and Stefánsson, 2004].

1.2. Contribution

This thesis presents a dataset obtained from the MRI that contains an enormous number of environmental measurements and tow information which will be used to investigate if these factors have a significant impact on the catch of either cod or

redfish.

Initially it needs to be considered which parts of the data are most useful for this type of investigation, *i.e.*, which species, areas, and factors are most likely to provide relevant results. The decision making of whether a factor is useful or not was evaluated by the amount and variety of the measurements for each attribute, which was done by analysing the data using the statistical software R. After the pre-processing phase, figures that demonstrate the relations between each factor and the log number of catch per tow were generated. It was decided to compare the log number of catch per tow to the weather- and sea condition, sea temperature, depth of the tow, barometer, luminosity, towing speed, towing time, towing length, vessel, and the degrees of latitude and longitude where these factors seemed to be most relevant depending on related work and the quality of the measurements. Each variable was plotted against the log number of catch per tow for two areas around the Icelandic sea to achieve greater accuracy and to be able to compare the results. Intervals had to be generated for each variable which was done by dividing the amount of observations (catch per tow) evenly which will most likely provide us with the most accurate results. When all figures have been generated, it will be possible to identify if there were any trends between certain factors and the overall catch performance for cod and redfish in these two areas. This analysis was used to make a linear model for further investigation of the significance of each factor and to find out how well they could explain the total variation of the catch for each species. The factors that show repeatedly significant effect on the catch performance were location (degree of latitude and longitude), depth of the tow, surface- and bottom temperature, vessels and towing time, but the effects are dependent on their rankings in the model. The effects of each factor also varies in terms of different areas and species. Similar results were obtain in this investigation as in Jenný's thesis but the model does not fit the cod catch data as well. This is most likely due to different approach in terms of research areas and that the dataset now contains a greater number of observations since the investigation has been ongoing so it is unlikely that the results will be exactly the same.

1.3. Overview

This thesis is organized as follows. In chapter 2 there will be a brief discussion about groundfish surveys in general. Also certain variables that are thought to have effect on the catch performance will be described using a causal loop diagram which is a causal diagram that aids in visualizing how different variables in a system are interrelated. A detailed description of the data is introduced in chapter 3 following a explanation of the pre-processing phase. In chapter 4 the methods used for this investigation are described. Then in chapters 5 all the figures generated are

1. Introduction

presented, describing the relationship between each variable and the catch rate. In chapter 6 the analysis of variance for the linear model generated is presented as well as the conclusion on which factors seem to have significant impact on the catch performance, all this is followed by a conclusion and discussion in chapter 7. Possible directions of future work are then presented in the last section of conclusion and discussion. For reference full results of analysis performed in software environment R are available in the appendix along with additional figures and tables.

2. Background

Groundfish trawl surveys are commonly conducted for the purpose of gathering statistics of the average catch per tow. This is used as an indicator of stock population which in turn is used in a stock assessment process [Brynjarsdóttir and Stefánsson, 2004]. These groundfish surveys often contain useful data for investigating the effect of different environmental variables on the catchability of different demersal species. It can be a difficult task to figure out which factors have the greatest effect on the catch performance of certain species because there are many causal relationships between the relevant variables that affect the catch, *i.e.*, many of the variables interfere with each other. The catch can depend on the population size of each species, the location of the tow, the time of the tow, environmental condition, etc.

2.1. Previous studies

Previous studies have demonstrated that certain environmental- and towing variables have a significant impact on catch performance, in which the catch data from the groundfish surveys are analysed using generalized linear models (GLM). In [Brynjarsdóttir and Stefánsson, 2004] the main goal was to test the effects of environmental variables on the expected cod catch data using GLM. The environmental factors that were considered are surface temperature, depth, bottom temperature, and wave height. The possible differences in catch rates between vessels and how much effect the location of the tow had on the catch performance were also considered. The most important environmental effects were found to be the bottom temperature, depth, and surface temperature, although most of the tested effects are found to be significant and the model used in the analysis fitted 45% of cod catch data. Similar study was performed and described in [Adlerstein and Ehrich, 2003] using GLM for the analysis. Their study contained an investigation of catch rates of North Sea cod in bottom trawl surveys within daytime hours during the summer of 1999. The effect of environmental conditions on catch rates were also investigated. The results of that research showed significant variation of catch rates during daytime, where catch rates in shallow areas were low in the early mornings but increased to a peak at around 14:00 h before declining again.

2. Background

In [Petrakis et al., 2001] is focused on the day-night and depth effects on catch rates during trawl surveys. In the article they mentioned that previous studies had shown that the daytime catches for herring, cod, haddock, beaked redfish, and American plaice are higher than the night-time catches. For many demersal fish, a positive correlation between the size of fish and depth has been reported. Furthermore, the density of fish may vary within the surveyed area since gradients can occur related to oceanographic parameters like water depth, salinity, temperature, *etc.* Their results showed that the time of day and depth are important determinants when considering variation in trawl catches. However, the effects vary by species. The difference between daytime trawl catches and catches during the night may be caused by fish behavior, particularly vertical migration and gear avoidance reactions. The variation of trawl catches between shallow and deep water indicates differences in the horizontal distribution of the species where preferred environment of each species is determined by many biotic and abiotic parameters.

It has been contemplated that the skipper plays an important role in fishing success. Some fishermen seem to catch more fish than others under the same conditions, and using the same equipments. In [Thorlindsson, 1988] an attempt is made to estimate the role of the skipper in fishing success. Data from the Icelandic summer herring fishery from the years 1959, 1960, and 1961 is used for the study. The skipper effect, measured by simple correlation between seasons explains anywhere from 35% to 49% of the variance in the catch. Of course factors like the size of the boat and time spent fishing need to be considered as they affect the fishing success. When these two relevant variables are considered the relationship remains strong which indicates that there is a significant difference between the catch performance of different skippers. However, it is mentioned in the article that these findings are limited to the data provided and that the skills required for fishing success are related to technology in use, the type of species being fished, the ecological conditions, and the social organization of the industry.

Many other studies have been carried out in which the effects of specific variables that are considered to have an influence on the catch performance are investigated. In this thesis, previous studies are considered when selecting which factors will be analysed and compared with the catch performance.

2.2. Factors that are considered to have an effect on the catch performance

Below is a Causal Loop Diagram (CLD) that was created to give a better overview of what factors are considered to have an effect on the likelihood of catching fish.

2.2. Factors that are considered to have an effect on the catch performance

All the factors that are included in the CLD are selected from a summary made from articles that cover similar subjects and they are of course limited to the data that the MRI provides.

The arrows in the diagram indicate the causal relationships between variables. In most cases, the arrows also contain positive and negative signs which will describe the interaction between two variables. The positive (+) sign at the arrowhead indicates that the effect is positively related to the cause, *e.g.*, if an increase in sea temperature causes the number of average catch per tow to increase, the arrow between those tow variables is marked with a positive sign [Stermann, 2001].

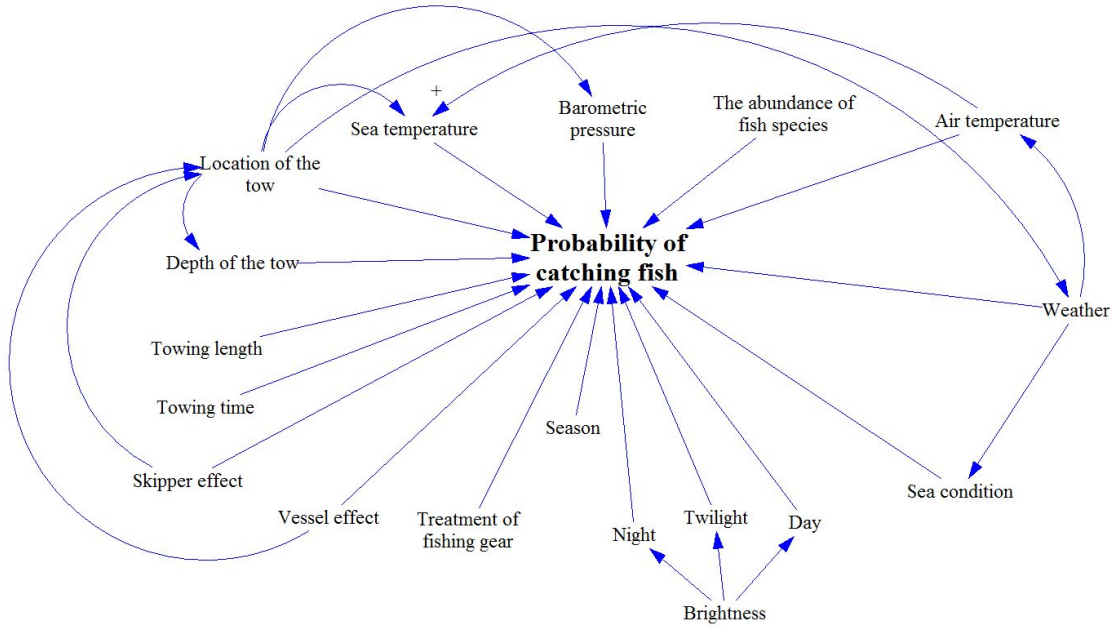


Figure 2.1: Causal loop diagram of the factors that are considered to have an effect on the catch performance

The condition of environmental variables introduced in figure 2.1, which are sea temperature, air temperature, barometric pressure, and sea condition, could all have an effect on the probability of catching fish. However, we do not know if increased sea temperature will increase or decrease the likelihood of catching fish. In these cases, the positive and negative signs are not always used, *i.e.*, the relations between the variables are unknown. It matters where the tow takes place in terms of the catch performance, but the exact reason why you catch more fish in certain spots than others is not obvious. It could be because of the different areas each vessel was allocated, that the skill and experience of the skipper resulted in a better haul, or that the environmental conditions that the tow took place in were optimal. It is also likely that the population of fish species, the condition of the fishing gear, available technology, the fishing ability of each vessel, and the seasonal effects, are crucial factors when it comes to catching fish. The CLD is not complete and is only used

2. Background

to give a better overview of what factors might influence the probability of catching fish. There is a need for more detailed investigation to add other factors and to explain the cause and effects between each factor more accurate. The majority of the variables in the figure are only considered to have an influence on the catch performance but it is not confirmed, so this thesis will hopefully shed some light on the effects of these factors.

2.3. General description of the species

In this section there will be a brief description of the species that will be analysed. It will only be covered the characteristics of the species that are considered to affect the results of this thesis.

2.3.1. Cod (*Gadus morhua*)

Cod is common in the sea all around Iceland. The cod is a demersal species that can live anywhere from a few meters depth down to 600 meters or more, but is most commonly found from 100 to 400 meters depth. It can swim to the surface when searching for food or during spawning season. Spawning occurs between March and May on the south coast and thereafter he seeks to reside within a sea temperature range of 5 to 7°C and at a depth between 50 to 150 meters. The growth of the cod depends on the temperature of the sea. For example, the cod reaches puberty between 4 to 6 years old in the warmer sea on the south coast but does not reach puberty until 6 to 9 years when living in the colder sea on the north coast [Matís, n.d.].

2.3.2. Redfish (*Sebastes marinus*)

Redfish is common in the sea all around Iceland but is more common on the west coast, specifically in the deep sea. Redfish is a demersal fish and is mostly found at a depth between 100 and 300 meters and within a sea temperature range of 3 to 8°C, but can be found all the way down to a depth of 1,000 meters. It is often found at the bottom of the ocean during daytime but swimming to the surface during the night. The redfish is a slow-growing fish and does not reach puberty until 14 to 16 years of age [Hafrannsóknarstofnun, n.d.].

2.4. The Icelandic groundfish survey

The groundfish survey area contains the Icelandic continental shelf inside the 500 meters depth contour, which covers the fishing grounds for the most important commercial species of demersal fish in Icelandic waters. In the beginning, the groundfish survey was primarily designed for cod and as a result only covered the grounds down to 500 m depth. The research area is divided into 10 different regions as can be seen on figure 2.2. There is also no constraint that states that the areas need to be continuous and therefore each one may be split up into several smaller areas [Stefánsson and Pálsson, 1997]. At first there were 590 towing stations distributed on these 10 areas based on per-estimated cod density patterns derived from commercial, as well as research vessels catch data. Stations were divided equally between fisherman and ichthyologists from the MRI and ichthyologists selected random positions for their stations. Fishermen were also provided stations and decided their position based on their knowledge and experience of fishing and the fishing grounds. Trawling is done during both day and night where the sampling is distributed uniformly over 24 hours, *i.e.*, a tow is equally probable to be performed at any given time.

Today, the groundfish trawl survey play a big role when it comes to stock assessment and catching advice for cod, haddock, redfish, catfish etc. Also the survey provides information on the spread of many other species and useful information on certain environmental and towing variables [Sólmundsson et al., 2016].

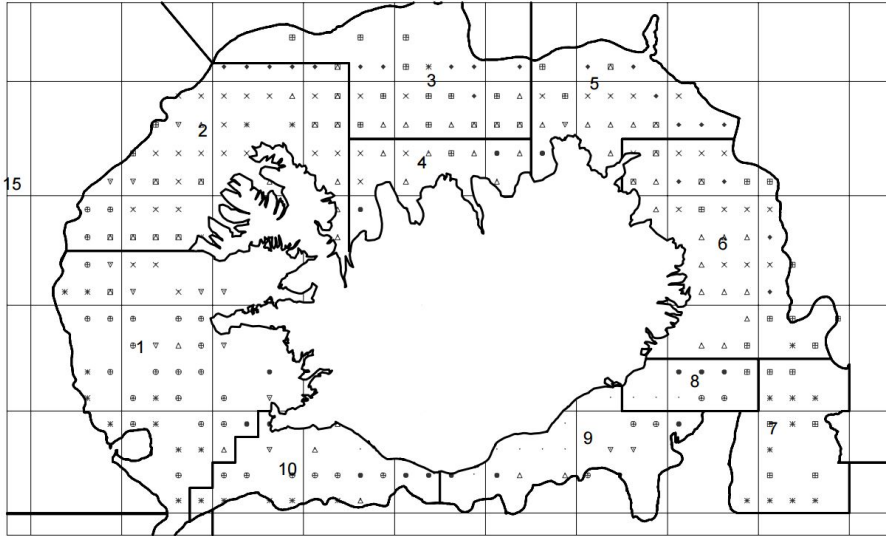


Figure 2.2: All areas in the Icelandic groundfish survey

2.5. Summary

In this chapter there is a brief discussion about previous studies that investigate the relation between certain variables and the catch performance of different species. A CLD is introduced to give a better overview of these relations. This chapter also covers a brief description of the species that will be analysed, which are cod and redfish and the Icelandic groundfish survey conducted by MRI is explained in details. In the next chapter there will be a detailed description of the dataset and the data pre-processing phase is covered.

3. Collected Data

The MRI began collecting data in 1985 and have collected data every year since then. The reason for these measurements in the beginning was to evaluate the size, condition, and spread of demersal fish. Figure 3.1 below shows all the areas where the data collection takes place. Today, there are two trawlers and two research vessels that are responsible for the data collection. A detailed description of the data is outlined in section 3.1 and then, in section 3.2, the data pre-processing phase is described, which includes any type of processing performed on the raw data to prepare it for another processing procedure.

3.1. Data description

For each tow the crew must fill out a station sheet which contains all the environmental measurements, trawl catch data, and the trawl station data. This information is then inserted into a database for further analysis and processing. The raw data that includes all stations and species consists of 59 variables and 354,544 observations where each observation represents a single tow. All the relevant data that the dataset provides and are thought to be relevant for this kind of investigation are listed below. There will be a more detailed description of them, along with the results and conclusions of the research.

The data collection can be divided into three categories: Trawl station data (tow information), trawl catch data, and environmental measurements. The recorded trawl station data contains position, time in hours and minutes at the beginning and end of each trawl haul, trawling direction, depth of the tow in meters, distance towed, towing time, and trawling speed. The tow starts when the trawl touches the bottom and ends when the hauling of the trawl starts. The geographical location of the station is registered as latitudinal and longitudinal coordinates (in degrees, minutes and seconds converted to decimal minutes) according to GPS calculations.

The trawl catch data included length measurements and age determination from examining otoliths in addition to sex determination. The sampling of otoliths was performed in a length-stratified procedure for each fish species in each of 10 sub-

3. Collected Data

areas. No weighing of fish was done at sea and the biomass calculation presented in this paper are based on constant length-weight relationships from earlier data [Pálsson et al., 1989].

The environmental factors recorded are wind direction, wind speed, sea condition, weather conditions, cloud coverage, air-, surface- and bottom temperature, barometric pressure, and ice conditions. In this thesis, only a few of the environmental factors listed above will be used for the investigation. The survey handbook also provides a detailed description of the data collection [Sólmundsson et al., 2016].

3.2. Data preprocessing

The statistical software R was used for the analysis and processing of the data. The R package *fjolt* that was created by MRI is used to access and analyse the data.

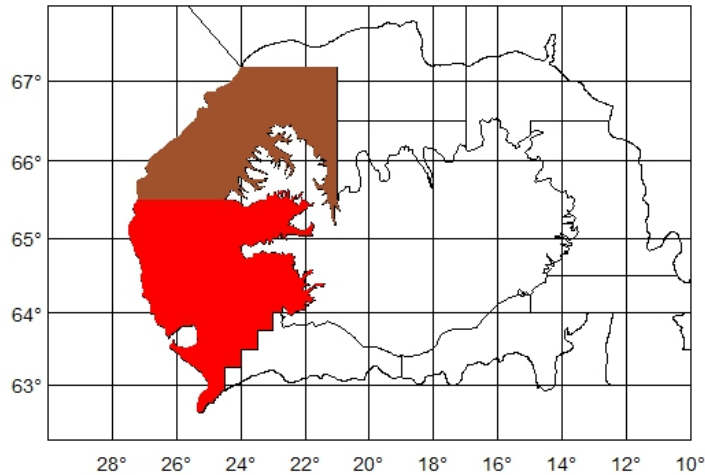


Figure 3.1: Geplot from R of the two areas that were used in this research. Area 1 is marked with red colour and area 2 with brown.

In the pre-processing phase the dataset was filtered by choosing appropriate samples of the areas around the Icelandic coast and fish species for the investigation. Simple data analysis was performed to discover the most useful parts of the data with re-

gards to missing values and amount of observations. In figure 3.1 the two areas that were chosen can be seen; both of which have similar characteristics with regards to sea temperature, depth, *etc.* Inside these two areas it was possible to investigate many species, but cod, redfish, haddock, and catfish, provided the most useful information. It was decided to investigate cod and redfish since the data related to those two species provided enough quality in order to obtain relevant results. When the data had been filtered it was possible to use four subsets for the investigation, where each contained the appropriate species and area. The catch volume in each tow was always greater than 0 in terms of both species and areas. All missing values were removed before performing calculations and while visualizing the data. The code generated for the pre-processing phase can be seen in Appendix A.

A histogram and a normal probability plot were created to investigate the distribution of the data as can be seen on figure 3.2 and figure 3.3. Figure 3.2 shows the distribution of the log transformed catch data for both cod and redfish in area 1 and figure 3.3 represents area 2.

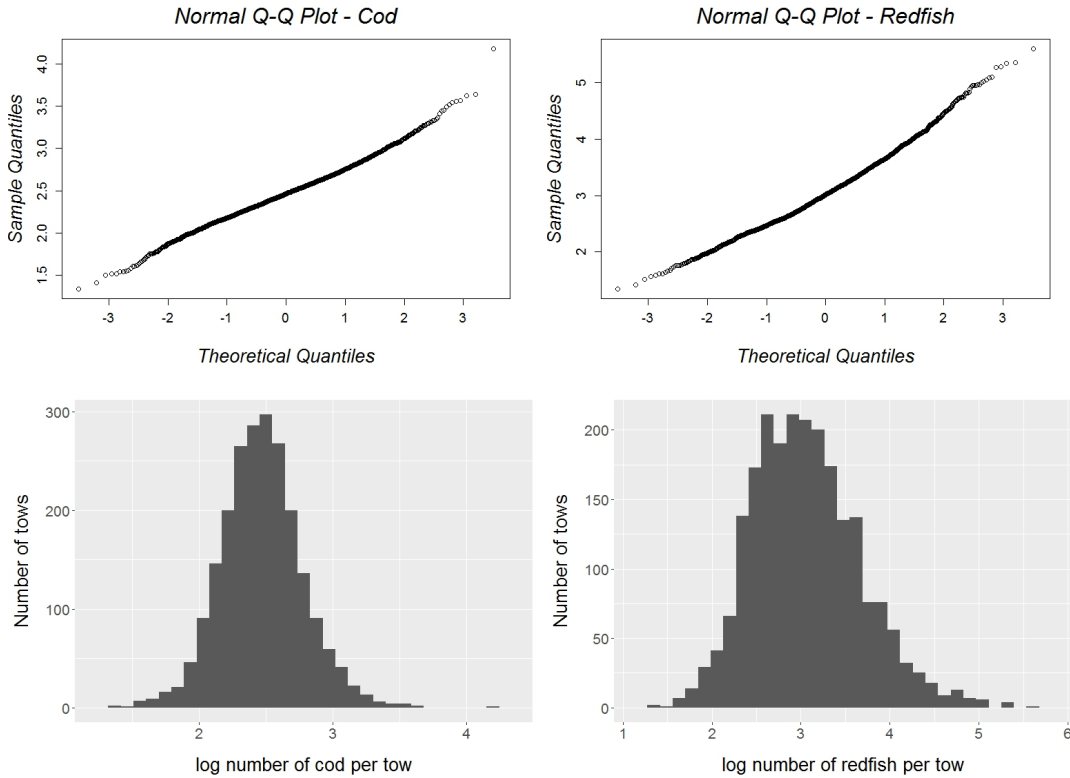


Figure 3.2: A normal probability plot and a histogram of the log transformed catch data for both redfish and cod in area 1.

3. Collected Data

These graphical techniques indicate that the raw catch data is skewed and therefore needs a transformation. Both histograms that show the frequency of the log-transformed catch data and the normal distribution plot indicate that a log-transformation can be used when generating a linear model since it is normally distributed, so the relationship between the number of catch per tow for both species and each variable will be viewed with a log-transformation of the catch hereafter.

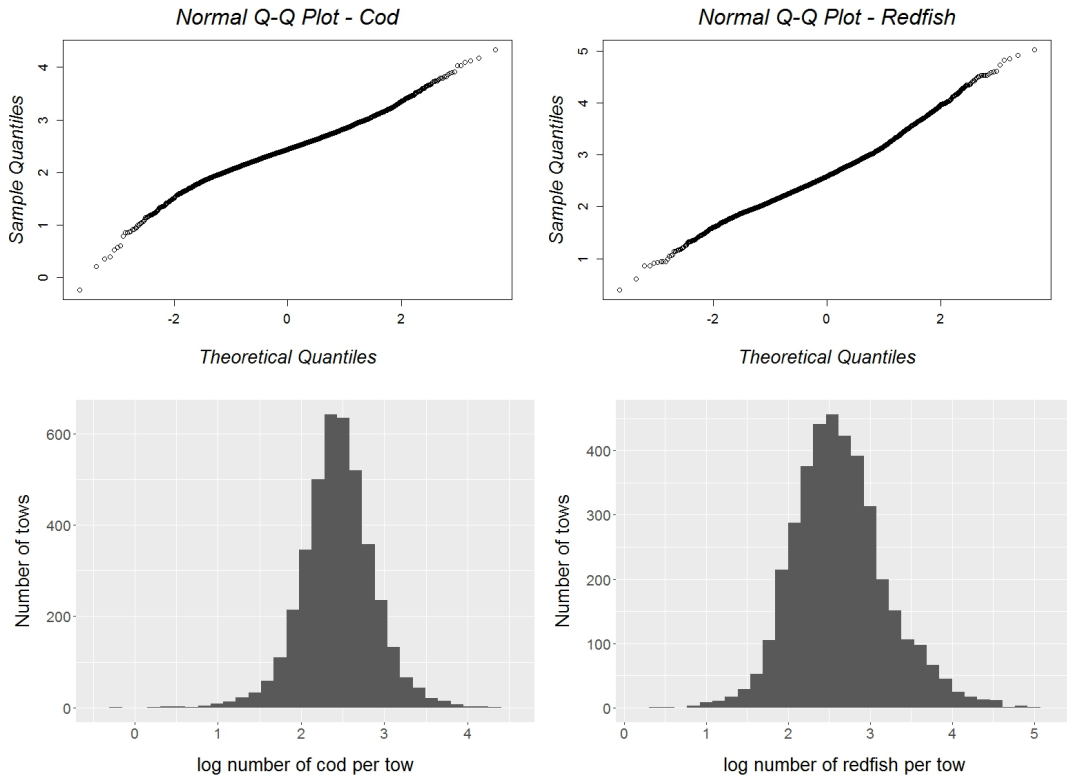


Figure 3.3: A normal probability plot and a histogram of the log transformed catch data for both redfish and cod in area 2.

3.3. Summary

This chapter described in details the data gathered by the MRI. All the processing done on the raw data is then introduced. In the next chapter the methodology used for this investigation will be introduced.

4. Methods

In this chapter an overview of the methodology used to generate a linear model in order to analyse the significance of each environmental factor against the catch rate will be viewed.

4.1. Linear regression

In statistics, linear regression is an approach for modelling the relationship between a dependent variable (the response) and one or more independent variables (the explanatory). For more than one explanatory variable, the process is called a multiple linear regression. The goal is to use a linear regression analysis to investigate how much of the variation in the response can be explained by the explanatory. The simplest way to describe this is by imagining that we have a specific dataset where linear regression can be used to fit a predictive model to the observed dataset containing the explanatory variables X and the response y . The fitted model can then be used to make a prediction of the values of y . In multiple linear regression, the model takes the form:

$$y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip} + \epsilon \quad i = 1, \dots, n \quad (4.1)$$

y_i is called the response, or the predicted value, $x_{i1}, x_{i2}, \dots, x_{ip}$ are the explanatory variables, or the predictors and β_0 is called the intercept. β is the regression coefficient and describes the effects between the response and the predictors. ϵ is the error variable or noise, this variable captures all other factors which influence the response variable y_i other than the predictors $x_{i1}, x_{i2}, \dots, x_{ip}$ [statistics, n.d.a].

The standard method to fit a linear regression model is known as least squares regression which is based on the vertical distance of the data points from the prediction line that is generated in a way to minimize the sum of squared residuals. Each vertical distance is the difference between a known value for the dependent variable Y and the value of the prediction \hat{Y} made by the created model [Pandis, 2016].

4. Methods

This difference between the observed values and the predicted values is known as a residual, the response residuals can be described using the following equation:

$$r_i = Y_i - \hat{Y}_i, \quad i = 1, 2, \dots, N \quad (4.2)$$

When building a model using linear regression all the variables need to be multivariate normal. This assumption can be viewed by generating a histogram. Histogram is an estimate of the probability distribution of a continuous variable. When the data is not normally distributed a non-linear transformation, *e.g.*, log-transformation might fix this issue. [Xiao et al., 2011].

4.1.1. ANOVA for regression

When determining the quality of a linear model an analysis of variance (ANOVA) table is often used. They provide useful information of how well the model fits the data and forms a basis for a test of significance. The ANOVA table consists of all the predictors where each of them is estimated by the degrees of freedom (DF), the sum of squares (SS), the mean sum of squares (MS), the F-statistic (F), and the p-value. In our case we are using a multiple linear regression where the objective is to fit a regression line for a response variable (the number of catches in our case) using more than one explanatory variable (all the environmental and tow factors). Simple linear regression uses only one explanatory variable when generating a linear model. The ANOVA calculations are almost the same as for simple linear regression calculations, except that the degrees of freedom describe the number of explanatory variables included in the model.

Table 4.1: Simple ANOVA table for multiple linear regression

| Source | Df | SS | MS | F-value | p-value |
|--------------|-------|--------------------------------|---------|---------|---------|
| Model | p | $\sum (\hat{y}_i - \bar{y})^2$ | SSM/DFM | SSM/DFM | |
| Error | n-p-1 | $\sum (y_i - \hat{y}_i)^2$ | | SSE/DFE | |
| Total | n-1 | $\sum (y_i - \bar{y})^2$ | SST/DFT | | |

In table 4.1 a simple ANOVA table for multiple linear regression can be seen. For p explanatory variables the model's degrees of freedom (DFM) is equal to p, the error degrees of freedom (DFE) is equal to $n - p - 1$ where n is number of observations in the dataset used to generate the linear model and the total degrees of freedom (DFT) is equal to $n - 1$ or the sum of DFM and DFE. SSM or the model's sum of squares is the variation in mean response y. The error sum of squares, or ESS, is the residuals error and the total sum of squares, or SST, represents the total variation in the response y. The mean square column is simply the sum of squares

divided by the degrees of freedom. The F-value provides a statistic for testing the hypothesis, which states that $\beta_j \neq 0, j = 1, 2, \dots, p$ against the null hypothesis that $\beta_1 = \beta_2 = \dots = \beta_p = 0$. Large values of the test statistics provide evidence against the null hypothesis which means that at least one of the explanatory variables are linearly related to the response variable. The p-value provides information about each explanatory variable, whether it has a significant effect on the response and how strong these effects are; lower p-values indicates that the explanatory variables have a greater significant effect on the response variable.

There is also a ratio $SSM/SST = R^2$ that the ANOVA table provides, it is known as the squared multiple correlation coefficient. This value is the proportion of the variation in the response variable that is explained by the explanatory variables, *i.e.*, how good the prediction is. However R^2 does not penalise for the number of explanatory used in a model, so it is not suitable for comparing models with different sets of explanatory variables [statistics, n.d.b].

4.1.2. Polynomial regression

Sometimes, a plot between the predictors and the response can suggest there is a non-linear relationship between them. Polynomial regression is a method that can be used to account for such a relationship. Polynomial regression fits a non-linear relationship between the conditional mean of the response y and the explanatory x denoted $E(y | x)$. A simple polynomial regression model with a single predictor X can be written as:

$$y = \beta_0 + \beta_1 X + \beta_2 X^2 + \dots + \beta_h X^h + \epsilon \quad (4.3)$$

where h is the degree of the polynomial. The function *poly* in *R* is useful if you want to get a polynomial of high degree, because it avoids to explicitly write the formula. The *poly* function is used in this investigation and the command $y = \text{poly}(x, z, \text{degree} = 2)$ would result in a second degree polynomial which can also be written as:

$$y = \beta_0 + \beta_1 x + \beta_2 z + \beta_3 x^2 + \beta_4 z^2 + \beta_5 xz \quad (4.4)$$

Even though polynomial regression fits a non-linear model to the data it still considered linear regression since it is linear in the regression coefficients $\beta_1, \beta_2, \dots, \beta_h$ [of Science, 2016].

4.2. Summary

This chapter explained the methodology used in order to perform and understand the statistical analyses introduced in this thesis. First the linear regression is explained, which is used for modelling the relationship between a dependent variable which are all the environmental- and towing factors and the response variable which is the catch rate of certain species in this case. The ANOVA table is then explained which is often used when determining the quality of a linear regression model. All this is followed by a explanation of polynomial regression. In the next chapter, box plots are used to visualize the relationship between the catch rate for both cod and redfish and all the environmental- and towing variables that were mentioned before.

5. Statistical summary

The advantage of using the groundfish survey data collected by MRI is that it contains a great variety of recorded measurements for every tow which can be used to build, and test, a linear regression model. However, it is not guaranteed that all the relevant factors which are likely to affect the catch rate are included in the groundfish survey dataset. So the analysis and results will be limited by the data that the groundfish survey provides. In this section the factors included in the dataset that are thought to have an effect on the catch performance of redfish and cod will be analysed.

5.1. Box plots analysis

All the figures displayed below are box plots of the log number of catches per tow against several environmental- and towing factors. Such plots can provide us with a hint of which factors are likely to affect the catch rate and in the process help suggest relevant explanatory (predictors) variables, when building a linear model. However, they are not helpful when the combined effects of several factors are examined because only one factor can be examined at a time. The red boxes represent area 1 and the brown boxes area 2. The coloured boxes show the middle 50% of the data or the first quartile to the third quartile (from 25% to 75% of the data). The black line represents the median for a certain interval, the median was preferred over the mean because it is less sensitive to extreme observations. Dotted lines are drawn to the extreme points but are not made larger than 1.5 times the height of the coloured box and then there are shown extreme points in a vertical line that exceed the dotted lines. The width of the boxes shows the variety of observations inside each interval which is often useful to investigate the quality of the results [Reese, 2005].

The intervals for all the continuous variables are generated by sorting every measurement of the catch data, from the lowest value to the highest and dividing them into equal intervals, or as evenly as possible. It was decided to have the same intervals for certain factors inside both areas to give a more accurate and comparable result. However, every area is different and provides different measurements, so each

interval will not include exactly the same amount of observations, but this method provided a relevant amount of measures in each interval. The exact values for each interval can be seen in Appendix A for all the variables.

5.2. Surface- and bottom temperature

As can be seen in figure 5.1 both bottom- and surface temperature have an effect on the catch rate, especially for redfish. A significant effect of bottom- and surface temperature for redfish in both areas can easily be seen and the log number of catch per tow (response) increases with increasing temperature. The optimal surface temperature seems to be somewhere between 6.5°C and 9°C but in terms of bottom temperature the optimal value is between 5.9°C and 8°C. This matches the description of the redfish before, where it was confirmed that the redfish is mostly found within a sea temperature range of 3 to 8°C. It is assumed that surface- and bottom temperature give similar results since they are dependent on each other, *i.e.*, bottom temperature increases with increasing surface temperature at the same location. A strong correlation is not obvious when analysing the catch rate for cod against sea temperature, but a relationship can be observed where the average catch per tow decreases with increasing temperature in area 1. The response seems to be dependent on the sea temperature squared in area 2, where the response seems to increase with increasing sea temperature. The increase is probably not linear but more akin to a parabolic increase. An optimal bottom temperature might be around -2°C and 2.6°C for cod but since the box is significantly smaller than the others, the results might be misleading as the number of observations are too few. However, it is hard to identify any significant difference in area 2.

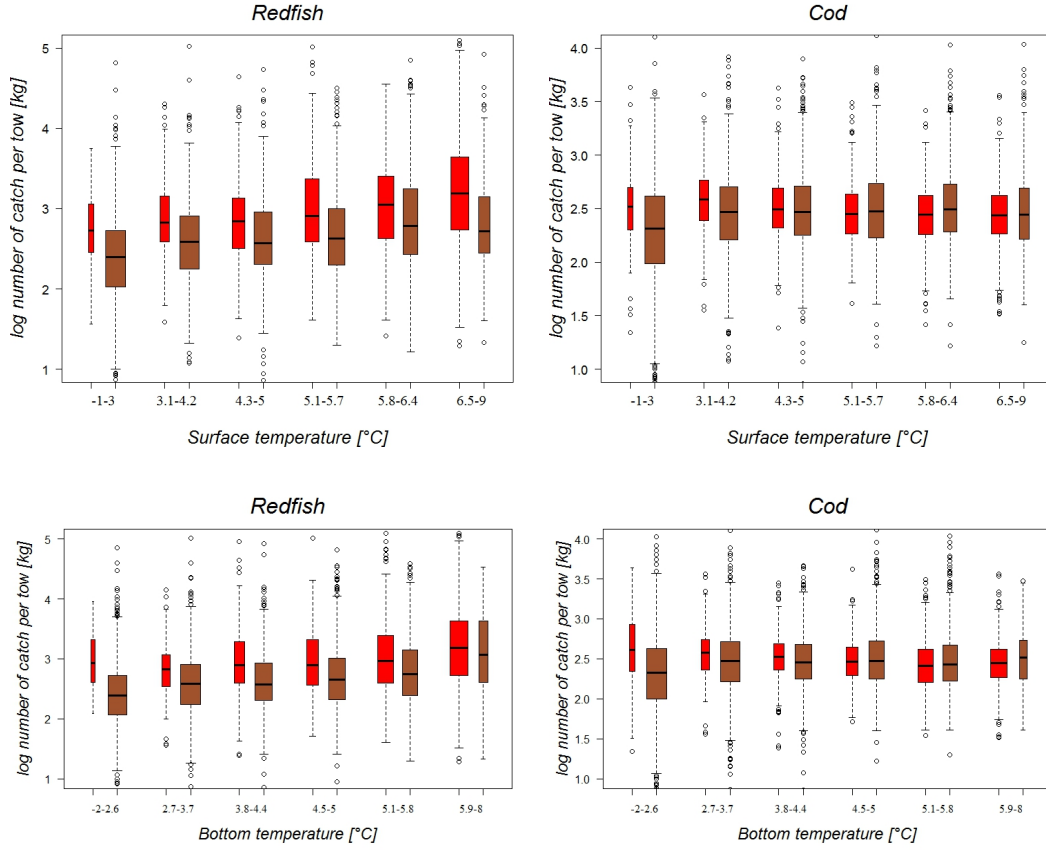


Figure 5.1: Box plots of the log number of catch per tow against surface- and bottom temperature for both redfish and cod. Surface- and bottom temperature are measured in $^{\circ}\text{C}$ and divided into six proper intervals. The red and brown boxes represent area 1 and area 2 respectively.

5.3. Depth of the tow

The depth is measured in the beginning and at the end of each tow. The data also provides the mean value of these two measurements and it was decided to use that value to get one representative depth quantity. Appropriate intervals for depth were generated using the method that was described above. An obvious relationship between depth and the catch rate can be seen in figure 5.2. In area 1 for both redfish and cod the catch rate seems to be dependent on the depth squared. The catch rate seems to be dependent on the depth squared for cod inside of area 2. However in both cases it is not clear that the depth squared gives a good fit.

5. Statistical summary

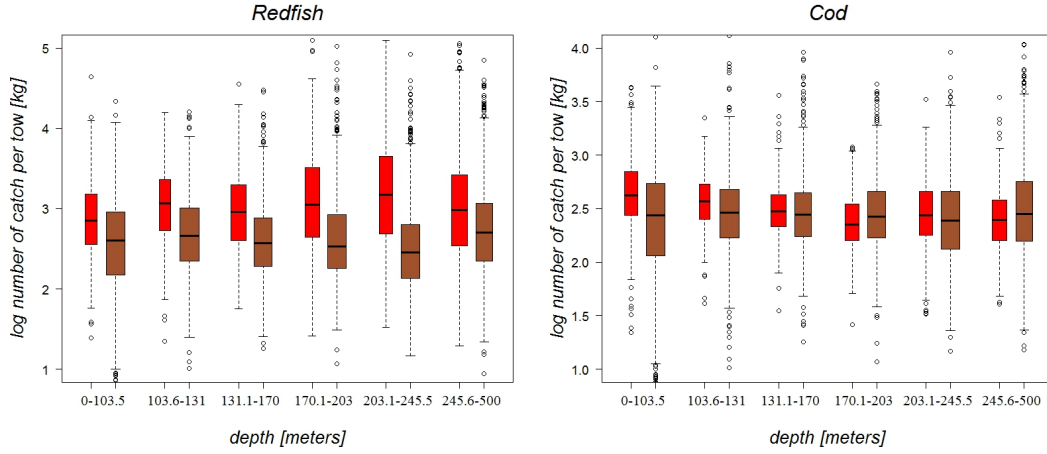


Figure 5.2: Box plots of the log number of catch per tow against depth of the tow for both redfish and cod. The depth is measured in meters and divided into six proper intervals. The red and brown boxes represent area 1 and area 2 respectively.

5.4. Weather and sea condition

Figure 5.3 shows box plots of the log number of catch per tow against weather- and sea conditions. Both the weather- and sea conditions are discrete variables so it is not possible to divide them into equal intervals, similar to what was done before. Because of this, certain intervals contain a small amount of observations which can lead to misleading results. Each value represents a certain weather- or sea condition, like clear skies or rain, and the condition gets worse as the value increases. These two variables are dependent on each other since a bad weather conditions will probably lead to poor sea conditions. It is impossible to see any striking relationships for both variables, except for the sea condition in area 1, where the optimal fishing condition seems to be between 0 and 1, which represents a calm sea. Most of the intervals contain a small amount of observations which is likely because it is impossible to fish under the worst weather- and/or sea conditions. The skipper can also choose when to tow and he will likely prefer certain conditions based on experience. It is also important to point out that it can be difficult to estimate these conditions; it depends on the individual performing the measurements and his daily mood.

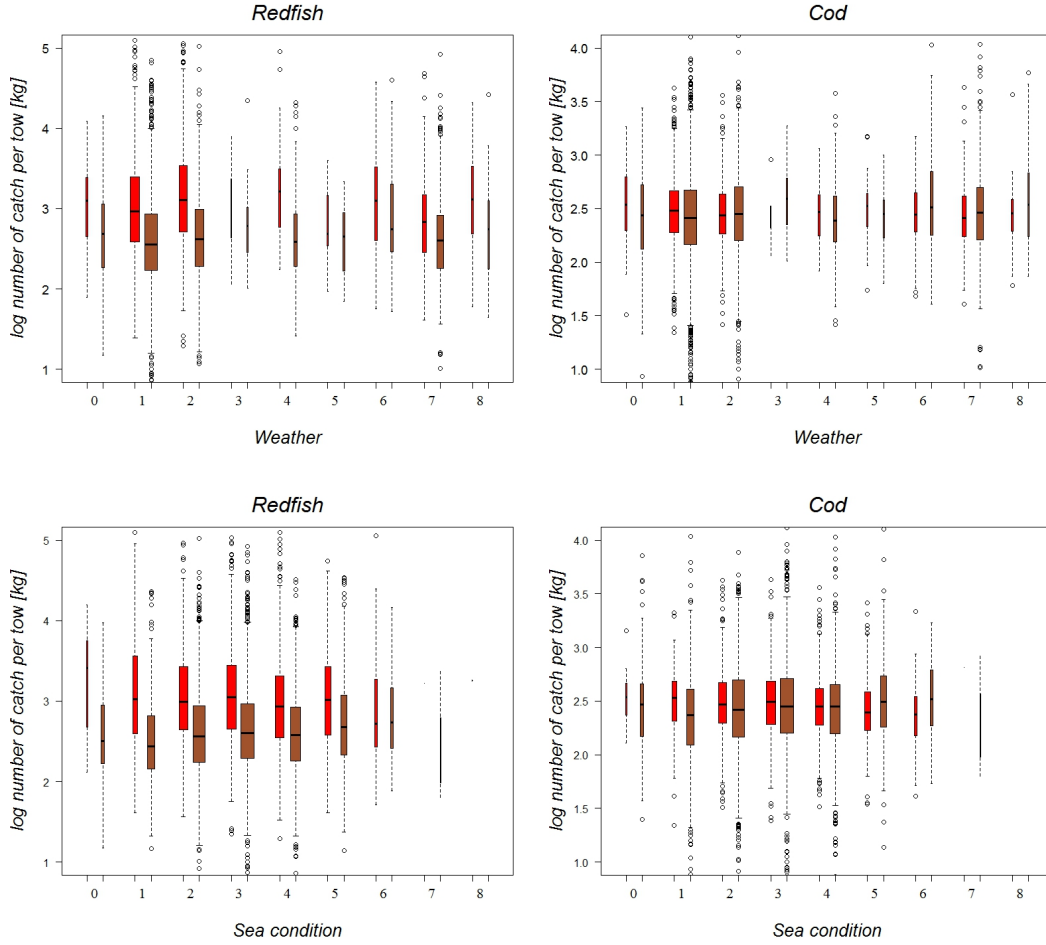


Figure 5.3: Box plots of the log number of catch per tow against weather- and sea conditions for both redfish and cod. Both variables are discrete so each interval represents a certain state of condition. The red and brown boxes represent area 1 and area 2 respectively.

5.5. Time of the tow

Luminosity was determined from the time of the tow. All the surveys take place in March and the brightness in March was determined from the time of sunrise and sunset in March. This variable is not completely accurate as the time of sunrise and sunset changed daily, but each interval stayed the same throughout the month. Three intervals were generated to define night tows, day tows, and tows that take place at twilight. It is assumed to be dark between 19:00 and 06:00, bright between 08:00 and 16:00 and twilight from 06:00 and 08:00 and then again from 16:00-19:00. It can be seen in figure 5.4 that the effect of luminosity on expected redfish catch differs for each area, the cod catch is highest when it is bright inside area 1 but

5. Statistical summary

lowest under the same conditions inside area 2. However, the expected cod catch is lowest when it is bright inside both areas.

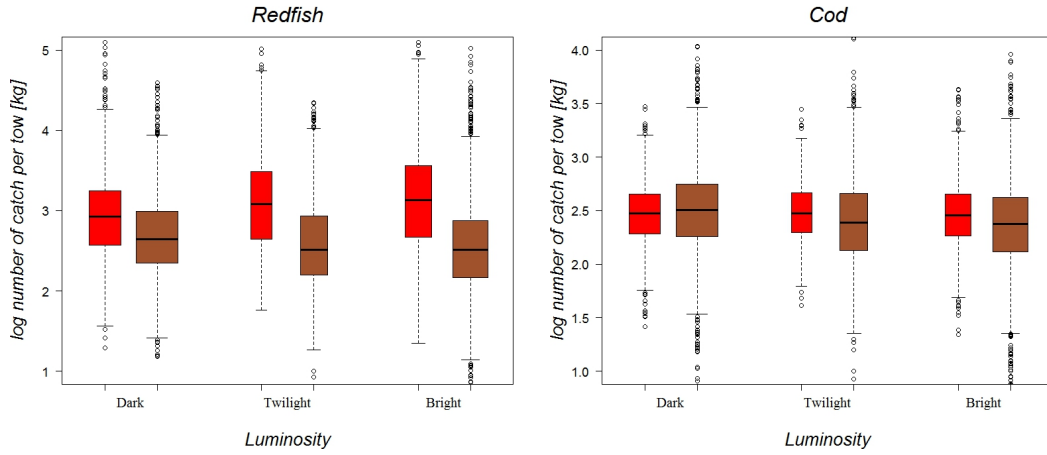


Figure 5.4: Box plots of the log number of catch per tow against luminosity for both redfish and cod. The luminosity is divided into three proper intervals, dark, twilight and bright. The red and brown boxes represent area 1 and area 2 respectively.

5.6. Barometric pressure

Figure 5.5 shows box plots of the log number of catch per tow against barometric pressure. The relationship between the response and the barometric pressure can be described with a parabola based on the plots. Even though it seems like the catch rate is affected by the barometric pressure for both species, this variable cannot be included in the model because there were too many records missing, or about 30% of the total amount of records.

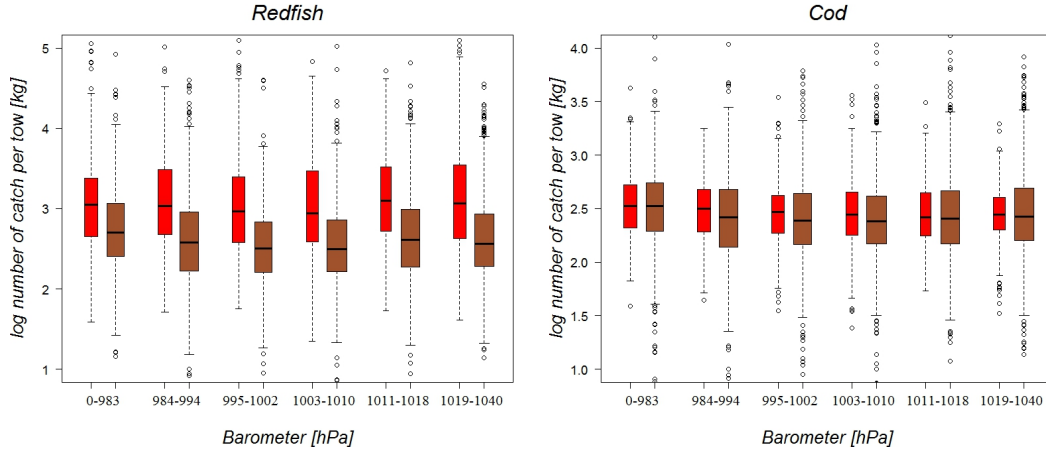


Figure 5.5: Box plots of the log number of catch per tow against barometric pressure of the tow for both redfish and cod. The barometer pressure is measured in hPa and divided into six proper intervals. The red and brown boxes represent area 1 and area 2 respectively.

5.7. Vessel effects

Figure 5.6 shows box plots of log number of catch per tow against the vessel number, which is a discrete variable, *i.e.*, each number represents a specific vessel. The cod- and redfish catch seems to change in terms of different vessels but it is hard to identify exactly where the effects come from. It can be because of different crew efforts or the capacity of the vessels. Each vessel was probably allocated certain locations to catch fish, which affects the catch rate. This variable can identify the skipper effect which were covered above, since captains are a constant factor of vessels as each captain has their own vessel.

5. Statistical summary

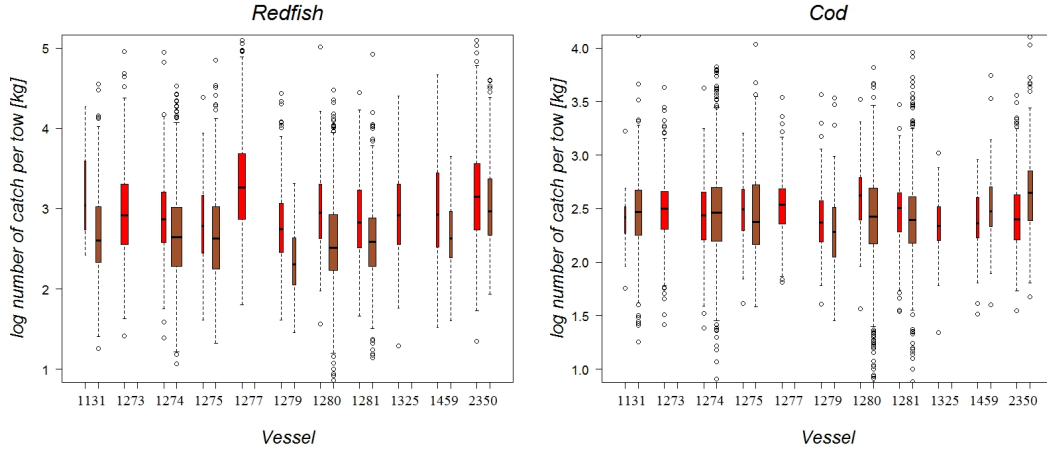


Figure 5.6: Box plots of the log number of catch per tow against vessel for both redfish and cod. The vessel variable is discrete so each interval represents a vessel number. The red and brown boxes represent area 1 and area 2 respectively.

5.8. Towing length and towing time

Figure 5.7 shows box plots of log number of catch per tow against towing length and towing time. It was considered to investigate the towing speed effects along with the other two variables, but that would likely prove to be unnecessary since towing length and towing time indirectly include the towing speed. Towing length is measured in nautical miles (Nm) where one nautical mile is approximately 1,852 meters. All these towing variables are dependent on each other, *i.e.*, they interfere with each other. Towing length and towing speed also give the same results where in most cases the towing length is around 4 Nm and is towed for 60 minutes, making the speed 4 Nm/hr. It can be seen that towing time and towing length show a linear relationship with the log number of catch per tow, for both redfish and cod. However, almost all the observations for towing length seem to lie around 4 nautical miles as mentioned already, so it is difficult to distinguish the effect of different sizes of towing lengths.

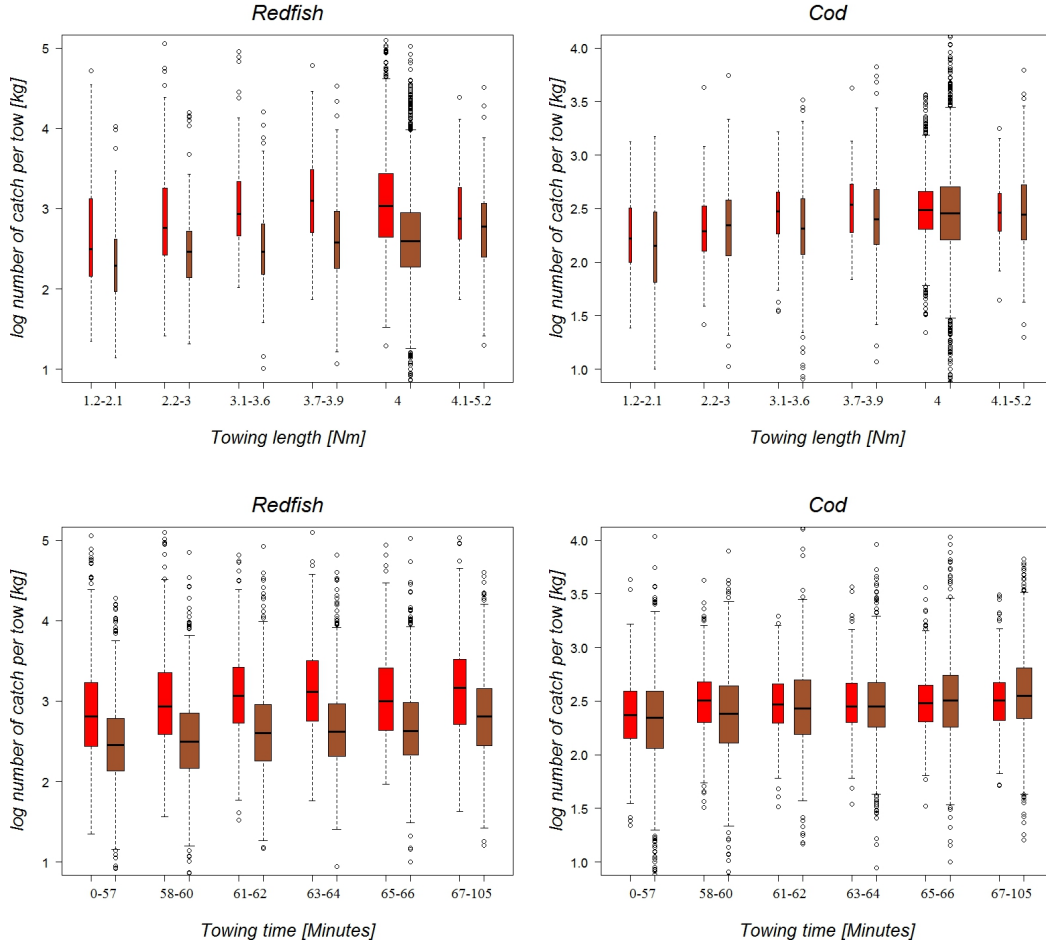


Figure 5.7: Box plots of the log number of catch per tow against towing length and towing time for both redfish and cod. Towing length is measured in nautical miles (Nm) and the towing time in minutes. Both variables are divided into six proper intervals. The red and brown boxes represent area 1 and area 2 respectively.

5.9. The degree of latitude and longitude

For each tow in the survey, the latitude and longitude is recorded both in the beginning and at the end of each tow. The mean values of both the latitude and longitude are used to get two position variables that are comparable with the log number of catch per tow. It is likely that the location of the tow has a significant effect on the catch performance. However, it is not obvious what causes the vessel to catch more fish in one location than another. It may be because certain sea- or environmental conditions at specific areas are more suitable for cod in one instance or redfish in another, however, other factors are likely in play. We can also come back to the vessel and skipper effects where the location of the tow is probably based

5. Statistical summary

on their decision. As can be seen on figure 5.8 each interval is only marked with the area, this is done because the latitude and longitude degrees would be unnecessarily complicated. The objective is not to see the exact location where most of the fish is caught, but to find out if there is any significant difference in catch performance between locations. Area 1 represents the south-western survey area and area 2 represents the north-western survey area. The cod catch can be described with a 2 or a 4 degree polynomial in terms of both the latitude and the longitude, but there is no obvious trend.

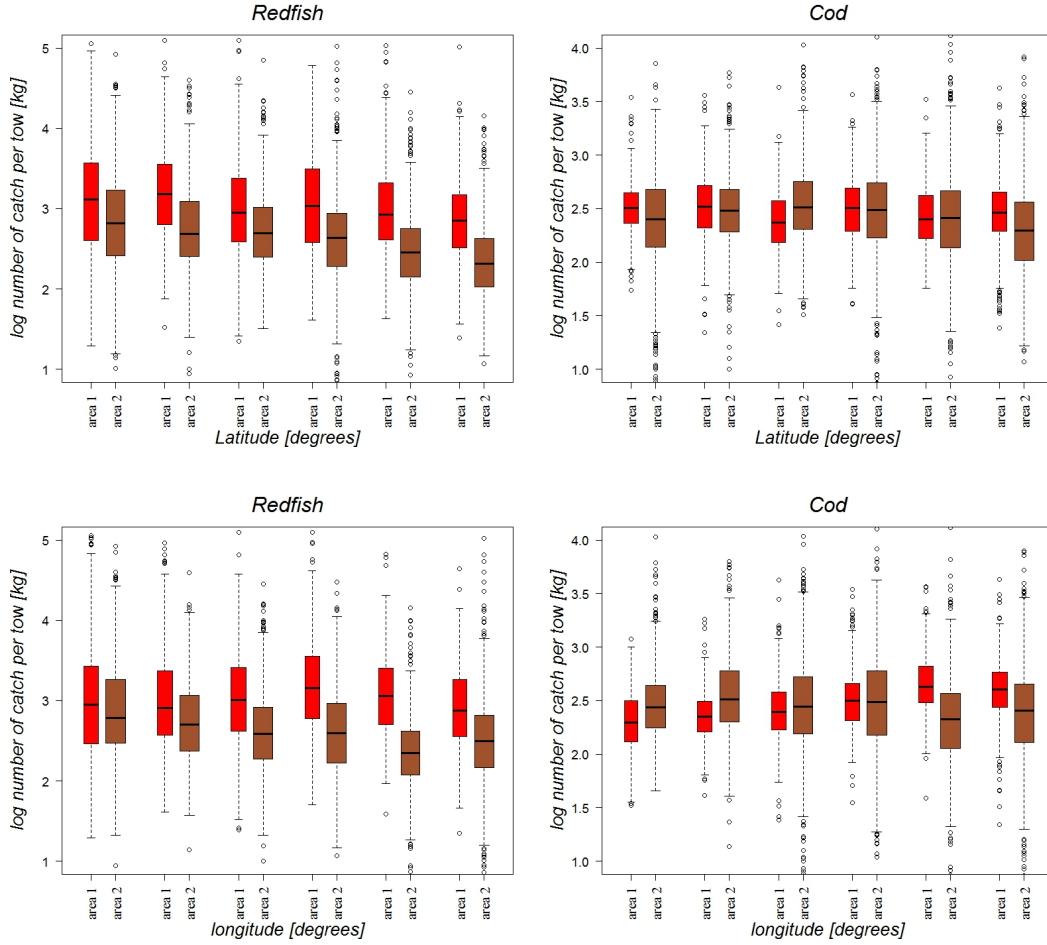


Figure 5.8: Box plots of the log number of catch per tow against the degree of latitude and longitude for both redfish and cod. The red and brown boxes represent area 1 and area 2 respectively.

It was sometimes hard to identify if there were any relations between the catch rate and the chosen explanatory variables using the box plots or what kind of relations they would be. However, an attempt was made to use the most appropriated polynomial to describe each relationship when generating the model, which will be the topic of the next chapter. The linear model analysis will give a clearer view when

investigating the combined effects of each factor in terms of the catch performance.

5.10. Summary

In this chapter the relationship between each variable and the catch rate of cod and redfish was visualized with box plots, along with discussion on how each factor affects the catch performance. Based on these plots, polynomials were applied to describe the relationship which will be used when generating a linear regression model. The topic of next chapter is using linear regression to investigate the combined effects of the variables on the catch rate.

6. Linear regression and analysis

In this chapter the analysis of variance for the linear model generated is presented. There was only one model generated but different parts (subsets) of the data are fitted to it, each part representing a specific species and area. The model was generated with the statistical summary above in mind. Another approach would be to generate four different models where each of them represents one species inside a certain area. However, one of the objectives of this thesis is to attempt to identify if there are any differences between areas in terms of the relationship between the response and the predictors and include environmental- and towing factors that can explain these differences. Polynomials were used in instances where the box plots showed that the relationship between the response and the predictors can be described with a nth degree polynomial. Weather, sea conditions, and vessels, are all discrete variables or qualitative variables where they are divided into appropriate categories. All the missing values were removed from the data before the model was fitted to each subset. The following model will be investigated:

$$\begin{aligned} \log(y_i) = & \beta_o + \text{poly}(\text{latitude}_i, \text{longitude}_i, 4) + \text{poly}(\text{depth}_i, 2) + \\ & \text{poly}(\text{surface temperature}_i, 2) + \text{poly}(\text{bottom temperature}_i, 2) + \text{log}(\text{light intensity}_i) + \\ & \text{factor}(\text{weather}) + \text{factor}(\text{sea condition}) + \text{factor}(\text{vessel}) + \\ & \text{towing speed}_i + \text{towing time}_i + \text{towing length}_i + \epsilon_i \end{aligned} \quad (6.1)$$

$\log(y_i)$ denotes the log number of certain species caught in tow i , β_o is the intercept and ϵ_i is the error variable, or noise, which captures all influence on the response variable that the explanatory variables cannot account for.

6.1. Areas analysed separately

Table 6.1: Total variation in percentages that each factor explains when it is the only factor in the model

| Source | Cod Area 1 | Cod Area 2 | Redfish Area 1 | Redfish Area 2 |
|----------------------------|---------------|---------------|-------------------|-------------------|
| poly(latitude,longitude,4) | 24 | 10.4 | 11.23 | 18.2 |
| poly(depth,2) | 7.1 | 0.54 | 4.8 | 0.5 |
| poly(surface temp,2) | 1.6 | 5.78 | 5.55 | 8 |
| poly(bottom temp,2) | 1.9 | 3.76 | 4.05 | 8.9 |
| luminosity | 0.005 | 0.22 | 1.13 | 0.03 |
| factor(weather) | 0.48 | 0.67 | 2.56 | 0.61 |
| factor(sea condition) | 0.87 | 0.75 | 0.88 | 1 |
| factor(vessel) | 4.68 | 2.1 | 8 | 6.54 |
| towing speed | 0.14 | 1.76 | 0.64 | 1.73 |
| towing time | 3.76 | 4.3 | 2.84 | 3.46 |
| towing length | 4.17 | 2.23 | 2.3 | 1.38 |

Table 6.2 provides information on how well the model fits the log-transformed catch data for cod in area 1. The number of observations for this part of the data is 2,234 tows. The last column (% expl.) shows the percentage of the total variation in the catch data that each explanatory variable explains. We can see that the factors that show the most significant effect on the cod catch in area 1 are the depth, both surface- and bottom temperature, towing time, vessel, and location (latitude,longitude). The location explains about 24% of the variation and the total model fit is approximately 35.8% of the variation, so the location has by far the greatest impact on the cod catch. But, having the *poly(latitude, longitude, 4)* predictor as the first term in the model might influence the results of other factors, since both depth and surface- and bottom temperature are dependent on the location where the tow takes place. So the majority of the depth, surface- and bottom temperature effects could be explained by the latitude and longitude location, however, they are still a useful part of the model since they are found to have a significant effect on the cod catch. Latitude and longitude still remains the factor that contributes by far the most to the variance explanation of the response. To emphasize this even further, when latitude and longitude is placed as the last predictor in the model, it still accounts around 16 % of the total variation. This applies for cod and redfish in both areas, despite the variation explanation being different between areas. It could also be sufficient to include only one temperature factor, since they are not independent, in fact, they are positively correlated, *i.e.*, the bottom temperature increases with increasing surface temperature at the same location. In the same vein, towing time, towing speed and towing length are all dependent on each other in some way. However, it seems like the majority of the effects of towing speed and towing length can be explained by

Table 6.2: Analysis of variance for cod in area 1

| Source | Df | SS | MS | F-value | p-value | % expl. |
|----------------------------|------|-------|-------|---------|-----------|---------|
| poly(latitude,longitude,4) | 14 | 50.2 | 3.6 | 58.3 | 0.000 *** | 24 |
| +poly(depth,2) | 2 | 1.2 | 0.59 | 9.5 | 0.000 *** | 0.56 |
| +poly(surface temp,2) | 2 | 0.76 | 0.38 | 6.2 | 0.002 ** | 0.36 |
| +poly(bottom temp,2) | 2 | 2.32 | 1.16 | 18.8 | 0.000 *** | 1.11 |
| +luminosity | 1 | 0.02 | 0.02 | 0.33 | 0.56 | 0.01 |
| +factor(weather) | 9 | 0.5 | 0.06 | 0.9 | 0.52 | 0.24 |
| +factor(sea condition) | 8 | 0.99 | 0.12 | 2 | 0.04 * | 0.47 |
| +factor(vessel) | 10 | 9.5 | 0.95 | 15.5 | 0.000 *** | 4.55 |
| +towing speed | 1 | 0.38 | 0.38 | 6.2 | 0.013 * | 0.18 |
| +towing time | 1 | 8.4 | 8.4 | 136.2 | 0.000 *** | 4 |
| +towing length | 1 | 0.45 | 0.45 | 7.4 | 0.007 ** | 0.22 |
| Total model | 51 | 74.7 | | | | 35.8 |
| Total residuals | 2182 | 134.3 | 0.062 | | | 64.2 |
| Total | 2233 | 209 | | | | |

the towing time, which makes a significant contribution to the variance explanation of the response.

The significant codes below represent the significant effect that each factor has on the catch rate of each specie inside area 1 and 2, where lower value indicates greater effects.

Significant codes: 0.000 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

The complete model in (6.1) explains 21% of the log-transformed catch data for cod in area 2, as can be seen in table 6.3. Factors that explain the highest percentage of the total variation are the location, depth, vessel, and towing time. Together, they explain approximately 18.2% of the variation, so all the other factors only account for around 2.8%. All the factors, except towing length and sea conditions, have a significant effect on the cod catch in area 2 based on the p-value in table 6.3. Again, the majority of depth, surface- and bottom temperature effects could be explained by the latitude and longitude polynomial.

In order to see the effects independently from the rankings of each factor in the model, an investigation was performed which shows the percentage of the total variation that each factor explains when it is the only factor in the model, which is the same as it would be the first factor in the combined model. The results can be seen in table 6.1. It is interesting that depth, surface- and bottom temperature, explain 7.1% , 1.6% and 1.9%, respectively, of the total variation when fitted separately for cod in area 1 but when combined, as in table 6.2 the total fit is only 2.03%.

Table 6.3: Analysis of variance for cod in area 2

| Source | Df | SS | MS | F-value | p-value | % expl. |
|----------------------------|------|-------|-------|---------|-----------|---------|
| poly(latitude,longitude,4) | 14 | 78.4 | 5.6 | 37.1 | 0.000 *** | 10.37 |
| +poly(depth,2) | 2 | 6.13 | 3.06 | 20.3 | 0.000 *** | 0.81 |
| +poly(surface temp,2) | 2 | 27.6 | 13.8 | 91.2 | 0.000 *** | 3.65 |
| +poly(bottom temp,2) | 2 | 2.24 | 1.12 | 7.42 | 0.000 *** | 0.3 |
| +luminosity | 1 | 1.13 | 1.13 | 7.5 | 0.006 ** | 0.15 |
| +factor(weather) | 9 | 4.8 | 0.54 | 3.55 | 0.000 *** | 0.64 |
| +factor(sea condition) | 8 | 2.18 | 0.27 | 1.8 | 0.07 . | 0.3 |
| +factor(vessel) | 9 | 17 | 1.9 | 12.5 | 0.000 *** | 2.25 |
| +towing speed | 1 | 4.8 | 4.8 | 32 | 0.000 *** | 0.64 |
| +towing time | 1 | 14.37 | 14.37 | 95 | 0.000 *** | 1.9 |
| +towing length | 1 | 0.01 | 0.01 | 0.09 | 0.76 | 0.002 |
| Total model | 50 | 158.8 | | | | 21 |
| Total residuals | 3950 | 597 | 0.15 | | | 79 |
| Total | 4000 | 755.8 | | | | |

However, for cod in area 2 a different result was obtained, 0.54%, 5.78% and 3.76%, respectively, when fitted separately but when combined, as in table 6.3, the total fit is 4.76%. From this we can derive that depth effects are not as strong in area 2, instead the sea temperature explains a greater proportion of the total variation. This confirms that which was concluded before, that the correlation between the temperature factors and the depth, surface- and bottom temperature effects, can be explained by the polynomial of the latitude and longitude factor. It could also mean that some of the temperature effects can be explained by differences in depth, *i.e.*, the temperature declines as the depth increases. It is worthy to mention that the depth effect is not as strong in area 2 as in area 1 when investigating the cod catch, the temperature seems to have a greater effect in area 2 than in area 1.

The towing time and towing length are likely dependent on each other since they explain 3.76% and 4.17% of the variation in cod catch in area 1, when fitted separately. However, when they are put together in a single model the towing time explains 4% while the towing length explains only 0.22%. This is because the towing time is placed first in the model, *i.e.*, the majority of the towing length effect on the cod catch is explained by the towing time which matches the statement mentioned above. However, towing speed is placed first of these three towing factors and it still only explains 0.18% of the variation, which means that it has no significant effect on the cod catch in area 1. The same result was obtained when the dataset for cod catch in area 2 was investigated.

According to the results from table 6.2, vessel effects on cod catch in area 1 are not dependent on any other factors since they explain almost the same percentage of the

variation regardless of whether it is the only factor in the model or if it is combined with all other factors, as in model 5.1. The same applies to area 2.

Table 6.4: Analysis of variance for redfish in area 1

| Source | Df | SS | MS | F-value | p-value | % expl. |
|----------------------------|------|-------|-------|---------|-----------|---------|
| poly(latitude,longitude,4) | 14 | 94.7 | 6.8 | 25.9 | 0.000 *** | 11.23 |
| +poly(depth,2) | 2 | 55.8 | 27.9 | 106.9 | 0.000 *** | 6.62 |
| +poly(surface temp,2) | 2 | 18 | 9 | 34.6 | 0.000 *** | 2.14 |
| +poly(bottom temp,2) | 2 | 1.65 | 0.83 | 3.2 | 0.04 * | 0.2 |
| +luminosity | 1 | 7.5 | 7.5 | 28.8 | 0.000 *** | 0.9 |
| +factor(weather) | 9 | 11.1 | 1.23 | 4.7 | 0.000 *** | 1.31 |
| +factor(sea condition) | 8 | 9.2 | 1.15 | 4.4 | 0.000 *** | 1.1 |
| +factor(vessel) | 10 | 35.3 | 3.5 | 13.5 | 0.000 *** | 4.2 |
| +towing speed | 1 | 0.41 | 0.41 | 1.6 | 0.2 | 0.05 |
| +towing time | 1 | 31.74 | 31.74 | 121.6 | 0.000 *** | 3.8 |
| +towing length | 1 | 1.6 | 1.6 | 6.2 | 0.01 * | 0.19 |
| Total model | 51 | 267 | | | | 32 |
| Total residuals | 2207 | 576.1 | 0.26 | | | 68 |
| Total | 2258 | 843.1 | | | | |

In table 6.4 the model fit of log-transformed catch data for redfish in area 1 can be seen. These results are similar to the ones seen in the cod catch data. The factors that are estimated to have a significant effect on the catch rate of redfish based on the analysis of variance in table 6.4 are the location, depth, surface temperature, luminosity, weather, sea conditions, vessel, and towing time. The total model fit is 32% and the number of observations are 2,259. However, the latitude and longitude polynomial does not have as great of an effect on the redfish catch as for the cod, where it explains 11.23% of the variation. Instead, the depth and surface temperature explain 8.8% of the redfish catch variation where they had a minimum effect on the cod catch in area 1, *i.e.*, together they explained 0.9% of the variation. The bottom temperature seems to be explained by the surface temperature factor, since it is placed first in the model, but it does not have an effect on the total fit to change the ranking. However, the bottom temperature still explains 0.2% of the variation and has still a significant effect based on the p-value, so it cannot be disregarded.

When the model is fitted to redfish in area 2 different results are obtained than when area 1 was investigated. The factors that are estimated to have a significant effect on the catch rate of redfish based on the analysis of variance in table 6.5 are the location, depth, surface- and bottom temperature, vessel, towing speed, and towing time. These factors together explain around 29.9% of the total variation, and the latitude and longitude polynomial still remains the most important factor in the variance reduction, explaining 18.2% of the total variation. All other factors only

Table 6.5: Analysis of variance for redfish in area 2

| Source | Df | SS | MS | F-value | p-value | % expl. |
|----------------------------|------|--------|-------|---------|-----------|---------|
| poly(latitude,longitude,4) | 14 | 228.6 | 16.3 | 71.8 | 0.000 *** | 18.2 |
| +poly(depth,2) | 2 | 12.7 | 6.4 | 28 | 0.000 *** | 1.01 |
| +poly(surface temp,2) | 2 | 45.2 | 22.6 | 99.5 | 0.000 *** | 3.6 |
| +poly(bottom temp,2) | 2 | 7.75 | 3.9 | 17 | 0.000 *** | 0.62 |
| +luminosity | 1 | 0.21 | 0.21 | 0.91 | 0.3 | 0.02 |
| +factor(weather) | 9 | 5.8 | 0.64 | 2.82 | 0.003 ** | 0.46 |
| +factor(sea condition) | 8 | 5.68 | 0.71 | 3.12 | 0.002 ** | 0.45 |
| +factor(vessel) | 9 | 50.69 | 5.63 | 24.8 | 0.000 *** | 4.04 |
| +towing speed | 1 | 5.53 | 5.53 | 24.33 | 0.000 *** | 0.44 |
| +towing time | 1 | 25.98 | 25.98 | 114.26 | 0.000 *** | 2.07 |
| +towing length | 1 | 0.19 | 0.19 | 0.85 | 0.35 | 0.02 |
| Total model | 50 | 388.4 | | | | 30.9 |
| Total residuals | 3813 | 867 | 0.23 | | | 69.1 |
| Total | 3863 | 1255.4 | | | | |

explain around 1% of the variation. The total model fit is 30.9% which is similar to the total fit obtained in area 2. The number of observations are 3,864 tows for this part of the data.

For both redfish and cod, the majority of the depth, surface- and bottom temperature effects were explained by the latitude and longitude polynomial. When looking at table 6.1 where the percentage of the total variation that each factor, independent from the ranking of them in the model, explain around 4.8%, 5.55% and 4.05% respectively, of the total variation when fitted separately for redfish in area 1. However, when fitted together, they explain 6.62%, 2.14% and 0.2%. This confirms the relationship between these factors and that depth has a greater effect on the redfish catch when fitted alongside all the other factors. The same applies to redfish in area 2, except that the temperature has a greater effect on the redfish catch than the depth, which is possibly due to the fact that the latitude and longitude polynomial explain a greater proportion of the total variation than in area 1. In other words, the majority of the depth effect can be explained by the latitude and longitude polynomial since the depth of the tow is dependent on the location of the tow. The vessel term explains 8% of the total variation in area 1 and 6.54% of the total variation in area 2 when it is placed as the first term in the model but it explains 4.2% in area 1 and 4.04% in area 2 of the total variation when fitted alongside other factors. This indicates that a small proportion of the vessel effects can be explained by some other factor in the model. The same results of the towing effects are obtained for both redfish and cod in both areas. The towing time and towing length are dependent on each other since the towing length is placed after the towing time in the model and only explains 0.19% in area 1 and 0.02% in area

2.

Based on these analyses the optimal model would be:

$$\begin{aligned} \log(y_i) = \beta_o + \text{poly}(\text{latitude}_i, \text{longitude}_i, 4) + \text{poly}(\text{depth}_i, 2) + \\ \text{poly}(\text{surface temperature}_i, 2) + \text{factor}(\text{vessel}) + \\ \text{towing time}_i + \epsilon_i \end{aligned} \quad (6.2)$$

where these factors showed the greatest effect on the fish catch when both species were considered. There were other factors that showed significant effects, but they explained such a small percentage of the total variation, so they were dismissed. When model 5.2 is fitted to the same subsets as before, it explains slightly less of the total variability of the response variable, but the model benefits from lower degrees of freedom, i.e it would likely provide more variety when fitting the model to different datasets. This can be seen in appendix B where additional ANOVA tables for the reduced model is shown. However, there is always a room for improvement and other factors could be tested in the model and perhaps achieve better result in terms of prediction of the fish catch.

6.2. Area 1 and area 2 combined

It is interesting that when model 6.1 above is fitted to cod and redfish, where area 1 and area 2 are combined (representing all the western area of the Icelandic coast) as can be seen in table 6.6 and 6.7, the model explains approximately 18.6% the total variation in the cod catch and around 34% of the total variation in the redfish catch. This could mean that some factors are missing if we are to explain the variation between areas, specifically for cod. The model fits the redfish catch data better when it consists of both areas. This is likely because the redfish is not as sensitive to changes in the ocean as the cod, which affects the catch performance as the two areas have different characteristics. The number of observations are similar for both species or 6235 for cod and 6123 for redfish.

The variance explanations in terms of each factor is similar as before for both species, where the factors that show the greatest effect on the catch performance are location of the tow, depth, surface- and bottom temperature, vessel effects and towing time.

6. Linear regression and analysis

Table 6.6: Analysis of variance for cod when area 1 and area 2 are combined

| Source | Df | SS | MS | F-value | p-value | % expl. |
|----------------------------|------|-------|------|---------|------------|---------|
| poly(latitude,longitude,4) | 14 | 73.9 | 5.3 | 41.5 | 0.000 *** | 7.65 |
| +poly(depth,2) | 2 | 10.1 | 5.1 | 39.8 | 0.000 *** | 1.05 |
| +poly(surface temp,2) | 2 | 37.5 | 18.7 | 147.3 | 0.000 **** | 3.9 |
| +poly(bottom temp,2) | 2 | 3.45 | 1.72 | 13.57 | 0.000 **** | 0.36 |
| +luminosity | 1 | 0.76 | 0.76 | 6 | 0.01 * | 0.08 |
| +factor(weather) | 9 | 2.95 | 0.33 | 2.58 | 0.006 ** | 0.31 |
| +factor(sea condition) | 8 | 1.98 | 0.25 | 1.94 | 0.05 * | 0.47 |
| +factor(vessel) | 12 | 18 | 1.5 | 11.8 | 0.000 *** | 1.87 |
| +towing speed | 1 | 5.5 | 5.5 | 43.4 | 0.000 *** | 0.57 |
| +towing time | 1 | 24.7 | 24.7 | 193.8 | 0.000 *** | 2.55 |
| +towing length | 1 | 0.3 | 0.3 | 2.4 | 0.12 | 0.03 |
| Total model | 53 | 179.2 | | | | 18.6 |
| Total residuals | 6181 | 786.8 | 0.13 | | | 81.4 |
| Total | 6234 | 966 | | | | |

Table 6.7: Analysis of variance for redfish when area 1 and area 2 are combined

| Source | Df | SS | MS | F-value | p-value | % expl. |
|----------------------------|------|--------|------|---------|-----------|---------|
| poly(latitude,longitude,4) | 14 | 515.65 | 36.8 | 143.5 | 0.000 *** | 21.9 |
| +poly(depth,2) | 2 | 64 | 32 | 124.6 | 0.000 *** | 2.71 |
| +poly(surface temp,2) | 2 | 56.2 | 28.1 | 109.5 | 0.000 *** | 2.38 |
| +poly(bottom temp,2) | 2 | 5.2 | 2.6 | 10.1 | 0.000 *** | 0.22 |
| +luminosity | 1 | 1.14 | 1.14 | 4.44 | 0.035 * | 0.05 |
| +factor(weather) | 9 | 10.96 | 1.22 | 4.74 | 0.000 *** | 0.46 |
| +factor(sea condition) | 9 | 8.4 | 0.93 | 3.63 | 0.000 *** | 0.36 |
| +factor(vessel) | 12 | 76.7 | 6.4 | 24.9 | 0.000 *** | 3.25 |
| +towing speed | 1 | 6 | 6 | 23.4 | 0.000 *** | 0.25 |
| +towing time | 1 | 56 | 56 | 218.18 | 0.000 *** | 2.38 |
| +towing length | 1 | 0.46 | 0.46 | 1.8 | 0.18 | 0.02 |
| Total model | 54 | 800.7 | | | | 34 |
| Total residuals | 6068 | 1558 | 0.26 | | | 66 |
| Total | 6122 | 2358.7 | | | | |

6.3. Summary

In this chapter the analysis of variance for the linear models generated were presented along with the conclusion on which factors have significant impact on the catch performance. The linear model was fitted to different parts of the data where each part represents a specific species and area. This was done to investigate if the results would stay consistent, independent of areas which indicates if other factors need to be included in the model to explain these differences. In the next chapter a conclusion and discussion is presented along with possible directions of future work.

7. Conclusion and Discussion

In terms of the statistical analysis done earlier there are certain factors that show repeatedly significant effects on the catch performance and therefore important determinants of the size composition of trawl catches. These factors are location, depth, surface- and bottom temperature, vessel effects, and towing time. Other factors that are included in the model, which are towing speed, towing length, weather- and sea conditions, and luminosity, only explain a small proportion of variation in the catch data which could occur due to the majority of these effects being explained by other factors in the model. This matches the conclusions of the box plots although it was sometimes difficult to identify exactly how the relationship between the response and each explanatory variable described itself. The latitude and longitude polynomial has the greatest effect on the catch performance in all cases, which indicates that the location of the tow is a critical factor when catching fish. However, the location is dependent on many other factors, *i.e.*, the catch rate can vary with respect to temperature, depth, and sea conditions, *etc.*, which are all factors that are influenced by the location of the tow. As mentioned earlier, the effects are not the same in terms of both species and areas, which indicate that some missing parameters need to be added to the model to explain these differences and achieve greater fitting results. The effects of different environmental and towing factors can also change depending on areas and species. It looks like the redfish is not as sensitive to changes in environmental conditions as the cod. Similar results are achieved when the model is fitted to redfish in all the western area (area 1 and 2 together) which points towards the cod being a more sensitive species than the redfish.

Differences in trawl catches of cod and redfish between each variable investigated may be attributed to fish behaviour. Their activity and position in relation to the bottom of the ocean may change, consequently affecting their availability to the bottom trawl. Some species might change position in terms of their feed location and some species are known to remain near the bottom during the day while swimming towards the surface during night. The preferred environment of a species is determined by many biotic and abiotic parameters, such as predator avoidance, sea temperature, salinity, depth, *etc.*. Variation in catch rates can also be due to changes in the fish population or fish catchability in a certain area. Since the location is a critical factor it is likely that the choice of skipper is important, since skippers possess unique experiences and intuition.

7. Conclusion and Discussion

The results of this thesis were similar to the ones found by other researchers like Jenný. Namely, that the environmental factors that were found to be most important are depth and bottom- and surface temperature. Jenný only investigated cod catch data and involved all the 10 areas around Iceland so her research was made with a different approach. However, her model explained 63% of the variation in cod catch data whereas my model only explained 35.8% for the cod catch data in area 1. The location also explained the greatest proportion of the variation in the catch data. This difference is likely because the data used in this thesis extends over a longer period of time and Jenný uses a different approach in terms of research areas. Since Jenný involved all 10 areas around Iceland, the catch data probably contains greater variation which would in most cases lead to less total variance explanation. However, Jenný included time (years) in the model which affects the total fit of the model, but including time was not thought to be relevant in this investigation since it would likely provide a model with less variety, *i.e.*, it fits a specific catch data well but would not give good results when tested on different catch data.

The analysis has shown that catch data for both redfish and cod from the bottom trawl surveys depend on many factors in addition to those who were investigated. These effects are not the same in terms of different species and areas. There seems to be a need for more sophisticated models to better explain the variation in both the redfish and cod catch data, but it can be difficult to identify these additional factors as we are working within a complex system.

7.1. Future work

It might be interesting to use an approach called system dynamics. System dynamics is an approach that could help us understand the non-linear behaviour of such a complex system. There are many variables that interfere with each other as we have shown. By drawing up a causal loop diagram, similar to figure 2.1 but with more details, could help us explain the relationship between each factor, how they affect each other and their causes. With that understanding it would help us to choose the relevant predictors in the model. The model analysis would then provide us with suitable information that could be used to simulate such a complex system. However, such a study would be a topic for another master thesis.

Bibliography

- Sara Adlerstein and Siegfried Ehrich. Patterns in diel variation of cod catches in north sea bottom trawl surveys. *Fisheries Research*, 63(2):169–178, 2003.
- Jenný Brynjarsdóttir and Gunnar Stefánsson. Analysis of cod catch from icelandic groundfish surveys using generalized linear models. *Fisheries Research*, 70(2): 195–208, 2004.
- Hafrannsóknarstofnun. *Karfi, Gullkarfi*. n.d. Retrieved from <http://www.hafro.is/undir.php?REFID=9&ID=39&REF=2&ID=5>.
- Fræðsluvefur Matís. *Þorskur*. n.d. Retrieved from <http://fraedsluvefur.rf.is/Undirflokkur/Undirflokkur/thorskur/>.
- PennState Eberly College of Science. *Polynomial regression*. 2016. Retrieved from <https://onlinecourses.science.psu.edu/stat501/node/324>.
- Nikolaos Pandis. Linear regression, 2016.
- G. Petrakis, D. N. MacLennan, and A. W. Newton. Day-night and depth effects on catch rates during trawl surveys in the north-sea. 58(1):50–60, 2001.
- O. Pálsson, E. Jonsson, A. Schopka, A. Stefansson, and B. Steinarsson. Icelandic groundfish survey data used to improve precision in stock assessments. *Hafrannsóknarstofnun (The Marine Research Institute in Iceland)*, 1989.
- R. Allan Reese. Boxplots. *Significance*, 2(3):134–135, 2005.
- J. Sólmundsson, B. Steinarsson, H. Karlsson, H. Björnsson, J. Pálsson, and V. Bogason. Handbók um stofnmælingu botnfiska á íslandsmiðum 2016. *Hafrannsóknarstofnun (The Marine Research Institute in Iceland)*, 2016.
- Department of statistics. *Multiple linear regression*. n.d.a. Retrieved from <http://www.stat.yale.edu/Courses/1997-98/101/linmult.htm>.
- Department of statistics. *ANOVA for regression*. n.d.b. Retrieved from <http://www.stat.yale.edu/Courses/1997-98/101/anovareg.htm>.
- Gunnar Stefánsson and Ólafur Pálsson. Bormicon. Technical report, Marine Research Institute, 1997.

BIBLIOGRAPHY

- John Sterman. System dynamics modeling: Tools for learning in a complex world. *California Management Review*, 43(4):8–25, 2001.
- Thorolfur Thorlindsson. The skipper effect in the icelandic herring fishery. *Human organization*, 1988.
- Xiao Xiao, Ethan P. White, Mevin B. Hooten, and Susan L. Durham. Use of log-transformation vs. nonlinear regression for analyzing biological power laws. *Ecology*, 92(10):1887–1894, 2011.

A. R-script

Below are all the R-scripts used in order to perform the pre-processing phase, generating all the figures and fitting the linear models to all the subsets.

```
## ----- ##
## Data preprocessing ##
## ----- ##
library(fjolt)
library(dplyr)
library(geo)
library(purrr)
library(plyr)
library(lattice)
library(jpeg)
library(ggplot2)

st <- stodvar %>%
  filter(synaflokkur == 30) %>%
  inside.reg.bcl() %>%
  #filter(area %in% c('1', '2'))

tmp.func <- function(x){
  afli.per.stod(st, teg = x, lfile = filter(all.le, synis.id
    %in% st$synis.id)) %>%
    mutate(teg = x)
}

st.m.afla <-
  1:31 %>%
  map(~tmp.func(.)) %>%
  bind_rows()

stAll = st.m.afla
stAll = do.call(data.frame, lapply(stAll, function(x)
  replace(x, is.infinite(x), NA)))
```

A. R-script

```
## ----- ##
## Dividing into proper intervals ##
## ----- ##
yht = subset(st2, teg == 1)
t = sort(yht$lon)

t[floor(length(t)/6)]
t[floor(2*length(t)/6)]
t[floor(3*length(t)/6)]
t[floor(4*length(t)/6)]
t[floor(5*length(t)/6)]
t[floor(6*length(t)/6)]
t[floor(7*length(t)/8)]
t[floor(8*length(t)/8)]

plot(t, cumsum(as.numeric(1:length(t))))

## ----- ##
## Box plot - Barometer ##
## ----- ##

#Species number
i = 1

#Area 1
lt11 = subset(st1, loftvog > 0 & loftvog <= 983 & teg == i)
lt12 = subset(st1, loftvog > 983 & loftvog <= 994 & teg == i)
lt13 = subset(st1, loftvog > 994 & loftvog <= 1002 & teg == i)
lt14 = subset(st1, loftvog > 1002 & loftvog <= 1010 & teg == i)
lt15 = subset(st1, loftvog > 1010 & loftvog <= 1018 & teg == i)
lt16 = subset(st1, loftvog > 1018 & loftvog <= 1040 & teg == i)

#Area 2
lt21 = subset(st2, loftvog > 0 & loftvog <= 983 & teg == i)
lt22 = subset(st2, loftvog > 983 & loftvog <= 994 & teg == i)
lt23 = subset(st2, loftvog > 994 & loftvog <= 1002 & teg == i)
lt24 = subset(st2, loftvog > 1002 & loftvog <= 1010 & teg == i)
lt25 = subset(st2, loftvog > 1010 & loftvog <= 1018 & teg == i)
lt26 = subset(st2, loftvog > 1018 & loftvog <= 1040 & teg == i)
```

```

boxplot(log10(lt11$afl),log10(lt21$afl),log10(lt12$afl),
log10(lt22$afl),log10(lt13$afl),log10(lt23$afl),
log10(lt14$afl),log10(lt24$afl),log10(lt15$afl),
log10(lt25$afl),log10(lt16$afl),log10(lt26$afl),
las = 2, col = c("red","sienna","red","sienna","red","sienna"),
at = c(1,2, 4,5, 7,8, 10,11, 13,14, 16,17),
names = c("", "", "", "", "", "", "", "", "", "", "", "", "" ),
ylim=c(1,4),varwidth = TRUE)

```

#Stationlabels

```

mtext("0-983", side=1, line=1, at=1.5, las=1, font=6, col="black")
mtext("984-994", side=1, line=1, at=4.5, las=1, font=6, col="black")
mtext("995-1002", side=1, line=1, at=7.5, las=1, font=6, col="black")
mtext("1003-1010", side=1, line=1, at=10.5, las=1, font=6, col="black")
mtext("1011-1018", side=1, line=1, at=13.5, las=1, font=6, col="black")
mtext("1019-1040", side=1, line=1, at=16.5, las=1, font=6, col="black")

```

#Axislabels

```

mtext("Barometer[hPa]", side = 1, line = 4, cex = 1, font = 3)
mtext("log_number_of_catch_per_tow[kg]", side = 2, line = 3, cex = 1,
font = 3)
mtext("Cod", side = 3, line = 1, cex = 2, font = 3)

```

```

## ----- ##
## Box plot - Luminosity##
## ----- ##

```

#Area 1

```

td11 = subset(st1,kl.hift>=0 & kl.hift <= 6 & teg==i)
td12 = subset(st1,kl.hift>6 & kl.hift <= 8 & teg==i)
td13 = subset(st1,kl.hift>8 & kl.hift <= 16 & teg==i)
td14 = subset(st1,kl.hift>16 & kl.hift <= 19 & teg==i)
td15 = subset(st1,kl.hift>19 & kl.hift <= 24 & teg==i)

```

```

td11 = c(td11$afl,td15$afl)
td12 = c(td12$afl,td14$afl)
td13 = c(td13$afl)

```

#Area 2

```

td21 = subset(st2,kl.hift>0 & kl.hift <= 6 & teg==i)
td22 = subset(st2,kl.hift>6 & kl.hift <= 8 & teg==i)
td23 = subset(st2,kl.hift>8 & kl.hift <= 16 & teg==i)

```

A. R-script

```
td24 = subset(st2 , kl.hift > 16 & kl.hift <= 19 & teg==i)
td25 = subset(st2 , kl.hift > 19 & kl.hift <= 24 & teg==i)

td21 = c(td21$afl , td25$afl)
td22 = c(td22$afl , td24$afl)
td23 = c(td23$afl)

## ----- ##
## Box plot - Weather##
## ----- ##

#Area 1
vt11 = subset(st1 , vedur == 0 & teg==i)
vt12 = subset(st1 , vedur == 1 & teg==i)
vt13 = subset(st1 , vedur == 2 & teg==i)
vt14 = subset(st1 , vedur == 3 & teg==i)
vt15 = subset(st1 , vedur == 4 & teg==i)
vt16 = subset(st1 , vedur == 5 & teg==i)
vt17 = subset(st1 , vedur == 6 & teg==i)
vt18 = subset(st1 , vedur == 7 & teg==i)
vt19 = subset(st1 , vedur == 8 & teg==i)

#Area 2
vt21 = subset(st2 , vedur == 0 & teg==i)
vt22 = subset(st2 , vedur == 1 & teg==i)
vt23 = subset(st2 , vedur == 2 & teg==i)
vt24 = subset(st2 , vedur == 3 & teg==i)
vt25 = subset(st2 , vedur == 4 & teg==i)
vt26 = subset(st2 , vedur == 5 & teg==i)
vt27 = subset(st2 , vedur == 6 & teg==i)
vt28 = subset(st2 , vedur == 7 & teg==i)
vt29 = subset(st2 , vedur == 8 & teg==i)

## ----- ##
## Box plot - Depth##
## ----- ##

#Area 1
dt11 = subset(st1 , dypi > 0 & dypi <= 103.5 & teg==i)
```

```
dt12 = subset(st1 , dypi > 103.5 & dypi <= 131 & teg==i)
dt13 = subset(st1 , dypi > 131 & dypi <= 170 & teg==i)
dt14 = subset(st1 , dypi > 170 & dypi <= 203 & teg==i)
dt15 = subset(st1 , dypi > 203 & dypi <= 245.5 & teg==i)
dt16 = subset(st1 , dypi > 245.5 & dypi <= 500 & teg==i)
```

#Area 2

```
dt21 = subset(st2 , dypi > 0 & dypi <= 103.5 & teg==i)
dt22 = subset(st2 , dypi > 103.5 & dypi <= 131 & teg==i)
dt23 = subset(st2 , dypi > 131 & dypi <= 170 & teg==i)
dt24 = subset(st2 , dypi > 170 & dypi <= 203 & teg==i)
dt25 = subset(st2 , dypi > 203 & dypi <= 245.5 & teg==i)
dt26 = subset(st2 , dypi > 245.5 & dypi <= 500 & teg==i)
```

```
## ----- ##
## Box plot – Surface temperature##
## ----- ##
```

#Area 1

```
yt11 = subset(st1 , yfirdordshiti >=-1 & yfirdordshiti <= 3 & teg==i)
yt12 = subset(st1 , yfirdordshiti > 3 & yfirdordshiti <= 4.2 & teg==i)
yt13 = subset(st1 , yfirdordshiti > 4.2 & yfirdordshiti <= 5 & teg==i)
yt14 = subset(st1 , yfirdordshiti > 5 & yfirdordshiti <= 5.7 & teg==i)
yt15 = subset(st1 , yfirdordshiti > 5.7 & yfirdordshiti <= 6.4 & teg==i)
yt16 = subset(st1 , yfirdordshiti > 6.4 & yfirdordshiti <= 9 & teg==i)
```

#Area 2

```
yt21 = subset(st2 , yfirdordshiti >=-1 & yfirdordshiti <= 3 & teg==i)
yt22 = subset(st2 , yfirdordshiti > 3 & yfirdordshiti <= 4.2 & teg==i)
yt23 = subset(st2 , yfirdordshiti > 4.2 & yfirdordshiti <= 5 & teg==i)
yt24 = subset(st2 , yfirdordshiti > 5 & yfirdordshiti <= 5.7 & teg==i)
yt25 = subset(st2 , yfirdordshiti > 5.7 & yfirdordshiti <= 6.4 & teg==i)
yt26 = subset(st2 , yfirdordshiti > 6.4 & yfirdordshiti <= 9 & teg==i)
```

```
## ----- ##
## Box plot – Bottom temperature##
## ----- ##
```

#Area 1

A. R-script

```
bt11 = subset(st1, botnhiti > -3 & botnhiti <= 2.6 & teg==i)
bt12 = subset(st1, botnhiti > 2.6 & botnhiti <= 3.7 & teg==i)
bt13 = subset(st1, botnhiti > 3.7 & botnhiti <= 4.4 & teg==i)
bt14 = subset(st1, botnhiti > 4.4 & botnhiti <= 5 & teg==i)
bt15 = subset(st1, botnhiti > 5 & botnhiti <= 5.8 & teg==i)
bt16 = subset(st1, botnhiti > 5.8 & botnhiti <= 8 & teg==i)
```

#Area 2

```
bt21 = subset(st2, botnhiti > -3 & botnhiti <= 2.6 & teg==i)
bt22 = subset(st2, botnhiti > 2.6 & botnhiti <= 3.7 & teg==i)
bt23 = subset(st2, botnhiti > 3.7 & botnhiti <= 4.4 & teg==i)
bt24 = subset(st2, botnhiti > 4.4 & botnhiti <= 5 & teg==i)
bt25 = subset(st2, botnhiti > 5 & botnhiti <= 5.8 & teg==i)
bt26 = subset(st2, botnhiti > 5.8 & botnhiti <= 8 & teg==i)
```

```
## ----- ##
## Box plot – Sea conditions ##
## ----- ##
```

#Area 1

```
s11 = subset(st1, sjor == 0 & teg==i)
s12 = subset(st1, sjor == 1 & teg==i)
s13 = subset(st1, sjor == 2 & teg==i)
s14 = subset(st1, sjor == 3 & teg==i)
s15 = subset(st1, sjor == 4 & teg==i)
s16 = subset(st1, sjor == 5 & teg==i)
s17 = subset(st1, sjor == 6 & teg==i)
s18 = subset(st1, sjor == 7 & teg==i)
s19 = subset(st1, sjor == 8 & teg==i)
```

#Area 2

```
s21 = subset(st2, sjor == 0 & teg==i)
s22 = subset(st2, sjor == 1 & teg==i)
s23 = subset(st2, sjor == 2 & teg==i)
s24 = subset(st2, sjor == 3 & teg==i)
s25 = subset(st2, sjor == 4 & teg==i)
s26 = subset(st2, sjor == 5 & teg==i)
s27 = subset(st2, sjor == 6 & teg==i)
s28 = subset(st2, sjor == 7 & teg==i)
s29 = subset(st2, sjor == 8 & teg==i)
```



```
## ----- ##
## Box plot – Towing speed##
## ----- ##
```

#Area 1

```
th11 = subset(st1 , toghradi >= 2 & toghradi <= 3.6 & teg==i)
th12 = subset(st1 , toghradi > 3.6 & toghradi <= 3.7 & teg==i)
th13 = subset(st1 , toghradi > 3.8 & toghradi < 3.9 & teg==i)
th14 = subset(st1 , toghradi > 3.8 & toghradi <= 3.9 & teg==i)
th15 = subset(st1 , toghradi > 3.9 & toghradi <= 4 & teg==i)
th16 = subset(st1 , toghradi > 4 & toghradi <= 5.2 & teg==i)
```

#Area 2

```
th21 = subset(st2 , toghradi >= 2 & toghradi <= 3.6 & teg==i)
th22 = subset(st2 , toghradi > 3.6 & toghradi <= 3.7 & teg==i)
th23 = subset(st2 , toghradi > 3.8 & toghradi < 3.9 & teg==i)
th24 = subset(st2 , toghradi > 3.8 & toghradi <= 3.9 & teg==i)
th25 = subset(st2 , toghradi > 3.9 & toghradi <= 4 & teg==i)
th26 = subset(st2 , toghradi > 4 & toghradi <= 5.2 & teg==i)
```

```
## ----- ##
## Box plot – Vessel##
## ----- ##
```

#Area 1

```
v11 = subset(st1 , skip == 1131 & teg==i)
v12 = subset(st1 , skip == 1273 & teg==i)
v13 = subset(st1 , skip == 1274 & teg==i)
v14 = subset(st1 , skip == 1275 & teg==i)
v15 = subset(st1 , skip == 1277 & teg==i)
v16 = subset(st1 , skip == 1279 & teg==i)
v17 = subset(st1 , skip == 1280 & teg==i)
v18 = subset(st1 , skip == 1281 & teg==i)
v19 = subset(st1 , skip == 1325 & teg==i)
v110 = subset(st1 , skip == 1459 & teg==i)
v111 = subset(st1 , skip == 2350 & teg==i)
```

#Area 2

```
v21 = subset(st2 , skip == 1131 & teg==i)
```

A. R-script

```
v22 = subset(st2 , skip == 1273 & teg==i)
v23 = subset(st2 , skip == 1274 & teg==i)
v24 = subset(st2 , skip == 1275 & teg==i)
v25 = subset(st2 , skip == 1277 & teg==i)
v26 = subset(st2 , skip == 1279 & teg==i)
v27 = subset(st2 , skip == 1280 & teg==i)
v28 = subset(st2 , skip == 1281 & teg==i)
v29 = subset(st2 , skip == 1325 & teg==i)
v210 = subset(st2 , skip == 1459 & teg==i)
v211 = subset(st2 , skip == 2350 & teg==i)
```

```
## ----- ##
## Box plot - Towing length##
## ----- ##
```

#Area 1

```
tl11 = subset(st1 , toglengd >= 1.2 & toglengd <= 2.1 & teg==i)
tl12 = subset(st1 , toglengd > 2.1 & toglengd <= 3 & teg==i)
tl13 = subset(st1 , toglengd > 3 & toglengd <= 3.6 & teg==i)
tl14 = subset(st1 , toglengd > 3.6 & toglengd <= 3.9 & teg==i)
tl15 = subset(st1 , toglengd > 3.9 & toglengd <= 4 & teg==i)
tl16 = subset(st1 , toglengd > 4 & toglengd <= 5.6 & teg==i)
```

#Area 2

```
tl21 = subset(st2 , toglengd >= 1.2 & toglengd <= 2.1 & teg==i)
tl22 = subset(st2 , toglengd > 2.1 & toglengd <= 3 & teg==i)
tl23 = subset(st2 , toglengd > 3 & toglengd <= 3.6 & teg==i)
tl24 = subset(st2 , toglengd > 3.6 & toglengd <= 3.9 & teg==i)
tl25 = subset(st2 , toglengd > 3.9 & toglengd <= 4 & teg==i)
tl26 = subset(st2 , toglengd > 4 & toglengd <= 5.2 & teg==i)
```

```
## ----- ##
## Box plot - Towing time##
## ----- ##
```

#Area 1

```
tt11 = subset(st1 , togtimi > 0 & togtimi <= 57 & teg==i)
tt12 = subset(st1 , togtimi > 57 & togtimi <= 60 & teg==i)
```

```

tt13 = subset(st1 , togtimi > 60 & togtimi <= 62 & teg==i)
tt14 = subset(st1 , togtimi > 62 & togtimi <= 64 & teg==i)
tt15 = subset(st1 , togtimi > 64 & togtimi <= 66 & teg==i)
tt16 = subset(st1 , togtimi > 66 & togtimi <= 105 & teg==i)

```

#Area 2

```

tt21 = subset(st2 , togtimi > 0 & togtimi <= 57 & teg==i)
tt22 = subset(st2 , togtimi > 57 & togtimi <= 60 & teg==i)
tt23 = subset(st2 , togtimi > 60 & togtimi <= 62 & teg==i)
tt24 = subset(st2 , togtimi > 62 & togtimi <= 64 & teg==i)
tt25 = subset(st2 , togtimi > 64 & togtimi <= 66 & teg==i)
tt26 = subset(st2 , togtimi > 66 & togtimi <= 105 & teg==i)

```

```

## ----- ##
## Box plot - Latitude##
## ----- ##

```

#Area 1

```

la11 = subset(st1 , lat >= 62.80767 & lat <= 63.93083 & teg==i)
la12 = subset(st1 , lat > 63.93083 & lat <= 64.2085 & teg==i)
la13 = subset(st1 , lat > 64.2085 & lat <= 64.50392 & teg==i)
la14 = subset(st1 , lat > 64.50392 & lat <= 64.8875 & teg==i)
la15 = subset(st1 , lat > 64.8875 & lat <= 65.0505 & teg==i)
la16 = subset(st1 , lat > 65.0505 & lat <= 65.49358 & teg==i)

```

#Area 2

```

la21 = subset(st2 , lat >= 65.40317 & lat <= 65.89667 & teg==i)
la22 = subset(st2 , lat > 65.89667 & lat <= 66.15258 & teg==i)
la23 = subset(st2 , lat > 66.15258 & lat <= 66.4375 & teg==i)
la24 = subset(st2 , lat > 66.4375 & lat <= 66.70083 & teg==i)
la25 = subset(st2 , lat > 66.70083 & lat <= 66.9225 & teg==i)
la26 = subset(st2 , lat > 66.9225 & lat <= 67.15083 & teg==i)

```

```

## ----- ##
## Box plot - Longitude##
## ----- ##

```

A. R-script

#Area 1

```
lo11 = subset(st1,lon >=-27.24183 & lon <= -26.32383 & teg==i)
lo12 = subset(st1,lon >-26.32383 & lon <= -25.53583 & teg==i)
lo13 = subset(st1,lon >-25.53583 & lon <= -24.81742 & teg==i)
lo14 = subset(st1,lon >-24.81742 & lon <= -24.239 & teg==i)
lo15 = subset(st1,lon >-24.239 & lon <= -23.59983 & teg==i)
lo16 = subset(st1,lon >-23.59983 & lon <= -21.95558 & teg==i)
```

#Area 2

```
lo21 = subset(st2,lon >=-26.81467 & lon <= -25.71742 & teg==i)
lo22 = subset(st2,lon >-25.71742 & lon <= -24.90325 & teg==i)
lo23 = subset(st2,lon >-24.90325 & lon <= -24.10825 & teg==i)
lo24 = subset(st2,lon >-24.10825 & lon <= -23.28525 & teg==i)
lo25 = subset(st2,lon >-23.28525 & lon <= -22.16 & teg==i)
lo26 = subset(st2,lon >-22.16 & lon <= -20.99137 & teg==i)
```

```
## ----- ##
## Linear regression model##
## ----- ##
```

```
library(ggplot2)
library(psc1)
library(boot)
```

```
st3 = subset(st2, teg == 5)
```

```
surfacetemp = st3$yfirbordshiti
bottomtemp = st3$botnhiti
depth = st3$dypi
weather = st3$vedur
luminosity = st3$kl.hift
barometer = st3$loftvog
seacondition = st3$sjor
vindhradi = st3$vindhradi
vindatt = st3$vindatt
heildarafli = st3$aflí
Logheildarafli = log10(st3$aflí)
latitude = st3$lat
longitude = st3$lon
year = st3$ar
vessel = st3$skip
```

```

towingspeed = st3$toghradi
towingtime = st3$togtimi
towinglength = st3$toglengd
towingdirection = st3$togstefna

#Data frame with all factors used
zinb <- data.frame(surfacetemp, bottomtemp, depth, luminosity, seacondition, we

#Removes all NA values
zinb = zinb[complete.cases(zinb),]

summary(zinb)

#Histogram and normplot
ggplot(zinb, aes(zinb$Logheildarafli)) + geom_histogram() +
  ylab("Number_of_tows") +
  xlab("log_number_of_redfish_per_tow") +
  theme(text = element_text(size=20),
        axis.title.y=element_text(margin=margin(0,20,0,0)),
        axis.title.x=element_text(margin=margin(20,0,0,0)),
        axis.text.x = element_text(angle=0, vjust=0.5))

qqnorm(zinb$Logheildarafli, main = "",
       xlab = "", ylab = "",
       plot.it = TRUE, datax = FALSE, cex.axis = 1.2)
#Axis labels
mtext("Theoretical_Quantiles", side = 1, line = 4, cex = 1.7, font = 3)
mtext("Sample_Quantiles", side = 2, line = 2.9, cex = 1.7, font = 3)
mtext("Normal_Q-Q_Plot_-_Redfish", side = 3, line = 1, cex = 2, font = 3)

#Generates model
summary(p1 <- lm(zinb$Logheildarafli ~ poly(zinb$lattitude, zinb$longitude)
+ factor(zinb$seacondition) + factor(zinb$vessel) + zinb$towingspeed + zinb$towingtime,
  data = zinb))
af = anova(p1)
afss <- af$`Sum Sq`
a = print(cbind(af, PctExp=afss/(sum(afss))*100))

#One factor at a time
summary(p1 <- lm(zinb$Logheildarafli ~ poly(zinb$lattitude, zinb$longitude)
+ factor(zinb$vessel) + zinb$towingtime, data = zinb))

```

A. R-script

```
af = anova(p1)
afss <- af$`Sum Sq`
a = print(cbind(af,PctExp=afss/(sum(afss))*100))

residual_1 = resid(p1)
plot(zinb$phat,residual_1,ylab="Residuals",xlab="Prediction_of_log-trans

#Save as a text file
capture.output(a, file = "CodSummary.txt")

#Compare models
vuong(p1, m1)

dput(coef(m1, "count"))
dput(coef(m1, "zero"))
```

B. Additional figures and tables

The tables below show the analysis of variance when investigating cod and redfish in both area 1 and area 2 using model (6.2).

Table B.1: Analysis of variance for cod in area 1

| Source | Df | SS | MS | F-value | p-value | % expl. |
|----------------------------|------|-------|-------|---------|-----------|---------|
| poly(latitude,longitude,4) | 14 | 50.2 | 3.6 | 57.6 | 0.000 *** | 24 |
| +poly(depth,2) | 2 | 1.2 | 0.59 | 9.4 | 0.000 *** | 0.56 |
| +poly(surface temp,2) | 2 | 0.76 | 0.38 | 6.2 | 0.002 ** | 0.36 |
| +factor(vessel) | 10 | 11.47 | 1.15 | 18.4 | 0.000 *** | 5.5 |
| +towing time | 1 | 8.2 | 8.2 | 131 | 0.000 *** | 4 |
| Total model | 29 | 71.7 | | | | 34.34 |
| Total residuals | 2204 | 137.2 | 0.062 | | | 65.66 |
| Total | 2233 | 209 | | | | |

Table B.2: Analysis of variance for cod in area 2

| Source | Df | SS | MS | F-value | p-value | % expl. |
|----------------------------|------|-------|------|---------|-----------|---------|
| poly(latitude,longitude,4) | 14 | 78.4 | 5.6 | 36.6 | 0.000 *** | 10.38 |
| +poly(depth,2) | 2 | 6.13 | 3.06 | 20 | 0.000 *** | 0.81 |
| +poly(surface temp,2) | 2 | 27.6 | 13.8 | 90.2 | 0.000 *** | 3.65 |
| +factor(vessel) | 9 | 17.5 | 1.9 | 12.7 | 0.000 *** | 2.32 |
| +towing time | 1 | 18.7 | 18.7 | 122 | 0.000 *** | 2.47 |
| Total model | 28 | 148.4 | | | | 19.6 |
| Total residuals | 3972 | 607.4 | 0.15 | | | 80.4 |
| Total | 4000 | 755.8 | | | | |

B. Additional figures and tables

Table B.3: Analysis of variance for redfish in area 1

| Source | Df | SS | MS | F-value | p-value | % expl. |
|----------------------------|------|-------|------|---------|-----------|---------|
| poly(latitude,longitude,4) | 14 | 94.7 | 6.8 | 24.7 | 0.000 *** | 11.23 |
| +poly(depth,2) | 2 | 55.8 | 27.9 | 102.2 | 0.000 *** | 6.62 |
| +poly(surface temp,2) | 2 | 18 | 9 | 33 | 0.000 *** | 2.14 |
| +factor(vessel) | 10 | 36.7 | 3.68 | 13.5 | 0.000 *** | 4.4 |
| +towing time | 1 | 29.2 | 29.2 | 107 | 0.000 *** | 3.5 |
| Total model | 29 | 234.6 | | | | 28 |
| Total residuals | 2229 | 608.5 | 0.27 | | | 72.2 |
| Total | 2258 | 843.1 | | | | |

Table B.4: Analysis of variance for redfish in area 2

| Source | Df | SS | MS | F-value | p-value | % expl. |
|----------------------------|------|--------|------|---------|-----------|---------|
| poly(latitude,longitude,4) | 14 | 228.6 | 16.3 | 71.1 | 0.000 *** | 18.2 |
| +poly(depth,2) | 2 | 12.7 | 6.4 | 28 | 0.000 *** | 1.01 |
| +poly(surface temp,2) | 2 | 45.2 | 22.6 | 98.5 | 0.000 *** | 3.6 |
| +factor(vessel) | 9 | 57 | 6.33 | 27.6 | 0.000 *** | 4.55 |
| +towing time | 1 | 31.7 | 31.7 | 138 | 0.000 *** | 2.52 |
| Total model | 28 | 375.4 | | | | 29.9 |
| Total residuals | 3835 | 880 | 0.23 | | | 70.1 |
| Total | 3863 | 1255.4 | | | | |