Accounting for potential nonlinearity between catch and effort using meta-analysis and applying GLM and GLMM to fishing data from deployments of fixed and mobile gear

by

Michelle Aljafary

A thesis submitted to the Faculty of Graduate Studies of

The University of Manitoba

in partial fulfilment of the requirements of the degree of

MASTER OF SCIENCE

Department of Biological Sciences

University of Manitoba

Winnipeg

ABSTRACT

My thesis examines nonlinearity between catch and effort. I use a meta-analysis of published literature and generalized linear mixed-effects models (GLMM) on both fixed and mobile gear fisheries of Atlantic Canada. The meta-analysis examines the proportionality of catch to effort using the slope of the reduced major axis (RMA) log-log regression, which accounts for "errors-in-variables". The GLMMs explored proportionality while accounting for variation among fishing vessels. Both analyses found evidence for disproportionality between catch and effort. Catch that increases disproportionally to effort could result from either facilitation or recruitment of effort into the fishery. Catch increases that are less than proportional are expected from competitive interactions among fishers or gear saturation. The GLMM also revealed that the level of aggregation (by set, trip, monthly, or annually) can affect the apparent proportionality between catch and effort. In general, catch and effort should not be considered to be proportional.

Key words: meta-analysis, generalized linear mixed-effects model (GLMM), catch, effort, error-in-variables, proportionality, disproportionality, mixed effects, RMA, OLS, aggregation.

ACKNOWLEDGEMENTS

I would like to thank the Saudi Cultural Bureau for providing the financial support for this research and the NSERC grant RPGIN-2014-04135 to Darren Gillis. Many thanks go to my professor Darren Gillis for his help and support during my research. I would also like to thank Gail Davoren and Saman Muthukumarana for their suggestions regarding this manuscript. Thanks to Mike McMahon and Peter Comeau at the Bedford Institution of Oceanography (BIO, Dartmouth, Nova Scotia) for providing the trawl fishery of southwest Nova Scotia research data (mobile gear). Thanks to Elmer Wade, who provided the Gulf of St. Lawrence snow crab fleet trap fishery data (fixed gear) for this research. Thanks to my lab mates Samantha Fulton and Kate Christie for their support and suggestions on my presentation.

DEDICATION

To my husband, Falaah Alshammari

You have lit up my life ever since you became a part of me. The days we've spent together, enjoyed and suffered. Your love, support, care, and faith are a special gift to me.

To my dad, Abdulrhman Aljafary, and mom, Mariam Aljafary

My life is fulfilled through your love. You never stop giving all you can and have. Thank you for your waiting patiently for me to come back home after all these years. Your love and prayers do not go to waste.

To my sisters, Marwah, Maali Aljafary, and my brothers, Mohammed, Mshari and Majed Aljafary

Neither space nor time can erase the great memories that we've shared since we were children. I count the days and the nights until we meet again, as we used to. The thought of seeing you again brings me great joy and happiness".

To my closest friend, Enas Almula

Saying farewell was very hard on you and me, but here we are, soon to be together again.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter One

Introduction

Introduction

Before 1980, most of the studies focused on fish biology, very little work was done regarding fleet dynamics and fishermen's behaviour. Since 1980, however, there has been a revival of interest in these topics. Currently, most such studies employ statistical analysis with advanced statistical methods. The dynamic behaviour of fishing fleets is the most important link between understanding fish biology and catch rates. Studying the behaviour of fishermen is the key to understanding why one catch is greater than another (Hilborn 1985). Despite years of research, there is still much debate over the validity and reliability of the use of commercial catch and effort data to estimate fish abundance. Currently, many, if not most, fish population analyses are based on a metric known as catch-per-unit-effort (CPUE). The effort is a measure of fishing activity, and nominal effort is the way it is represented (number of vessels, hours fished, etc.). It is one of several data sources contributing to research among the following: virtual population analysis (VPA), statistical catch-at-age (SCA), and, more recently, integrated analyses such as stock synthesis. There are a number of variables for fishing effort, such as the number of vessels or the number of anglers. To measure effort, the number of vessels is multiplied by time, whether the time is in days or hours of fishing. All this information can be acquired from the available fisheries data. If the available information includes CPUE and catch number, then we can calculate effort by dividing catch number by CPUE. Having CPUE can be helpful for acquiring any incomplete data from either catch or effort (Ricker 1975).

CPUE is most often used as an index of abundance, though abundance is not the only factor that affects CPUE. For example, CPUE often varies among years and areas fished due changing distributions and abundance. In addition, the size of vessels and number of active vessels in response to market forces can also influence the relationship between catch and effort.

Most previous work has used standardization of CPUE to remove the impact of such factors in order to obtain independent data from the original data (e.g. Kimura 1981, Vignaux 1996, Maunder 2001, Campbell 2004). Most fishery studies that assess fish stocks (VPA, and integrated analyses) have used methods that include fitting models to an index of abundance, usually CPUE. A number of studies have used statistical methods and concluded that CPUE is inaccurate and may be misleading in providing indices of abundance (Saville and Oe 1980, Gillis and Peterman 1998, Salthaug and Godø 2000). Harley et al. (2001) used meta-analysis to reveal that there is a common disproportionality between CPUE and abundance. They represented this relationship as a power curve, based on the value of a shape parameter, the exponent β. Despite the fact that CPUE has been recognized as being an inaccurate metric for reflecting abundance, it continues to be used for this very purpose, to represent abundance (Harley et al. 2001, Maunder et al. 2006).

A number of studies have standardized catch and effort data for different reasons, such as to remove the impact of factors such as fishing location, target species, and environmental conditions, because each of these factors is measured differently (Su et al. 2008). Standardization can be performed in various ways, for example via a log-linear model or using advanced methods such as a generalized linear model. When using the log transformation, logging both sides of the equation will convert the multiplication to addition. This transformation makes it easier to deal with the variables and helps to remove as much variability as possible from the factors. It is often assumed that CPUE is proportional to abundance, however, using CPUE = C/f (C = catch, f = nominal effort) also assumes that catch is proportional to effort (Gillis 2003). If this is in fact untrue, then CPUE may not accurately reflect abundance.

Many studies have used catch and effort data for a variety of reasons. Catch and effort data have been used to investigate factors that affect catch, such as the choice of fishing location, the target species, and environmental conditions (Su et al. 2008). Catch and effort data have also been examined to determine the fishing tactics employed. Fishing tactics are the decisions made before each fishing operation that determine such things as fishing location, gear, and target species. Fishing location and seasonal effects are some of the important factors that influence the choice of fishing tactics (Pelletier and Ferraris 2000). The chosen search strategy is helpful in interpreting catch rate and relates to the area fished and the time spent fishing; an example is the San Diego tuna fleet (Orbach 1977). There are two strategies used: “hunter” and “chaser”. Hunters are fishermen that rely on themselves, are more knowledgeable, and have more experience. The chasers, on the other hand, are typically greater in number and used radio information to follow the hunters, thus moving into aggregations of boats. However, within each group skippers have different amounts of relevant knowledge (Orbach 1977).

Fishing tactics and strategies have been used to better understand the catch rate. The fishing tactics employed by a skipper at sea during a fishing trip can be considered as having two components: the first component is the métier, which is the choice of target species, in area fished, and gear used, selected for the fishing trip by the skipper, and the second relates to the indices that show how the skipper behaved during the fishing trip, such as aggregation or movement within the area. These fishing tactics and strategies have a high impact on catch rates (Marchal et al. 2006).

My thesis will examine whether the amount of fish caught is proportional to nominal effort. The model that has been used in current fisheries is shown in the equation below (Maunder and Punt 2004):

$$C_i = \beta_0 \cdot f_i \cdot N_i \tag{1.1}$$

Where C_i is the catch, β_0 is a constant (catchability), f is the fishing effort, and N is the abundance, which may be represented as density.

There are many reasons for disproportionality between catch and effort: the changing behaviour of fishing fleets through competition, facilitation among vessels, or a numerical response to fish availability, among others. Competition is defined as individual usage or defense of a resource to reduce the availability of that resource to any other individuals. Competition is one of the main reasons for why some natural foragers dominate others and among fishing vessels it may cause variation in success among vessels.

Competition results when different forgers target the same species, either at different times of the day, on different days, or even during different seasons, such that there is no real interaction between them but the species may be reduced by one of them, which, in turn, affects the other. This indirect competition is known as exploitation competition. In the late 1800's, lake sturgeon (*Acipenser fulvescens*) was overexploited as fishermen targeted it as a preferred species. By the early 1900's, this species' numbers had been greatly reduced. Moreover, the heavy exploitation had a similar negative effect on lake herring (*Leucichthys artedi*) in Lake Erie. Then, in the 1920's and 1930's, it was whitefish (*Coregonus clupeaformis*) in Lake Huron that were overexploited. All these overexploited species encountered a sharp decline in their catch (Smith 1968). Similar declines in catch and abundance can be observed within seasons and fishing areas on a shorter time scale.

For the target species with low abundance, the relationship between catch and nominal effort can be non-intuitive. For example, with decreasing abundance it may take a trawl or a trap longer to catch the same amount of fish. If they attempt to fish until their gear is saturated nominal effort, in the term of hours, will increase while the catch remain the same and overall catch-per-unit effort will decline.

Another form of competition is called gear competition (Rothschild 1967), or competition among fish for space in the gear. In this case, longer fishing times catch proportionately less fish because fish already in the nets, traps, or on the hooks limit the capture of additional fish. This will occur more rapidly when fish populations are high or locally aggregated, but can be mitigated by retrieving gear over shorter time intervals. It can also occur when gear becomes saturated by unwanted species, limiting the capture of target species (Dauk and Schwarz 2001).

Interference competition is a direct interaction among competitors. It occurs when one forager physically interferes with another while that one is trying to access a resource (Stillman et al. 1997, Delong and Vasseur 2013). When the number of vessels increases, the additional interactions may result in a reduction in catch due to factors such the gear (nest, hooks, traps, etc.) of one fisherman interfering with that of others. This is independent from fish population abundance and can rapidly reverse when vessel number decline. For example, setting new nets near one that is already in operation can scare the fish away and decrease the catch rate from the first net. Also, when there are a number of vessels fishing schooling fish, they disperse the school, reducing fishing success. When vessels are targeting the same species, their search operations can overlap. The evidence of interference was from the sole (*Solea solea)* and plaice (*Pleuronectes platessa)* when targeted by the Dutch beam trawl fleet. Immediate decrease in catch estimated to be14% when vessels increased in the open north area even when doubling fishing effort (Poos and Rijnsdorp 2007). The catch rate increased when vessel density declined by 10% due to the reduction in interactions (less competition) (Rijnsdorp et al. 2000). An experiment done with chinook salmon found that, when vessel density increases, the catch rate of chinook salmon decreases. Vessel densities were randomized in the experiment, so these changes represent competitive effects rather than local depletion of fish. Thus, variation in vessel density may exert a substantial influence on catch rates during the commercial fishing season (Abrahams and Healey 1993).

The Ideal free distribution (IFD) in fleet dynamics is a behavioural hypothesis that examines the spatial relationship between foragers and their resources. It starts with the assumption that, in an ideal world, all foragers would have the same knowledge of their environment and there is no cost or restriction in the movement between targeted areas. In addition to these assumptions, the model assumes that foragers have the same ability in foraging and in accessing resources and that they reduce each other's success through competitive interactions (Gillis and Peterman 1998). When the assumptions are reasonable, based on IFD, the benefits will be distributed equally between foragers, and the proportion of foragers at any site will equal the proportion of resources available. When any one of these assumptions is violated, however, this will affect IFD and changes in the prediction will occur. These violations will provide a new way to look at predicted distributions. If foragers lack knowledge about the environment, then they will need to search among various foraging sites to acquire new information to improve their success. If they have different abilities and costs they will not follow an IFD. If some foragers are more successful, have knowledge about their area and can defend it, then this will produce a new ideal dominance distribution through asymmetrical interference (Gillis and Peterman 1998, Abrahams and Healey 1990, Gillis et al. 1993, Gillis 2003, Swain et al. 2003).

Regulations such as limited fishing seasons, gear restrictions, and others may be another reason for catch decline. These regulations increase the competition for a limited available catch. Fishermen will see this interference as a form of restrictive regulations rather than as a form of competition. Because of the restrictive regulations, if anything goes wrong for the fishermen, such as a wrong choice of area or if gear stops working, this will ultimately reduce the catch because of the limited fishing season.

Another of the many possible reasons for a low catch is traveling time and searching for an area with aggregation of the target species. Some fishermen fish in new areas to get to know them and acquire information. When they fish in an area with low numbers of fish, even if they increase their fishing effort, catch will not increase accordingly. Fishermen in the salmon gillnet fishery, for example, go fishing early in the season in order to test their gear and get it in good working order so they don't lose their catch later on due to malfunction. Moreover, some fishermen fish for the purpose of feeding themselves or their families, even when fish are less abundant, and, once they have caught enough to last for the rest of the year, they stop fishing. Also, when fishermen do not want to increase their costs, they fish in an area close to home, which may be less abundant; here, again, catch will not increase in direct correlation to effort. Areas more desirable and less risky may also have a lower catch rate, even if fishermen increase their efforts there (Hilborn and Walters 1992).

There are different types of response to changes in prey density. One of them is the numerical response, such as aggregative responses where increases in local prey numbers lead to predators moving into an area. Numerical responses occur in the case of British Columbia's salmon gillnet fleet. The movement of boats there was in response to fish abundance (Millington 1984). The numerical response accrues when fishermen devote effort in response to quantitative measures of fishing success, such as fish density. The numerical response can be stronger when the time necessary for fishermen to get to the prey is shorter. For example, rainbow trout lakes attract 2.5 times more fishermen than other lakes, because rainbow trout lakes are close to the center of British Columbia, making travel shorter, which, in turn, attracts higher numbers of fishermen to travel from the greater Vancouver area (Post et al. 2002). Fishing location affects the distribution of fishermen. The more success had in an area, the more predators will be attracted to that area; this is a positive numerical response.

From the analysis of models on the predation process, it has been found that the distribution of foragers' search time and catches are dependent on the distribution of prey. In addition, a change in abundance is not reflected in catch per fishing day, except when the school density is proportional to abundance (Paloheimo and Dickie 1964). The difficulty lies in determining the reasons for the increase in catch, and therefore dollar returns, resulting from increased effort. There are two possible reasons for this. The first one is that gross dollar returns of individual vessels increase when vessels fish for more days because the increased experience improves their efficiency: a learning response. The second reason is that gross dollar returns per vessel increase when more vessels fish in the same area without interfering with each other, which can result from a facilitation response where vessel share information or assist in fish capture. The gross dollar returns from an individual vessel may not increase when the number of vessels fishing in the same area increases. This may be because of competition for preferred trawling sites (Lapointe 1989).

Catch variability may occur because of long-term shifts in the distributions of fish, or because of changes in fleet composition due to having found new, alternative fishing opportunities. For example, in the case of the English sole fishery, some of the changes in catch were because of the changes in cod abundance. When English sole catchability and, thus gross dollar returns decreased during the early 1960's, it was because the abundance of English sole had decreased, while cod abundance had increased (Lapointe 1989). When the search targets two different species, catch rates may vary as a result of switching to alternative prey. For example, if foraging time is T, then when $T = T_1 + T_2$, where T_1 is time spent foraging for species one and T_2 is time spent foraging for species two, then when foragers spend more time searching for the first species, the time foraging for the second species will decrease. This will reduce the catch rate of the second species in the fishery statistics, which do not identify the species toward which effort is directed (Lapointe 1989).

Meta-analysis in Ecology

Originally, reviewing research followed two main methodologies. The first is called “narrative review” and attempts to synthesize the results of a large body of research into a single consistent narrative; the reviews handled by this method do not include details of the source or the search terms used to obtain studies (Roberts et al. 2006). The second methodology is called vote counting and refers to a process where the significant results from various studies are counted (Koricheva et al. 2013). In order to avoid the bias potentially afflicting these two methods, the statistical techniques of meta-analysis have been developed (Gates 2002).

The term “meta-analysis” was coined in 1976 by Gene V. Glass. Meta-analysis is one of the methods of research synthesis, which can be conducted qualitatively or quantitatively: qualitatively in the form of narrative review and quantitatively by using statistical methods (Koricheva et al. 2013). Meta-analysis uses techniques that collect analytical results and integrate the findings to get a wider understanding of a specific question. Combining the findings from a number of studies strengthens the evidence for a specific question or hypothesis. This method was first used in medicine and social sciences. Meta-analysis in ecology began to be performed early in 1990s (Koricheva et al. 2013) on questions such as the value of life history variable and the presence of competition (Jarvinen 1991, Gurevitch et al. 1992). It has grown in use, investigating topics such as value of endangered species (Richardson and Loomis 2009) and patterns in habitat loss and species decline (Bender et al. 1998).

Meta-analysis uses tools that are powerful and unbiased, which helps in summarizing the results of studies for the same question or in the same area. The most helpful aspect of meta-analysis is that it represents the outcomes of each study on a common scale. This scale is what is called “effect size”; this includes information about the effects under investigation from each study (Koricheva et al. 2013). The process of meta-analysis begins with a precise question being asked on a specific topic. One develops criteria for a process of selection, starting with the initial search. In case of too broad a question, the question and the criteria should be refined in order to proceed with the full research. If sufficient data are available, a proper meta-analysis can be conducted. In the case of insufficient data, one can explain the gap in the area of interest. To conduct a meta-analysis, one must choose the effect size and the moderator, design the data file, extract the data, choose the model, choose the software, perform the analysis, check the bias, calculate the effect sizes and confidence intervals for each study, calculate the mean effect size and the confidence interval, test for heterogeneity, and, finally, interpret the results (Koricheva et al. 2013). It is important to have clear criteria for selection of studies, whether to include or exclude studies that seem untrustworthy or studies with samples that do not represent the target population (Gates 2002).

A component of research synthesis is the systematic review method, which one conducts on a specific topic or question of interest. The systematic review can be done by describing the steps used to select studies and by explicitly identifying the criteria used. This is helpful to others who are interested in similar studies. The systematic review can be done with or without meta-analysis, depending on whether or not there are sufficient data to conduct the meta-analysis. The systematic review alone shows the current studies, available knowledge, and gaps in the reviewed area (Koricheva et al. 2013).

One important step in meta-analysis is the collection of the effect size from each study. The effect size in one study is independent from that of the others. The best way of estimating the true effect size is to identify and select independent studies from which one can trust the findings. As all the studies chosen will contribute to the overall estimate of the true effect, a study with precise findings should be weighted more heavily than others; this will increase its influence on the overall estimate. Meta-analysis also has some bias issues when it comes to the selection of studies. When poor methodology is used in one or more of studies used, meta-analysis can produce misleading, inaccurate results.

The difference in the effect size between studies can be because of various factors: biological factors, such as when selecting studies that are based on different areas, or statistical factors, such as when selecting studies that use different methods to acquire their findings. It is best to identify these factors when doing a meta-analysis (Gates 2002). There are two kinds of effect size: fixed effects and random effects. In the case of fixed effects, the effect size in the population will be fixed but with unknown constants, so the true effect size is assumed to be the same for all studies in the meta-analysis. This is called a homogenous case. However, with random effects, the population effect size may differ randomly from study to study, so each study has a different effect size. Statistically, the difference between calculating the fixed effect and the random effect is found in the standard error, related to the effect size. In the case of fixed-effects, the standard error will be calculated and the variability estimated within the study. Fixed effects ignore any other factors that explain the variability. On the other hand, the standard error in the random-effects model will be calculated both within a given study and between studies. The random effect accounts for errors from sampling a population. These errors arise from the variability estimated within a study and the variability estimated between studies. The standard error that results from random effects will be larger than in the case of fixed effects (Raudenbush 1994, Field 2001). The random variation acknowledged to exist among studied effects is due to additional, unknown factors. This can be accounted for by allowing this variation and assuming that the difference between true effect sizes for different studies is because of random variation around the overall mean effect. This mean effect represents the population of studies. When there is no variation between studies, the variance will be equal to 0, and if the true effect size is equal to the overall mean of the true effect size, then the random-effects model will be converted to a fixed-effects model (Koricheva et al. 2013).

To find out whether the studies selected for a meta-analysis are consistent or inconsistent, the test of heterogeneity should be performed. This detects whether there is any difference in the results of the studies. The test of heterogeneity assesses the hypothesis that all studies have the same effect size. This is computed via the Cochran's Q test, which produces a summary of the squared deviations of all the effect sizes in the studies to be examined in the meta-analysis. Meta-analysis with a small number of studies may not detect the heterogeneity. In case of an insignificant result, this cannot be considered evidence of homogeneity; it is more accurate when there are more studies and, more importantly, when each study has a large sample size. There is an alternative approach that has been developed that quantifies the effects of heterogeneity. This approach provides a measure for the inconsistency in the studies' results. This approach is represented in I^2, which is the percentage of total variation across studies due to heterogeneity. This can be calculated by $I^2 = 100\% \times (Q - df) / Q$, where Q is Cochran's and df represents the degrees of freedom. I^2 can be between 0% and 100%, where 0% represents no heterogeneity, and this percentage increases as the heterogeneity increases (Higgins et al. 2003). Heterogeneity has a major impact on a meta-analysis. The results of a heterogeneity test are derived from mathematical criteria: I^2 is the percentage of total variation across studies due to heterogeneity. Alternatively, H can be provided as the proportion of total variation in study estimates that is due to heterogeneity. Either H or I^2 is important to calculate when publishing the results of a meta-analysis because this helps show the impact of heterogeneity (Higgins and Thompson 2002).

When using two different measures of the true effect, it is important to check one against the other to find out how they vary and to detect any biases that may exist. Bivariate data are commonly analyzed using linear regression models. Model I regression, which is also called ordinary least square regression (OLS), is used to predict or estimate the relationships among variables. In this approach, estimating the parameter, i.e. the coefficient β in the regression line, can be done by minimizing the sum square of the distance between the observed value and the fitted line (Imbrie 1956). In OLS, it is assumed that X is fixed and Y varies, assuming that X can be measured without error (Smith 2009). Usually, however, both X and Y varies and is subject to error. However, when X is under the control of the experimenter it can be treated as fixed, even though there is uncertainty about its true value due to factors such as measurement error (Berkson 1950). It is also assumed that Y is normally distributed (Quinn and Keough 2002). Model II regression theory is more useful in case of errors in the independent variables. Nowadays, it is widely recognized that OLS is not the best fit in all situations. For example, when using Model I regression with the existence of errors in the independent variables, the value of the parameter, the coefficient β in the regression line, will be lower than otherwise expected. Reduced major axis regression (RMA), also known as geometric mean regression

(GM), has been recognized to be an appropriate alternative (LaBarbera 1989, Smith 2009, Quinn and Keough 2002). The GM Model II method is the preferred way of estimating the relationship between X and Y (Ricker 1973). The calculation of the slope coefficient β is simply the division of the standard deviation of Y by the standard deviation of X. This is equivalent to minimizing the sum of the triangle area constructed by fitting the horizontal and vertical lines dropping from the observed value to the fitted line (Quinn and Keough 2002). The standard error of the slope for RMA is the same as the standard error of the OLS model (Imbrie 1956, Laws and Archie 1981, Quinn and Keough 2002).

GLM and GLMM in fisheries

Generalized linear model is introduced in ecology in 1990 to deal with data exposed to errors (Aebischer and Coulson 1990, Austin et al. 1990) and in twenty first century GLM played an increasing role in the analysis of fisheries data (Maunder and Punt 2004). Goñi et al. (1999) used a generalized linear model to analyze hake catch rates to support the indices of abundance. These data were provided by a Spanish trawl fishery between 1991 and 1996. They tried to determine fishing power (the catchability associated with the characteristic of a vessel) using vessel tonnage. They also included year and month in the model used. They found that vessel tonnage and year have a significant effect on catch rate. The total variation in the data, which consisted of 54%, was explained by vessel size. Hake abundance fluctuated across the years. Ye et al. (2001) used a generalized linear model on data from a Kuwait driftnet fishery. They used both gamma and Bernoulli distributions for non-zero catches. They found that the decrease of the catch rate of silver pomfret *(Pampus argenteus)* changed fishermen's fishing behaviour toward hilsa shad *(Tenualosa ilisha)*. Dick (2004) discussed the distribution errors in the mean response.

He also used a generalized linear model. He evaluated different distributions to fit the data using the Akaike information criterion (AIC), these being lognormal, gamma, Weibull, log-logistic, and inverse Gaussian. He found that the AIC is an effective technique for discriminating among potential error distributions using moderate and large sample sizes. The AIC favored log-normal over other distributions. GLMs have also been used to examine fisheries survey methodology. The NAFC's (Northwest Atlantic Fisheries Centre) standard survey bottom trawl was used in an experiment conducted in 2001 to estimate trawl efficiency to capture snow crab *(Chionoecetes opilio)*. A second trawl was placed underneath the main one to capture any escaped crab. The results showed that the model had over-estimated true efficiency (Dawe et al. 2010).

A generalized linear mixed model (GLMM) is used in ecology to deal with non-normal data in the presence of random effect, which is often used to account for variation among individuals (Milsom et al. 2000, Pawitan et al. 2004, Vergara et al. 2007). A generalized linear mixed model (GLMM) has been used in a number of studies. Helser (2004) used a GLMM to examine fishing power among chartered industry-based vessels and on a research trawler, the *FRV Miller Freeman*, for bottom trawl surveys on the upper continental slope of the U.S. West coast. A GLMM can be helpful when using the year as a fixed effect and treating vessels as random effects. Two distributions were used: discrete distribution for a non-zero haul and continuous distribution for non-zero catch rates; both applied to four ground fish species. The best model was chosen based on which produced the smallest AIC value, which was the effect of random vessels and year. Using fixed effects is not appropriate; however, using vessels as a random effect is reasonable. Bishop et al. (2004) compared the results of various models, including a generalized linear mixed model, with the standardization of fishing effort and the relative index of abundance. He used different vessels' covariates in the models. The use of random effects did not make a difference in the results. The standard error was different from one model to another because of the difference in statistical efficiency. A GLMM can deal with spatial and temporal correlations. Baum and Blanchard (2010) used generalized linear mixed-effects models to estimate trends in the relative abundance of Northwest Atlantic oceanic and large coastal sharks between 1992 and 2005. The catch rate of eight kinds of sharks was standardized. There was a decline in the number of shark types, stabilization in the abundance of mako and thresher sharks, and an increase in the tiger shark population. Overall, the GLM and GLMM approaches are becoming standard methods in fisheries analysis in the way that linear regression and associated methods were used in the 20th century.

Fisheries Studied

Snow crab trap fishery

Snow crab (*Chionoecetes opilio*), also known as queen crab, is distributed in the North Atlantic and North Pacific Oceans. It is the most valuable species in the Atlantic region. Snow crabs live in both sandy and muddy, cold-water environments. The size of mature male snow crab is greater than that of the female (Paul 1992), with males in the Gulf fishery growing up to 16.5 cm CW (carapace width) while females are generally less than 9.5 cm in width (pers. Comm. Elmer Wade, Gulf Fisheries Centre, Fisheries and Oceans Canada). Fisheries are restricted to landing mature males with a width of 9.5 cm or more. Landing female snow crabs is forbidden in the fishery, in order to insure reproduction. Snow crabs are fished using various kinds of traps, such as conical, pyramidal and rectangular. There are a number of restrictions applied to snow crab fisheries, such as the number of licenses and traps allowed to a single fishery, seasonal limits, designated areas, and vessel size permitted in relation to the water's depth. Landed snow crabs are preserved in ice or in a circulation of salted water to ensure good quality in the long run. High numbers of snow crabs are harvested from Newfoundland and the Gulf of St. Lawrence. Snow crabs can be harvested from the spring to the fall (Weston 2011).

The Scotian Shelf 4X trawl fishery

The trawl fishery of division 4X captures several ground fish species including haddock (*Melanogrammus aeglefinus*), redfish (*Sebastes mentella*), pollock (*Pollachius virens L.*), cod (*Gadus morhua*) and halibut (*Reinhardtius hippoglossoides*). Regulations were introduced to the fishery in late 1960 to reduce overharvesting and catch decline in haddock. These regulate fishing seasons, spawning areas, fish size, fishing gear, species and catch. The fishery was closed when catch peaked between March and May. However, the closure was based on spawning and not on the catch. This closure was to avoid the disruption of fish by gear during breeding. In this closure all gears other than pelagic fishing gears that cannot reach the bottom of the sea, such as purse seines and midwater trawls, was prohibited. By-catch of haddock and cod during the closure was limited to 1% of total catch (Halliday 1988). Today, haddock and redfish are the major species taken by trawl in division 4X (pers. comm. Peter Comeau, Bedford institute of Oceanography, Dartmouth, NS).

Gear types

The choice of gear to be used in a fishery is an important decision. The gear and its method of deployment define the nominal effort that is integral to traditional calculations of catch per unit effort (CPUE) and the expected relationship between fishing activities and catch. Gear is selected depending on the target species and the characteristics of area to be fished. The choice of gear may depend on the depth of the water, mesh size, and the speed of the vessels to be used (Squires 1987). There are fixed and mobile types of fishing gear. Fixed gear remains stationary in one location over time and includes such things as traps, pounds, gill nets, long lines, seiners, trammel nets, small ring nets, and drifters (Stergiou and Erzini 2002). Snow crabs can be captured by traps, of which there are varying kinds used by fisheries. The conical trap with a larger mesh size is favoured by fishermen for its efficiency in catching more crabs within the legal size. Traps are made out of steel in order to be able to capture crabs from the very bottom of the water. The depth at which the traps are used can vary from 50 to 280 m. Harvested crabs can be stored in ice or salted water (Rose 1992, Chiasson et al. 1993, Xu and Millar 1993, Winger and Walsh 2011). Mobile gear is towed and dragged behind a boat and includes such things as trawls and dredges. A number of studies have discussed the impact of using trawls and dredges on the seabed (Messieh et al. 1991, Jones 1992, Watling and Norse 1998). Trawling by dragging nets over the seabed is a very common method of catching fish and is used by vessels with a length of between 10 and 130 m. Trawls often catch not only the target species, but also anything else in their way. Some trawls are dragged along the seabed when the target species is in deep water, such as in the case of Atlantic cod and shrimp (Watling and Norse 1998). There are various kinds of trawls, such as the otter trawl and beam trawl. The otter trawl has a cone shape and is held by two otter boards to keep the trawl open. These boards weigh thousands of kilograms. It is constructed with a number of panels that help to prevent species from escaping the net. The opening end of the otter trawl also consists of a ground rope, which is supported by rubber discs, bobbins, spacers, etc. The ground gear captures as many ground species as possible and prevents the trawl from being damaged while dragging it along the seabed. Otter trawls are made of wood, aluminum, or both. The important species most often fished for using otter trawls are cod, yellowtail and other flounders, haddock, redfish, and pollock. To begin fishing, the vessel simply releases the net into the water and drags it for several hours. The net is then pulled out of the water and the fish are sorted and packed in ice. The beam trawl has a steel beam instead of the two otter boards to keep the trawl mouth open (Squires 1987, Watling and Norse 1998).

My thesis will examine proportionality between catch and effort data by exploring the slope of the log-log regression using meta-analysis and the slope of the exponent of effort using generalized linear mixed-effects model in the Gulf of St. Lawrence snow crab data and 4X Scotian Shelf trawl fishery data to account for variation among fishing vessels.

References

Abrahams, M. V., & Healey, M. C. (1990). Variation in the competitive abilities of fishermen and its influence on the spatial distribution of the British Columbia salmon troll fleet. *Canadian Journal of Fisheries and Aquatic Sciences*, *47*(6), 1116-1121.

Abrahams, M. V., & Healey, M. C. (1993). Some consequences of variation in vessel density: a manipulative field experiment. *Fisheries Research*, *15*(4), 315-322.

Aebischer, N. J., & Coulson, J. C. (1990). Survival of the kittiwake in relation to sex, year, breeding experience and position in the colony. *The Journal of Animal Ecology*, 1063-1071.

Austin, M. P., Nicholls, A. O., & Margules, C. R. (1990). Measurement of the realized qualitative niche: environmental niches of five Eucalyptus species. *Ecological monographs*, *60*(2), 161-177.

Baum, J. K., & Blanchard, W. (2010). Inferring shark population trends from generalized linear mixed models of pelagic longline catch and effort data. *Fisheries Research*, *102*(3), 229-239.

Bender, D. J., Contreras, T. A., & Fahrig, L. (1998). Habitat loss and population decline: a meta-analysis of the patch size effect. *Ecology*, *79*(2), 517-533.

Berkson, J. (1950). Are there two regressions? *Journal of the american statistical association*, *45*(250), 164-180.

Bishop, J., Venables, W. N., & Wang, Y. G. (2004). Analysing commercial catch and effort data from a penaeid trawl fishery: a comparison of linear models, mixed models, and generalised estimating equations approaches. *Fisheries Research*, *70*(2), 179-193.

Campbell, R. A. (2004). CPUE standardisation and the construction of indices of stock abundance in a spatially varying fishery using general linear models. *Fisheries Research*,*70*(2), 209-227.

Chiasson, Y. J., Vienneau, R., DeGrace, P., Campbell, R., Hebert, M., & Moriyasu, M. (1993). *Evaluation of catch selectivity of modified snow crab (Chionoecetes opilio) conical traps*. Department of Fisheries and Oceans, Gulf Fisheries Centre.

Dauk, P. C., & Schwarz, C. J. (2001). Catch estimation in the presence of declining catch rate due to gear saturation. *Biometrics*, 287-293.

Dawe, E. G., Walsh, S. J., & Hynick, E. M. (2010). Capture efficiency of a multi-species survey trawl for Snow Crab (*Chionoecetes opilio*) in the Newfoundland region. *Fisheries Research*, *101*(1), 70-79.

Delong, J. P., & Vasseur, D. A. (2013). Linked exploitation and interference competition drives the variable behavior of a classic predator–prey system. *Oikos*, *122*(10), 1393-1400.

Dick, E. J. (2004). Beyond ‘lognormal versus gamma’: discrimination among error distributions for generalized linear models. *Fisheries Research*, *70*(2), 351-366.

Field, A. P. (2001). Meta-analysis of correlation coefficients: a Monte Carlo comparison of fixed-and random-effects methods. *Psychological methods*, *6*(2), 161.

Gates, S. (2002). Review of methodology of quantitative reviews using meta‐analysis in ecology..*Journal of Animal Ecology*, *71*(4), 547-557.

Gillis, D. M. (2003). Ideal free distributions in fleet dynamics: a behavioral perspective on vessel movement in fisheries analysis. *Canadian Journal of Zoology*, *81*(2), 177-187.

Gillis, D. M., & Peterman, R. M. (1998). Implications of interference among fishing vessels and the ideal free distribution to the interpretation of CPUE. *Canadian Journal of Fisheries and Aquatic Sciences*, *55*(1), 37-46.

Gillis, D. M., Peterman, R. M., & Tyler, A. V. (1993). Movement dynamics in a fishery: application of the ideal free distribution to spatial allocation of effort. *Canadian Journal of Fisheries and Aquatic Sciences*, *50*(2), 323-333.

Glass, G.V. (1976). Primary, secondary and meta-analysis. Educational Researcher 5, 3-8.

Goñi, R., Alvarez, F., & Adlerstein, S. (1999). Application of generalized linear modeling to catch rate analysis of Western Mediterranean fisheries: the Castellón trawl fleet as a case study. *Fisheries research*, *42*(3), 291-302.

Gurevitch, J., Morrow, L. L., Wallace, A., and Walsh, J. S. (1992). A meta-analysis of competition in field experiments. *American Naturalist 140*, 539-572.

Halliday, R. G. (1988). Use of seasonal spawning area closures in the management of haddock fisheries in the Northwest Atlantic. *NAFO Scientific Council Studies*, *12*, 27-36.

Harley, S. J., Myers, R. A., & Dunn, A. (2001). Is catch-per-unit-effort proportional to abundance? *Canadian Journal of Fisheries and Aquatic Sciences*, *58*(9), 1760-1772.

Helser, T. E., Punt, A. E., & Methot, R. D. (2004). A generalized linear mixed model analysis of a multi-vessel fishery resource survey. *Fisheries Research*, *70*(2), 251-264.

Higgins, J., & Thompson, S. G. (2002). Quantifying heterogeneity in a meta-analysis. *Statistics in medicine*, *21*(11), 1539-1558.

Higgins, J., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *Bmj*, *327*(7414), 557-560.

Hilborn, R. (1985). Fleet dynamics and individual variation: why some people catch more fish than others. *Canadian Journal of Fisheries and Aquatic Sciences*, *42*(1), 2-13.

Hilborn, R., & Walters, C. J. (1992). Quantitative fisheries stock assessment: choice, dynamics and uncertainty. *Reviews in Fish Biology and Fisheries*, *2*(2), 177-178.

Imbrie, J. (1956). Biometrical methods in the study of invertebrate fossils. Bulletin of the AMNH; v. 108, article 2.

Jarvinen, A. (1991). A meta‐analytic study of the effects of female age on laying‐date and clutch‐size in the Great Tit Parus major and the Pied Flycatcher Ficedula hypoleuca. *Ibis*, *133*(1), 62-67.

Jones, J. B. (1992). Environmental impact of trawling on the seabed: a review. *New Zealand Journal of Marine and Freshwater Research*, *26*(1), 59-67.

Kimura, D. K. (1981). Standardized measures of relative abundance based on modelling log (cpue), and their application to Pacific ocean perch (Sebastes alutus). *Journal du Conseil*, *39*(3), 211-218.

Koricheva, J., Gurevitch, J., & Mengersen, K. (eds.). (2013). *Handbook of meta-analysis in ecology and evolution*. Princeton University Press.

LaBarbera, M. (1989). Analyzing body size as a factor in ecology and evolution. *Annual review of ecology and systematics*, 97-117.

Lapointe, M. F. (1989). *Numerical and functional responses of British Columbia trawler fishermen* (Doctoral dissertation, M.Sc. thesis, University of British Columbia, Vancouver).

Laws, E. A., & Archie, J. W. (1981). Appropriate use of regression analysis in marine biology. *Marine Biology*, *65*(1), 13-16.

Marchal, P., Andersen, B., Bromley, D., Iriondo, A., Mahévas, S., Quirijns, F., & Ulrich, C. (2006). Improving the definition of fishing effort for important European fleets by accounting for the skipper effect. *Canadian Journal of Fisheries and Aquatic Sciences*, *63*(3), 510-533.

Maunder, M. N. (2001). A general framework for integrating the standardization of catch per unit of effort into stock assessment models. *Canadian Journal of Fisheries and Aquatic Sciences*, *58*(4), 795-803.

Maunder, M. N., & Punt, A. E. (2004). Standardizing catch and effort data: a review of recent approaches. *Fisheries Research*, *70*(2), 141-159.

Maunder, M. N., Sibert, J. R., Fonteneau, A., Hampton, J., Kleiber, P., & Harley, S. J. (2006). Interpreting catch per unit effort data to assess the status of individual stocks and communities. *ICES Journal of Marine Science: Journal du Conseil*, *63*(8), 1373-1385.

Messieh, S. N., Rowell, T. W., Peer, D. L., & Cranford, P. J. (1991). The effects of trawling, dredging and ocean dumping on the eastern Canadian continental shelf seabed. *Continental Shelf Research*, *11*(8), 1237-1263.

Millington, P. (1984). The fishermen as predator: numerical responses of British Columbia gillnet fishermen to salmon abundance.

Milsom, T. P., Langton, S. D., Parkin, W. K., Peel, S., Bishop, J. D., Hart, J. D., & Moore, N. P. (2000). Habitat models of bird species' distribution: an aid to the management of coastal grazing marshes. *Journal of Applied Ecology*, *37*(5), 706-727.

Orbach, M. K. (1977). *Hunters, seamen, and entrepreneurs: the tuna seinermen of San Diego*. Univ of California Press.

Paloheimo, J. E., & Dickie, L. M. (1964). Abundance and fishing success. *Rapports et Proces Verbaux des Réunions du Conseil International pour l'Exploration de la Mer*, *155*.

Paul, A. J. (1992). A review of size at maturity in male Tanner (Chionoecetes bairdi) and king (Paralithodes camtschaticus) crabs and the methods used to determine maturity. *American Zoologist*, *32*(3), 534-540.

Pawitan, Y., Reilly, M., Nilsson, E., Cnattingius, S., & Lichtenstein, P. (2004). Estimation of genetic and environmental factors for binary traits using family data. *Statistics in medicine*, *23*(3), 449-465.

Pelletier, D., & Ferraris, J. (2000). A multivariate approach for defining fishing tactics from commercial catch and effort data. *Canadian Journal of Fisheries and Aquatic Sciences*, *57*(1), 51-65.

Poos, J. J., & Rijnsdorp, A. D. (2007). An "experiment" on effort allocation of fishing vessels: the role of interference competition and area specialization. *Canadian Journal of Fisheries and Aquatic Sciences*, *64*(2), 304-313.

Post, J. R., Sullivan, M., Cox, S., Lester, N. P., Walters, C. J., Parkinson, E. A. & Shuter, B. J. (2002). Canada's recreational fisheries: the invisible collapse? *Fisheries*, *27*(1), 6-17.

Quinn, G. P., & Keough, M. J. (2002). *Experimental design and data analysis for biologists*. Cambridge, UK: Cambridge University Press.

Raudenbush, S. W. (1994). Random effects models. Pages 301– 321 *in* H. Cooper and L. V. Hedges (eds.). *The handbook of research synthesis*. New York: Russel Sage Foundation.

Richardson, L., & Loomis, J. (2009). The total economic value of threatened, endangered and rare species: an updated meta-analysis. *Ecological Economics*, *68*(5), 1535-1548.

Ricker, W. E. (1973). Linear regressions in fishery research. *Journal of the Fisheries Board of Canada*, *30*(3), 409-434.

Ricker, W.E. (1975). Computation and interpretation of biological statistics of fish populations. *Bull. Fish. Res. Board Can*. No. 191.

Rijnsdorp, A. D., Dol, W., Hoyer, M., & Pastoors, M. A. (2000). Effects of fishing power and competitive interactions among vessels on the effort allocation on the trip level of the Dutch beam trawl fleet. *ICES Journal of Marine Science: Journal du Conseil*, *57*(4), 927-937.

Roberts, P. D., Stewart, G. B., & Pullin, A. S. (2006). Are review articles a reliable source of evidence to support conservation and environmental management? A comparison with medicine. *Biological Conservation*, *132*(4), 409-423.

Rose, G. A. (1992). Indices of Total Stock Biomass in the "Northern" and Gulf of St. Lawrence Atlantic Cod (*Gadus morhua*) Stocks Derived from Time Series Analyses of Fixed Gear (Trap) Catches. *Canadian Journal of Fisheries and Aquatic Sciences*, *49*(2), 202-209.

Rothschild, B. J. (1967). Competition for gear in a multiple-species fishery. *Journal du Conseil*, *31*(1), 102-110.

Salthaug, A., & Godø, O. R. (2000). Analysis of CPUE from the Norwegian bottom trawl fleet. *ICES*.

Saville, A., & Oe, U. (1980). Factors affecting the reaction of pelagic fish to exploitation and requiring a new approach to assessment and management [fishing pressure, Norwegian spring-spawning herring, recruitment failure, stock size]. *Rapports et Proces Verbaux des Reunions*.

Smith, S. H. (1968). Species succession and fishery exploitation in the Great Lakes. *Journal of the Fisheries Board of Canada*, *25*(4), 667-693.

Smith, R. J. (2009). Use and misuse of the reduced major axis for line-fitting. *American Journal of Physical Anthropology*, *140*(3), 476-486.

Squires, D. (1987). Public regulation and the structure of production in multiproduct industries: an application to the New England otter trawl industry. *The Rand Journal of Economics*, 232-247.

Stergiou, K. I., & Erzini, K. (2002). Comparative fixed gear studies in the Cyclades (Aegean Sea): size selectivity of small-hook longlines and monofilament gill nets. *Fisheries Research*, *58*(1), 25-40.

Stillman, R. A., Goss-Custard, J. D., & Caldow, R. W. G. (1997). Modelling interference from basic foraging behaviour. *Journal of Animal Ecology*, 692-703.

Su, N. J., Yeh, S. Z., Sun, C. L., Punt, A. E., Chen, Y., & Wang, S. P. (2008). Standardizing catch and effort data of the Taiwanese distant-water longline fishery in the western and central Pacific Ocean for bigeye tuna, *Thunnus obesus*. *Fisheries Research*, *90*(1), 235-246.

Swain, D. P., & Wade, E. J. (2003). Spatial distribution of catch and effort in a fishery for snow crab (*Chionoecetes opilio*): tests of predictions of the ideal free distribution. *Canadian Journal of Fisheries and Aquatic Sciences*, *60*(8), 897-909.

Vergara, P., I Aguirre, J., & Fernández - Cruz, M. (2007). Arrival date, age and breeding success in white stork Ciconia ciconia. *Journal of Avian Biology*, *38*(5), 573-579.

Vignaux, M. (1996). Analysis of vessel movements and strategies using commercial catch and effort data from the New Zealand hoki fishery. *Canadian Journal of Fisheries and Aquatic Sciences*, *53*(9), 2126-2136.

Watling, L., & Norse, E. A. (1998). Disturbance of the seabed by mobile fishing gear: a comparison to forest clearcutting. *Conservation Biology*, *12*(6), 1180-1197.

Weston, R. (2011). *Report On The Snow Crab Industry In The Atlantic Provinces And In Quebec*. House of Commons Standing Committee on Fisheries and Oceans. Canada.

Winger, P. D., & Walsh, P. J. (2011). Selectivity, efficiency, and underwater observations of modified trap designs for the snow crab (*Chionoecetes opilio*) fishery in Newfoundland and Labrador. *Fisheries Research*, *109*(1), 107-113.

Xu, X., & Millar, R. B. (1993). Estimation of trap selectivity for male snow crab (*Chionoecetes opilio*) using the SELECT modeling approach with unequal sampling effort. *Canadian Journal of Fisheries and Aquatic Sciences*, *50*(11), 2485-2490.

Ye, Y., Al-Husaini, M., & Al-Baz, A. (2001). Use of generalized linear models to analyze catch rates having zero values: the Kuwait driftnet fishery. *Fisheries Research*, *53*(2), 151-168.

Chapter Two

Meta-analysis of published data relating catch to nominal effort

Abstract

My study analyzes ninety-one previously published cases, drawn from the broader literature via a search of 3246 journal articles and reports, to examine proportionality between catch and effort in fisheries' data. The formal meta-analysis for this project examines proportionality through the slope of the log-log regression of catch on effort. To account for any "errors-in-variables" bias, reduced major axis (RMA) regressions are used to define the relationship. I investigated the role of fishery type, effort unit, gear, gear type, and target species as moderator variables in a mixed-effects meta-analysis. The typical disproportionality that was expected from facilitation and a numerical response of effort to favorable catches was indicated by the meta-analysis, with the observed slope of the log-log regressions exceeding one. However, in specific fisheries, the slopes can be higher, lower or close to one, suggesting that the effects of vessel behaviour (interference competition, facilitation, or numerical responses) should be considered for each fishery, individually. Among moderator variables (covariates) fishery type was shown to impact the proportionality between catch and effort

Key words: Meta-analysis, catch, effort, error-in-variables, proportionality, disproportionality, mixed-effects, RMA, OLS.

Introduction

Traditionally, catch in a fishery is assumed to be proportional to both the abundance of the species pursued (target) and the amount of fishing activity (effort). This assumption is the basis for the use of catch-per-unit-effort (CPUE) as both an index of abundance (Richards and Schnute 1986) and as a data series in more complex methods such as stock synthesis (Methot and Wetzel 2013). However, many processes know to play a role in fishing activities could invalidate the assumption of proportionality. The disproportionality could result from the changing behaviour of fishing fleets through competition, fishing facilitation, or a numerical response to fish availability and other factors. There are however a number of studies that relate the disproportionality between catch and effort to interference competition. This occurs when one forager directly interacts with another, or interferes by dispersing prey, while the second is trying to access a resource (Stillman et al. 1997). When vessels aggregate in an area, interference among vessels may also become more intense. In this case, a doubling of effort may not double the catch due to interference; catch may increase to a lesser extent. Overall, this would contribute to a trend where the observed increase in catch with effort decelerates. This is because vessels will negatively affect each other in their ability to find fish. For example, Abrahams and Healey (1993) tested some factors that affect the distribution of B.C. salmon trawlers. The vessels caught chinook salmon and found that, when vessel density increased, the catch rate of chinook salmon decreased, so the degree of change in catch rate in relation to the changes in vessel density was significant. The result of Abrahams' study suggests that variation in vessel density may exert a substantial influence on catch rates during the commercial fishing season. Gillis (2003) also took a similar point of view, when he discussed foragers based on ideal free distributions in fleet dynamics (IFD). The IFD is a behavioural hypothesis that examines the spatial relationship between foragers and their resources and assumes interference competition (Poos and Rijnsdorp 2007) exists among foragers (vessels). In a commercial fishery like the Pacific herring (*Clupea pallasi*) fishery, fish may disperse in response to fishermen setting their nets or moving their vessels, reducing fishing success. Congestion of vessels can also affect fishing gear through loss or damage (by entanglement with other gear). This will also decrease the overall catch (Gillis and Peterman 1998)

An alternative form of disproportionality can result from facilitation (Lapointe 1989) and numerical responses (Solomon 1949, Millington 1984, Ledbetter 1986). In general, numerical responses occur when the predator density changes in relation to changes in prey density (Post et al. 2008). In fisheries, numerical responses result from additional vessels being mobilized when fish are more abundant. For example, when vessels try to allocate fishing effort in an area and have the ability to exploit fish aggregations, this will influence their fishing efficiency and the catch rate. In the case of facilitation, the fish harvest becomes proportionately higher when more fishing effort is applied. For example, even when vessel density increases, fishers can increase their catches by communicating with each other about the best areas to fish in. In addition, having more vessels in a water body facilitates the coverage of larger areas, increasing the chance to catch fish and thus the effectiveness of nominal fishing effort.

A synthesis of published fisheries data can provide insight into the generality of proportionality between catch and effort. Meta-analysis, first introduced by Gene V. Glass (1976), and is becoming a popular method of research synthesis. It can be conducted qualitatively or quantitatively: qualitatively in the form of a narrative review, and quantitatively by using statistical methods (Koricheva et al. 2013). Meta-analysis has become a common tool for the synthesis of previous research in ecology and fisheries. It uses techniques that collect an analytical result and then integrate these findings with others to get a wider understanding of a specific area of study. Combining the findings from a number of studies strengthens the evidence for a specific question or hypothesis. Gurevitch et al. (2001) have reviewed the history of meta-analysis and discussed its application in the field of ecology and evolution. Today, there are many studies using meta-analysis in ecology and evolutionary biology (Gurevitch et al. 1992, Arnqvist and Wooster 1995, Adams et al. 1997, Bender et al. 1998, Koricheva 2002, Harrison 2011). There are also a growing number of papers using meta-analysis for fisheries (Côté et al. 2001, Worm and Myers 2003, Evans et al. 2011). For instance, Worm and Myers (2003) applied meta-analysis in the context of population interactions across the North Atlantic Ocean. They used data for Atlantic cod (*Gadus morhua*) and Northern shrimp (*Pandalus borealis*) drawn from the NAFO (Northwest Atlantic Fisheries Organization, Dartmouth, Nova Scotia, Canada) database to test for patterns in the catches of target species. Catch and effort data from different fisheries have been extensively analyzed for various purposes. The purpose of meta-analysis is to combine all these data into one study to arrive at a conclusion or theory in response to the question asked. In my study, I will collect and synthesize data from published fisheries studies to perform a meta-analysis that addresses the questions: is catch generally proportional to effort in published fisheries data, and does the relationship vary substantially among fisheries?

Methods

The data sources used in this study were collected using the Google Scholar search engine, the International Commission for the Conservation of Atlantic Tuna (ICCAT) website (http://www.iccat.int), and the International Council for the Exploration of the Sea (ICES) website (http://www.ices.dk). Initially, searching the databases with the search terms “catch” and "effort" identified potential studies within these sources. To find out if a given paper discussed both catch and effort, I examined its summary in the data source. If the summary referenced both catch and effort, I then reviewed the paper to determine if that study explicitly reported these catch and effort data. If not, I excluded it from my research. I reviewed 1000 summaries using the Google Scholar database and found 606 with catch and effort data. I reviewed 2246 summaries from the ICCAT website and found 1155 papers with catch and effort data. I did not find any time series data containing both catch and effort data using the ICES website. Finally, I selected only the documents from the list with long-term studies (studies that had a time series of data from the same site over multiple years). In total, only 91 out of the 3246 potentially relevant cases in the databases met all of my search criteria.

The meta-analysis for this project used two measures of effect size. The first measure was the ordinary least square regression slope (OLS) as an effect size, and the standard errors of the slopes as sample variability. The second measure was the reduced major axis slope (RMA). The RMA slope has the same standard errors as the ordinary least square regression slopes (Sokal and Rohlf 1995, Quinn and Keough 2002, Price and Phillimore 2007). The Model II regression can be found by minimizing the sum of the triangle areas constructed by vertical and horizontal lines from each observed value to the fitted line. The RMA measure accounts for error-in-variables that can bias OLS slopes (Quinn and Keough 2002, p. 101). For my analysis, I used both random and mixed-effects meta-analysis models to examine these effect size measures. A mixed-effects model allows for a combination of both fixed and random effects. There are two levels to this model: the parameter being studied (the fixed-effects model) and the variations between different studies (the random effect) (Konstantopoulos 2006).

The catch model that has been used in current fisheries analysis is shown in the equation below (Maunder and Punt 2004):

$$C_i = \beta_0 \cdot f_i \cdot N_i \tag{2.1}$$

where C_i is the catch, β_0 is a constant (catchability), f is the fishing effort, and N is the abundance, which may be represented as density. I can modify equation (2.1) by adding powers, which allows for potential nonlinearity in the relationships between catch and either effort or abundance.

$$C_i = \beta_0 \cdot {f_i}^{\beta 1} \cdot {N_i}^{\beta 2} \tag{2.2}$$

In equation (2.2), proportionality is reflected in the variables $\beta 1$ and $\beta 2$. I examine proportionality by testing whether β is equal to one, greater than one, or less than one, $i = 1$, 2,…, n, i is the observation of the slope. I used log transformation on both catch and effort to convert the multiplication into addition. Equation (2.2) will be rewritten as

$$log\ C_i = log(\beta_0) + \beta_1\ log(f_i) + \beta_2 \log(N_i) \tag{2.3}$$

where β_1 equal one. These slopes are the effect sizes in my meta-analysis. In this form, the effect sizes are independent of the measurement scale of the original catch and effort data. I do not have abundance but I added year effect to reflect inter-annual changes in fish availability.

I also examined fishery covariates (moderator variables) that could be determined from the original sources: fishery type (such as artisanal, commercial, or research); nominal effort measure (such as hours, sets, trips, or hooks); gear type used (such as mobile, fixed, or a mixture of both); gear (such as gillnet, trawl, electrofishing, purse-seine, pole and line, long line, or mixed); target species (yes or no); and species category (tuna, such as bigeye tuna (*Thunnus obesus*), shark, such as blue shark (*Prionace glauca*), shrimp, such as brown tiger prawns (*Penaeus esculentus*), redfish (*Sebastes mentella*); billfish, such as swordfish (*Xiphias gladius*), salmon, such as chinook salmon (*Oncorhynchus tshawytsch*), haddock (*Melanogrammus *aeglefinus*), herring, such as Atlantic herring (*Clupea harengus*), cod *(Gadus morhua)*, or Nile perch (*Lates niloticus*)). Equation (2.4) will allow me to discover the effect of these covariates on the slope. The alteration of the slope will indicate the effect of the covariates on the relationship between catch and effort. Each study generated a slope, which was the effect size in the meta-analysis to which the moderator was applied.

$$Y_{ij} = \mu + m_j + \alpha_i \tag{2.4}$$

where Y_{ij} is the observed slope, μ is the overall mean, m_j is the moderator, α_i is the random effect for the study, i is the observation (the slope from the study i), and j is the value of the moderator as a fixed effect. In a statistical review, heterogeneity is any variability in study outcomes that occurs among studies. In my case, the heterogeneity is the variability among studies between true effects (slopes). If I have high heterogeneity, the studies I have selected may be too different from each other to combine. For interpreting heterogeneity, Higgins et al. (2003) suggest that we categorize as follows: 0% = no heterogeneity, 25% = low heterogeneity, 50% = moderate heterogeneity, and 75% = a high level of heterogeneity. In this study, I am interested in knowing if it is best to use a fixed-effects model or a random-effects model. Fixed-effects assume the same true effect size in each study, and the sample variability estimated within each study. A random-effects model allows for the true effect size to vary among studies and assumes these differences between effect sizes is because of random variation around the mean of effect size. Random effects include both variation between studies (due to random differences in true effect size) and within study (due to sample variability or sample error). The mixed effect model combines both fixed and random effects.

I used a mixed-effects model in this study to allow me to determine the effect of moderators on the size of the average slope (Viechtbauer 2010). Fishery data may have different slopes depending on the known and unknown factors in each case. The true effects (slopes) are assumed to be distributed normally, with a variance of τ^2, μ_i~N (0, τ^2). To fit a mixed-effects model, I needed to estimate τ^2, the amount of residual heterogeneity among true effects (slope). In my case, I have estimated τ^2 with a restricted maximum-likelihood estimator (Viechtbauer 2010) using the R programing language with the *metafor* package (R Core Team 2014). *Metafor* can fit both random- and mixed-effects models, providing a variety of diagnostics and allowing the examination of moderador variables as fixed effects (Viechtbauer 2010).

The moderator variables used were: fishery type (including artisanal, commercial, and research), effort unit (including hours, sets, boats, days, hooks, and trips), and gear type (including fixed, mobile, and a mixture of both). I combined the reduced major axis regression slope as effect size (random effect) with various moderator variables (fixed effects). Fitting moderators as fixed effects in a mixed-effects model creates "dummy variables" to represent the individual levels in the moderator variables. For example, fishery type has three levels: artisanal, commercial, and research. One of these levels becomes the reference level (β_0). By default, the first level of the moderator will be the reference level. β_0 will estimate the reference level slope, β_1 and $\beta_2$will indicate how much lower or higher the slope is than the reference level. The slope for levels other than the base level was calculated by adding the parameter estimate to the intercept. The 95% confidence interval was calculated as shown:

$$\beta_i \pm 1.96\ se(\beta_i)$$

where β_i is the parameter estimate, se(β_i) is the standard error of the parameter, 1.96 is the quantile of the normal distribution, and i is the observation of the slope. If the slope is different in each of the studies, this will result in significant heterogeneity and support the representation of slope as a random-effect. For this, I am going to test for heterogeneity. If τ^2 is greater than 0, then there is heterogeneity among the true effects. I^2 composed of both heterogeneity among studies and sampling variability. The test of moderators (QM) will indicate if coefficients (moderator levels excluding the reference) are equal to 0 or greater than 0. This test will show if the moderators contribute to the heterogeneity in the slope.

To begin, I examined the correlation between log catch and log effort, which represents the strength of their relationship I expected that there would be a strong relationship between catch and effort. To confirm that, I analyzed the data using random-effects models with Fisher's z-transformation of correlations as the effect size. Fisher's z-transformation was developed to deal with a non-normally distributed Pearson's r variable through a normally distributed z variable. I transformed the average of z values back into the Pearson's correlation for display purposes (Silver and Dunlap 1987, Alexander 1990). I used the back transformation as effect size with a random-effects model and a mixed-effects model. I then used the mixed-effects model with a number of moderator variables: fishery type, target species, effort type, effort unit, gear type (mobile or fixed), and gear to examine influences on the strength of the correlation. I fit the model with and without one of the moderators (full and reduced model) and used likelihood ratio tests to compare them.

Results

To begin, I examined the correlation between catch and nominal effort in my set of ninety-one studies. There is generally a high correlation between log transformed catch and effort. Overall, the mean of the correlations was 0.75. Figure 2.1 also shows how the strength of the relationship between catch and effort varied from one study to another. This was consistent with the results of the mixed-effects models using Fisher's z (transformed correlation) as effect size (Figure 2.2). In Figure 2.2, the forest plot shows Fisher's z-transformed correlations of log catch and log effort as measured by various studies. There is clear heterogeneity from one study to another. The total variability in the strength of the relationship, which is composed of both heterogeneity among studies and sampling variability, was shown as the percentage: $I^2 = 86.12\%$. The heterogeneity in the correlations was estimated to be $\tau^2 = 0.4671$ (Table 2.1). This result suggests that there is high heterogeneity among studies in the strength of the relationship between catch and effort. The funnel plot (Figure 2.2) appears symmetric and the regression test for the funnel plot did not suggest the asymmetry that would indicate selection bias in the studies used: $p = 0.6598$.

The strength of the correlations varied among the moderator variables. The test of moderators (QM) using Fisher's z (transformed correlation) showed that the strength of the relationship between catch and effort increased when fishermen caught their intended target ($p=0.0237$, Table 2.2). Fishery type and gear were not significant using the QM test (Table 2.2). The likelihood-ratio tests (LTR) of possible moderators agreed with the QM test for target species, but also indicated that gear type was a significant moderator (Table 2.2). However, a more complex model containing both target and gear type was not a significant improvement over using target species alone as a moderator (LRT (df= 2)= 5.0619, p-value = 0.0245)

I examined the proportionality between catch and effort using random (no moderator) (Table 2.3) with both OLS and RMA slopes as effect size. The OLS slope was generally less than one (mean = 0.73 for all studies) but greater than one (mean = 1.15 for all studies) for the RMA slope. The 95% confidence intervals for each regression excluded one, which is the slope associated with proportionality. The total variability in slopes, which is composed of both heterogeneity among studies and sampling variability, was shown as a percentage: I^2= 56.98% (Table 2.1) (Higgins and Thompson 2002). The heterogeneity in the true slope was estimated to be τ^2 = 0.1912 (Table 2.1). I followed Higgins et al. (2003) in their categorization of the heterogeneity; if the heterogeneity among the studies exceeds 50%, then this is classified as moderate and the estimates of these studies should not be treated as a single common value, which would correspond to a fixed effect model for slopes in the absence of moderators.

Moderator variables on the regression slopes were investigated in more detail with the RMA regressions. The outcomes of the mixed-effects model using RMA slope as effect size show that the test of moderator (QM) produced a p-value of 0.0320 for fishery type and a p-value of 0.0668 for gear type. The confidence intervals for-this moderator suggested that catch increases disproportionately faster than effort in artisanal fisheries. Gear type approached significance when using mixed gear (both fixed and mobile gear in the fishery) (Table 2.4) though the reasons for this are unclear from the level of detail available for these fisheries from the literature. Other potential moderators (whether or not the species was targeted, effort unit, specific gear used) were not significant. When RMA slope was used with fishery type as a moderator variable, the total variability in the slopes, which is composed of both heterogeneity among studies and sampling variability, was shown as a percentage: I^2 = 56.67%. The heterogeneity in the true slope was estimated to be τ^2 = 0.1884 (Table 2.1). Once again, the analysis suggested that a single "typical" slope for the log catch on log effort regressions would not represent the variability in proportionality among fisheries.

I have used a forest plot and a funnel plot to show the results of analyzing the RMA slopes of my ninety-one studies, without a moderator variable of fishery. In the forest plot, confidence intervals vary in width and the inclusion of one (proportionality). The funnel plot appears to be symmetrical, indicating that there is no clear bias in study selection. This was also supported by the regression test which did not suggest asymmetry: $p = 0.6068$ (Figure 2.3).

Table 2.1: Heterogeneity measures for meta-analysis. The outcomes of random- and mixed-effects models of fishery type using OLS slope as effect size. k = 91, where k is the number of studies. τ^2 is the heterogeneity in the true slope. The estimator is REML. SE is the standard error. I^2 is composed of both heterogeneity among studies and sampling variability.

Model Effect size ~ moderator	τ^2	SE	I^2
Fisher's z ~ fishery type	0.4680	0.0905	86.14%
Fisher's z	0.4671	0.0893	86.12%
OLS slope ~ fishery type	0.1352	0.0458	48.43%
OLS slope	0.1341	0.0452	48.16%
RMA slope ~ fishery type	0.1884	0.0554	56.67%
RMA slope	0.1912	0.0554	56.98%

Table 2.2: The outcomes of the mixed-effects models using Fisher's z (trans. correlation) as effect size. n is the number of studies. SE is the standard error. QM and LRT are the tests of moderators.

Moderator	n	Fisher's z	SE	CI $_{lower\ 95\%}$	CI $_{upper\ 95\%}$
Non	90	0.9623	0.0823	0.8010	1.1235
Fishery type: test of moderators: QM (df = 2) = 1.5449, p-value = 0.4619					
Fishery type: test of moderators: LRT (df= 2) =1.5889, p-value = 0.4518					
Artisanal	10	0.7593	0.2315	0.3056	1.2129
Commercial	74	0.9717	0.2489	0.4839	1.4595
Research	6	1.2490	0.4012	0.4626	2.0353
Gear: test of moderators: QM (df = 8) = 15.4850, p-value = 0.0504					
Gear: test of moderators: LRT (df= 2) = 15.6023, p-value = 0.0484					
Bait boat	2	1.8103	0.4995	0.8312	2.7893
Electrofishing	2	0.0656	0.7388	-1.3823	1.5135
Gillnet	8	0.4793	0.5564	-0.6112	1.5697
Long line	25	0.9602	0.5225	-0.0637	1.9842
Mix	9	0.9156	0.5522	-0.1666	1.9978
Pole and line	3	1.5086	0.6369	0.2602	2.7569
Purse-seine	18	1.132	0.5316	0.0901	2.1739
Shore angling	1	2.2222	0.8776	0.502	3.9423
Trawl	22	0.9052	0.5236	-0.1209	1.9314

Table 2.2 (continued from previous page)

Target species: test of moderators: QM (df = 1) = 5.1190, p-value = 0.0237

Target species: test of moderators: LRT (df= 2)= 5.0619, p-value = 0.0245

Yes	77	1.4093	0.2130	0.9917	1.8268
No	13	0.8891	0.2299	0.4385	1.3397

Table 2.3: The outcomes of the random-effects model using OLS, RMA slopes, and Fisher's z as effect size with no moderator variables.

Effect size	K	Criteria	Slope	SE	CI $_{\text{lower 95\%}}$	CI $_{\text{upper 95\%}}$
Fisher's z	90	All studies	0.9623	0.0823	0.8010	1.1235
OLS	91	All studies	0.7278	0.0603	0.6097	0.8460
OLS	60	N>10 years	0.7474	0.0700	0.6103	0.8846
RMA	91	All studies	1.1498	0.0661	1.0203	1.2794
RMA	60	N>10 years	1.1636	0.0812	1.0045	1.3227

Table 2.4: The outcomes of the mixed-effects model using RMA slope as effect size to estimate the slopes of the moderator variables. k = 91. n is the number of studies. SE is the standard error. QM is the test of moderators.

Moderator	n	Slope	SE	CI $_{lower\ 95\%}$	CI $_{upper\ 95\%}$
Fishery type: test of moderators: QM (df = 2) = 6.8843, p-val = 0.0320					
Artisanal	22	1.6775	0.2156	1.2549	2.1001
Commercial	63	1.0881	0.2268	0.6435	1.5326
Research	6	1.278	0.4202	0.4545	2.1015
Effort unit: test of moderators: QM (df = 5) = 4.5990, p-val = 0.4667					
Hours	20	1.1468	0.1585	0.8362	1.4574
Sets	12	0.8893	0.234	0.4306	1.348
Boats	17	1.2187	0.214	0.7993	1.638
Days	8	1.0287	0.273	0.4937	1.5637
Hooks	20	1.3263	0.2086	0.9174	1.7351
Trips	14	1.1004	0.2273	0.6549	1.5458
Gear type: test of moderators: QM (df = 2) = 5.4114, p-val = 0.0668					
Mix	9	1.5633	0.2292	1.1142	2.0125
Fixed	35	1.2232	0.2528	0.7278	1.7186
Mobile	47	1.034	0.246	0.5519	1.516

Figure 2.1: Forest plot showing the result of correlation between catch and effort of ninety-one studies ordered by RMA slope.

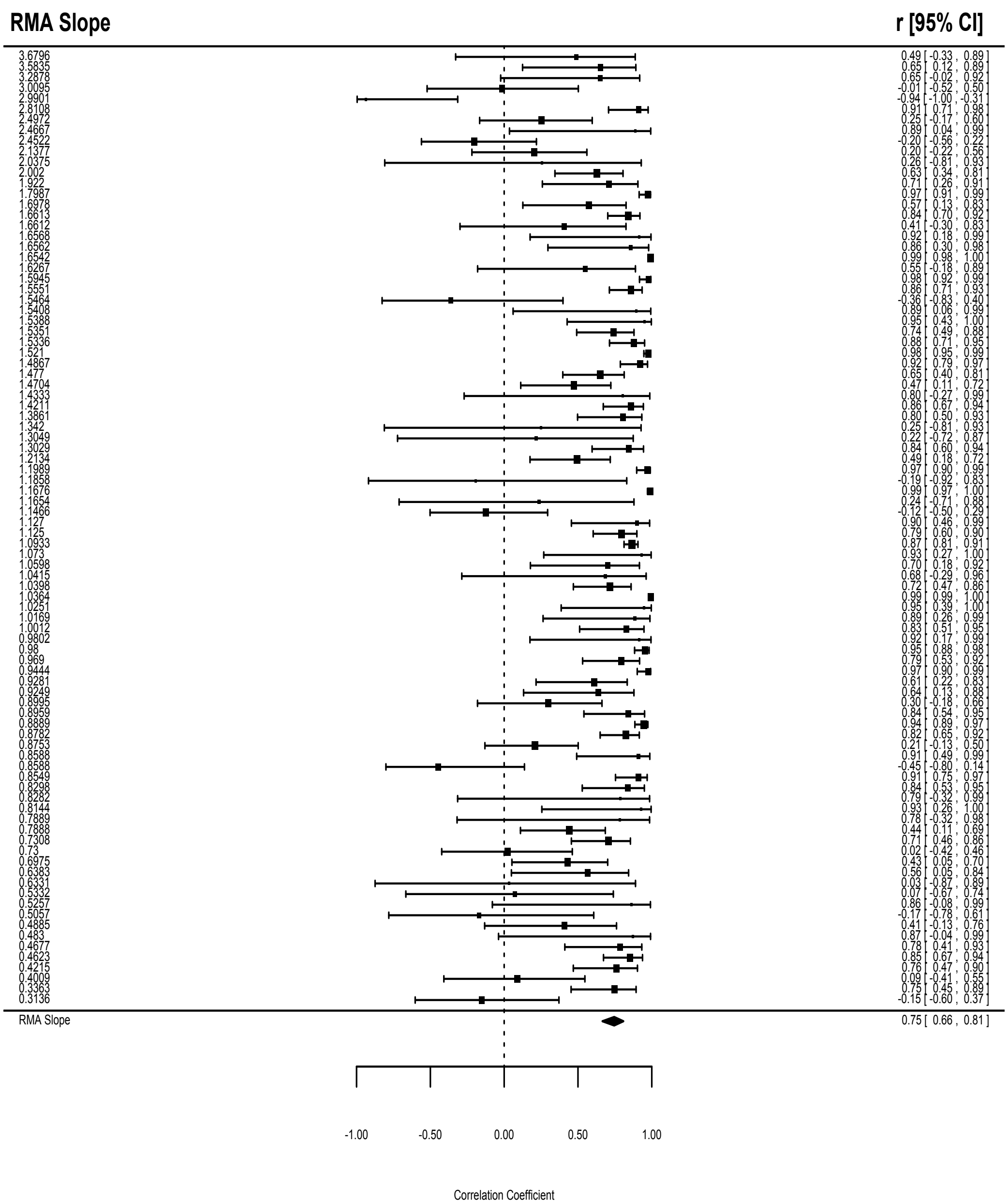


Figure 2.2: Forest, and funnel, showing the results of Fisher's z of ninety-one studies.

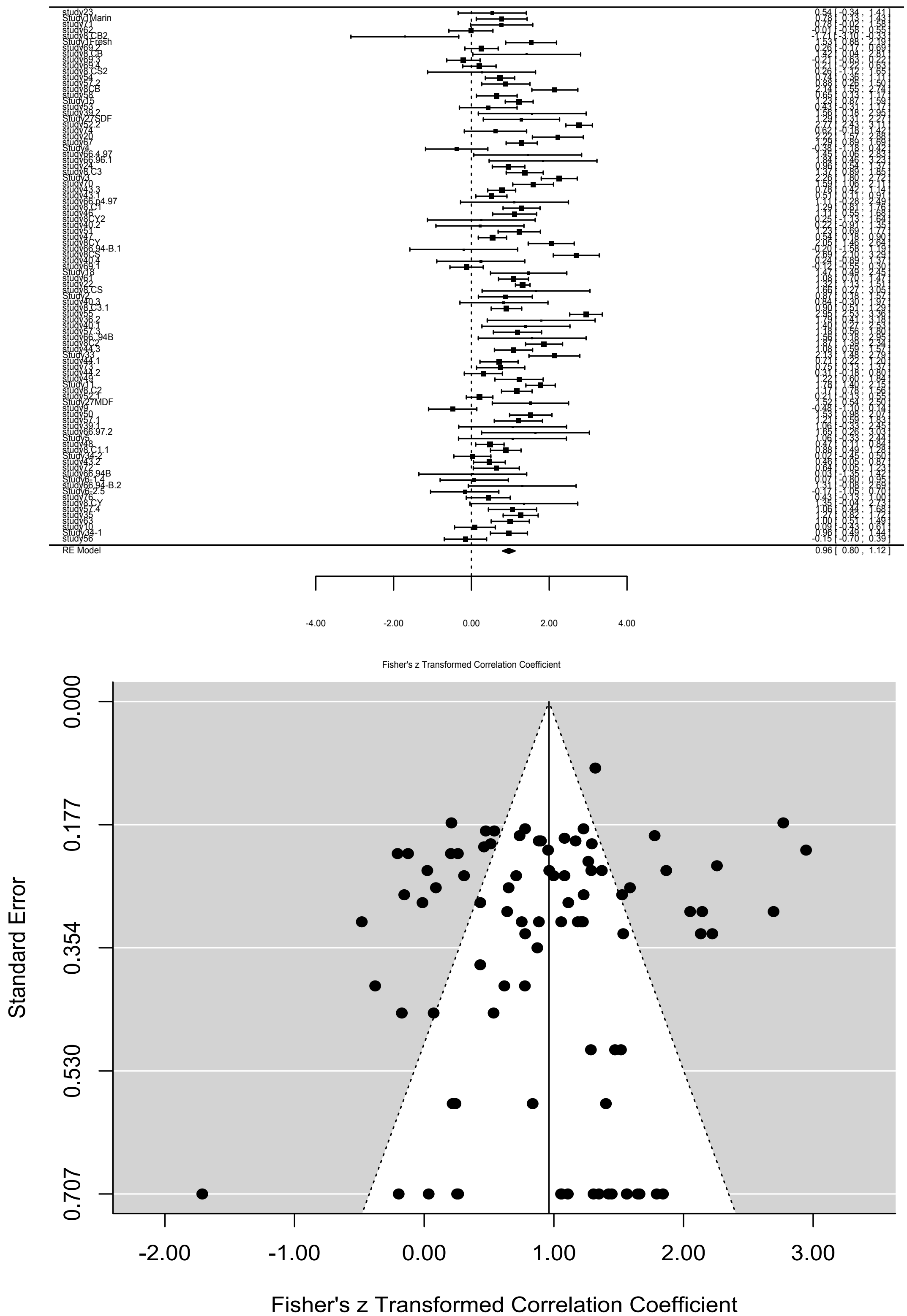
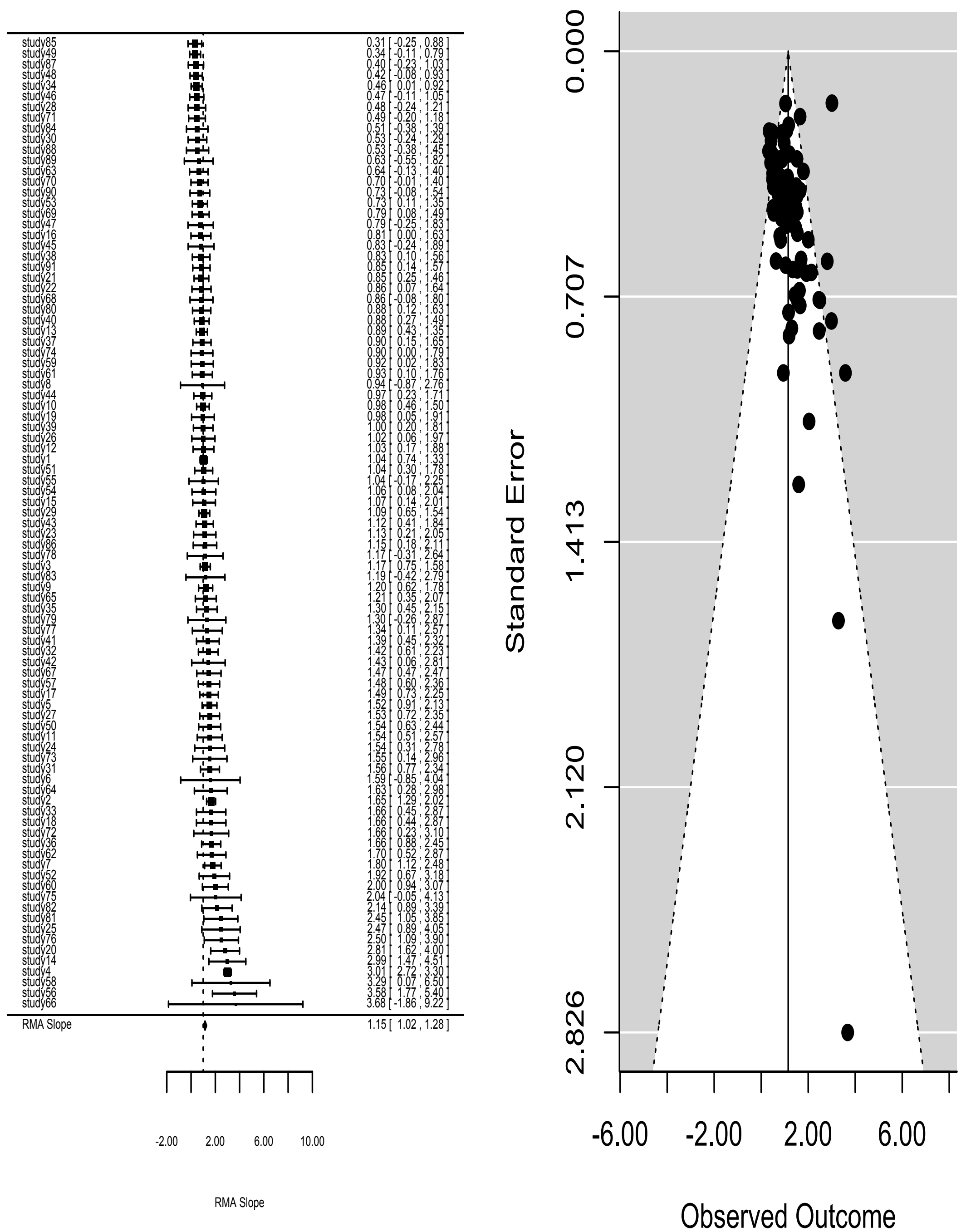


Figure 2.3: Forest plot and funnel plot showing the ninety-one studies without moderator variable, using random-effects model and RMA slope as the effect size.



Discussion

My results show a general tendency for a disproportionate relationship between catch and effort, with much variation among fisheries. There are many possible reasons for disproportionality in the changing behaviour of fishing fleets, such as interference competition (Abrahams and Healey 1993), fishing facilitation (Stuster 1978), or a numerical response to fish availability (Post et al. 2002). The general trend that I observed when using reduced major axis slope (catch increase more than effort) could be caused by either facilitation or numerical responses among fish harvesters. In facilitation, fish harvest becomes more efficient when more fishing effort is applied. For example, Lapointe (1989) found that gross dollar returns per vessel increased when more vessels fished in the same area.

An increase in fleet efficiencies due to information exchange has been observed in a wide variety of fisheries. Stuster (1978) illustrated that many successful U.S. fisheries, such as the west coast crews, use radio transmission to exchange information. They can increase their catch by half through this use of direct communication. Facilitation can also occur through information sharing; for example, Lapointe (1989) found that information sharing helps vessels to compare among areas and helps in predicting seasonal gross dollar returns per vessel. In the San Diego tuna fleet, fish harvesters used two operative strategies; the first known as “the hunter” and the second as “the chaser”. Hunters relied on themselves, were knowledgeable, and had more experience. On the other hand, chasers used radio information to follow the hunters and aggregate around them (Orbach 1977). Vessels can also increase their catch by apparent facilitation For example, when vessels and fish are concentrated in the same area, fish harvesters may be able to successfully locate aggregations of fish, even as overall abundance declines (Deng et al. 2005).

Numerical responses (Post et al. 2008) may also cause catch to increase disproportionately with fishing effort. Numerical responses refer to an increase in the number of vessels or of fishing activity by each vessel when fish are more abundant. The numerical responses can be observed in the way that increased abundance attracts more vessels to fish in the area, such that the number of vessels increases in response to increase in catch. Numerical responses are also shown in fisheries studied by the Inter-American Tropical Tuna Commission, where efforts are concentrated on an area of high fish density (Gulland 1956). Hilborn and Ledbetter (1979) found that vessels responded to salmon densities as well as to the presence of other vessels along the B.C. coast.

In my results, the measure of nominal effort did not affect the proportionality of the relationship between catch and effort. However, one might expect that more highly resolved effort measures would capture more detail and patterns in fleet dynamics. McConnell et al. (1995) discussed the influence of heterogeneity on the estimated value of catch rate among fish harvesters. They found that fishermen who fished long hours, had more experience, or fished in close areas had high historic catch rates. Thus, more detailed effort measures would be expected to differ in proportionality, but I did not see this trend among the fisheries examined.

In each study, various forms of effort were measured, such as hooks, hours, trips, and other types of effort. In some studies, catch was measured by number of fish, and in others it was measured in kilograms. My use of the log-log transformation successfully converted these different measures to the common metric of proportionality. The different units only impact the constant, allowing slopes to provide a common, dimensionless metric. These slopes can be compared to each other among different fisheries, as they are numbers with no units. Using the log transformation on both catch and effort changes multiplication to addition in the basic catch equation (Maunder et al. 2006). This transformation linearizes the relationship and makes the variability in the response (catch) more homoscedastic. More current standardizations use generalized linear models with log links and gamma distributed responses to achieve the same result, but the transformation method is more readily employed for meta-analysis.

The only strong effect of moderators was seen with the fishery type (for slope) and target species (for correlation). It is not possible to definitively say why artisanal fisheries demonstrated slopes significantly greater than 1. However, this may be a result of the greater flexibility in numerical responses. Fishers who have alternative activities, such as farming or other vocations on shore, may only enter the fishery when conditions are very good, while industrialized commercial fishing may be required to continue maximally through a range of fish abundances. The association of target species with stronger relationships between catch and effort is reasonable, given that target species should be taken with greater reliability by knowledgeable fish harvesters.

The biological, geographical, and technological uniqueness of fisheries beyond the documented moderator variables was represented through random- and mixed- effects meta-analysis. Treating individual fisheries as random effects is reasonable when the relationship between catch and effort is expected to be truly different among them. As my results illustrate, each case produced a different slope (there was not just one slope for all the cases) and there was a distribution for all of these slopes, where, using the random-effects model, the mean effect size of the distribution was significantly greater than one. Each fishery study was different, so, not only were there various fishery types, but also various types of effort, in addition to other factors that made each of the data sets extracted from these studies unique. Koricheva et al. (2013) clarify some of these points, explaining that the random variation acknowledged among study results is due to unmeasured factors. It can be accounted for by assuming that the difference between true effect sizes for different studies is because of random variation (due to the unmeasured factors) around the overall mean effect. This mean effect then represents the population of studies. Field (2001) presented a brief tutorial on meta-analytic methods in which he discussed two simulations that compare these methods. He defined fixed- and random-effects models, as well as discussing the statistical differences between them. He thereby illustrated that random effects cannot control type I error in meta-analysis with 15 or fewer studies but can account for that when a meta-analysis contains a large number of studies.

Variation in the measurements of both catch and effort could lead to negative bias in the observed slopes, due to error-in-variables effects (Quinn and Keough 2002) on the estimated relationship. This became visible when comparing the OLS (ordinary least squares) and RMA (reduced major axis) regression slopes. As expected, my results clarify the difference between OLS and RMA; they present OLS with a lower slope estimate in comparison to RMA. As illustrated by Smith (2009), RMA operates on the assumption that the independent variable is also measured with error. The use of two different measures may help in illustrating how these measures can vary, in order to detect any biases that may occur. These biases are especially important to be aware of when employing standard likelihood methods, because they do not consider errors-in-variables. Model II regression theory provides more meaningful results in case of errors in the independent variables. When Model I regression is used in the presence of errors in the independent variables, the value of the parameter will be closer to 0. It has therefore been widely recognized that OLS is not the best fit in all situations, because OLS assumes that error is present only in the dependent variables. The suitable situation for use of OLS is when there is no error in the variables. RMA has been recognized to be an appropriate alternative (Smith 2009).

My results showed that using OLS would have resulted in a completely different answer to my question and changed my interpretation of the results. In fact, the estimation of slope was less than 1 (as shown by the confidence interval) when using the random-effects model and OLS for all studies examined. These results would be interpreted as interference competition. Taking into account error-in-variables, using RMA resulted in a higher estimation of slope that is significantly greater than 1. These results change the interpretation, showing that the reason for disproportionality is more likely due to fishing facilitation or numerical responses among fish harvesters.

In my examination, the correlation between catch and effort was high, as it was expected to be. The result is consistent with an underlying positive relationship between catch and effort; more fishing tends to land more fish. For many fisheries, the fit of the data was examined by analyzing the correlation between catch and effort. Báez et al. (2007) illustrated this in their examination of the data of traditional Spanish drifting longline boats in the Balearic Sea (western Mediterranean). They found that catch correlated to effort by observing that the number of swordfish captured correlated to fishing effort but not to distance to the coast or to depth.

To examine the consistency of the catch-effort relationship among fisheries, I tested heterogeneity among the results of the studies. Heterogeneity will affect the way that I interpret, and draw conclusions about the presence of a general trend among the selected studies. I was able to detect moderate heterogeneity from the effect sizes in these studies, demonstrating variation and supporting my use of a random-effects model. Field (2001) noted that if there were heterogeneity between effect sizes, the results would show the standard errors in a random-effects model to be greater than the standard errors in a fixed-effects model. Koricheva (2013) illustrated that the use of a random-effects model can account for part of the variation in the data.

The relevance of heterogeneity is consistent with the results of previous fishery studies. Goodyear (2003) illustrated that there is both spatial and temporal heterogeneity in the catch rate, which is related to the changes in sea surface temperatures. Moreover, Johnson and Libecap (1982) discussed heterogeneity among fishermen regarding their ability to fish, as was detected from catch data for the bay shrimp season in the fall of 1978. Fishermen's heterogeneity was based on their success over time (higher catch than average), due to differing amounts of knowledge about setting nets and the location of the shrimp.

I have found that the disproportionality observed in the slope of log-log regressions that exceed 1 maybe occur due to the following reasons: numerical response of effort to higher catch and fishing facilitation. To better understand the disproportionality further studies need to consider collecting abundance data in order to provide a clear indication of the relationship between catch and effort.

In this chapter, I used a meta-analysis to determine if there was a consistent or typical relationship across studies between nominal effort and fishing success. I reviewed many sources of published and "grey" literature for the meta-analysis. I cite in my paper the ninety-one of these sources that were suitable for my analysis. The research examined the slope data within each of the ninety-one studies and analyzed the effect size using two different measures. This study identified that fishing catch values were not proportional to nominal effort by using a random-effects model to illustrate the effects on catch. The random-effects model, using OLS for all the studies examined, revealed that the estimation of slope was less than 1 and that the confidence interval was significant, whereas the results of RMA showed a higher estimation of slope and the confidence interval was significantly greater than 1. My findings suggest that the disproportionality may be because of facilitation and numerical responses. Using various moderator variables to check for disproportionality, I found that some of the moderator variables used such as fishery type, and target species were significant. Additional potential statistical causes for these patterns are developed in Chapter 4.

References

Abrahams, M. V., & Healey, M. C. (1993). Some consequences of variation in vessel density: a manipulative field experiment. *Fisheries Research*, *15*(4), 315-322.

Adams, D. C., Gurevitch, J., & Rosenberg, M. S. (1997). Resampling tests for meta-analysis of ecological data. *Ecology*, *78*(4), 1277-1283.

Alexander, R. A. (1990). A note on averaging correlations. *Bulletin of the Psychonomic Society*, *28*(4), 335-336.

Arnqvist, G., & Wooster, D. (1995). Meta-analysis: synthesizing research findings in ecology and evolution. *Trends in Ecology & Evolution*, *10*(6), 236-240.

Báez, J. C., Real, R., García-Soto, C., De la Serna, J. M., Macías, D., & Camiñas, J. A. (2007). Loggerhead turtle by-catch depends on distance to the coast, independent of fishing effort: implications for conservation and fisheries management. *Marine Ecology Progress Series*, *338*, 249-256.

Bender, D. J., Contreras, T. A., & Fahrig, L. (1998). Habitat loss and population decline: a meta-analysis of the patch size effect. *Ecology*, *79*(2), 517-533.

Côté, I. M., Mosqueira, I., & Reynolds, J. D. (2001). Effects of marine reserve characteristics on the protection of fish populations: a meta ‐ analysis. *Journal of Fish Biology*, *59*(sA), 178-189.

Deng, R., Dichmont, C., Milton, D., Haywood, M., Vance, D., Hall, N., & Die, D. (2005). Can vessel monitoring system data also be used to study trawling intensity and population depletion? The example of Australia's northern prawn fishery. *Canadian Journal of Fisheries and Aquatic Sciences*, *62*(3), 611-622.

Evans, L., Cherrett, N., & Pemsl, D. (2011). Assessing the impact of fisheries co-management interventions in developing countries: A meta-analysis. *Journal of environmental management*, *92*(8), 1938-1949.

Field, A. P. (2001). Meta-analysis of correlation coefficients: a Monte Carlo comparison of fixed- and random-effects methods. *Psychological methods*, *6*(2), 161.

Gillis, D. M. (2003). Ideal free distributions in fleet dynamics: a behavioral perspective on vessel movement in fisheries analysis. *Canadian Journal of Zoology*, *81*(2), 177-187.

Gillis, D. M., & Peterman, R. M. (1998). Implications of interference among fishing vessels and the ideal free distribution to the interpretation of CPUE. *Canadian Journal of Fisheries and Aquatic Sciences*, *55*(1), 37-46.

Glass, G. V. (1976). Primary, secondary, and meta-analysis of research. *Educational researcher*, 3-8.

Goodyear, C. P. (2003). Spatio-temporal distribution of longline catch per unit effort, sea surface temperature and Atlantic marlin. *Marine and freshwater research*, *54*(4), 409-417.

Gulland, J.A. (1956). On the fishing effort in English demersal fisheries. *Fish. Invest. London*, *20* (5), series 2, 1±41.

Gurevitch, J., Curtis, P. S., & Jones, M. H. (2001). Meta-analysis in ecology. *Advances in Ecological Research*, *32*, 199-247.

Gurevitch, J., Morrow, L. L., Wallace, A., & Walsh, J. S. (1992). A meta-analysis of competition in field experiments. *American Naturalist*, 539-572.

Harrison, F. (2011). Getting started with meta-analysis. *Methods in Ecology and Evolution*, *2*(1), 1-10.

Higgins, J., & Thompson, S. G. (2002). Quantifying heterogeneity in a meta-analysis. *Statistics in medicine*, *21*(11), 1539-1558.

Higgins, J., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *Bmj*, *327*(7414), 557-560.

Hilborn, R. and M. Ledbetter (1979). Analysis of the British Columbia salmon purse-seine fleet: dynamics of movement. *J. Fish. Res. Bd. Can.*, *36*, 384-391.

Johnson, R. N., & Libecap, G. D. (1982). Contracting problems and regulation: the case of the fishery. *The American Economic Review*, 1005-1022.

Konstantopoulos, S. (2006). Fixed and mixed effects models in meta-analysis (No. 2198). IZA Discussion Papers.

Koricheva, J., Gurevitch, J., & Mengersen, K. (eds.) (2013). *Handbook of meta-analysis in ecology and evolution*. Princeton University Press.

Koricheva, J. (2002). Meta-analysis of sources of variation in fitness costs of plant antiherbivore defenses. *Ecology*, *83*(1), 176-190.

Lapointe, M. F. (1989). *Numerical and functional responses of British Columbia trawler fishermen* (Doctoral dissertation, M.Sc. thesis, University of British Columbia, Vancouver).

Ledbetter, M. (1986). *Competition and information among British Columbia salmon purse seiners*. Max Ledbetter.

Maunder, M. N., & Punt, A. E. (2004). Standardizing catch and effort data: a review of recent approaches. *Fisheries Research*, *70*(2), 141-159.

Maunder, M. N., Sibert, J. R., Fonteneau, A., Hampton, J., Kleiber, P., & Harley, S. J. (2006). Interpreting catch per unit effort data to assess the status of individual stocks and communities. *ICES Journal of Marine Science: Journal du Conseil*, *63*(8), 1373-1385.

McConnell, K. E., Strand, I. E., & Blake-Hedges, L. (1995). Random Utility Models of Recreational Fishing: Catching Fish Using a Poisson-Process. *Marine Resource Economics*, *10*, 247-262.

Methot, R. D., & Wetzel, C. R. (2013). Stock synthesis: A biological and statistical framework for fish stock assessment and fishery management. *Fisheries Research*, *142*, 86-99.

Millington, P. (1984). The fishermen as predator: numerical responses of British Columbia gillnet fishermen to salmon abundance.

Orbach, M. K. (1977). *Hunters, seamen, and entrepreneurs: the tuna seinermen of San Diego*. Univ. of California Press.

Poos, J. J., & Rijnsdorp, A. D. (2007). An "experiment" on effort allocation of fishing vessels: the role of interference competition and area specialization. *Canadian Journal of Fisheries and Aquatic Sciences*, *64*(2), 304-313.

Post, J. R., Persson, L., Parkinson, E. V., & Kooten, T. V. (2008). Angler numerical response across landscapes and the collapse of freshwater fisheries. *Ecological Applications*, *18*(4), 1038-1049.

Post, J. R., Sullivan, M., Cox, S., Lester, N. P., Walters, C. J., Parkinson, E. A. & Shuter, B. J. (2002). Canada's recreational fisheries: the invisible collapse? *Fisheries*, *27*(1), 6-17.

Price, T. D., & Phillimore, A. B. (2007). Reduced major axis regression and the island rule. *Journal of Biogeography*, *34*(11), 1998-1999.

Quinn, G. P., & Keough, M. J. (2002). *Experimental design and data analysis for biologists*. Cambridge, UK: Cambridge University Press.

R Core Team (2014). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org.

Richards, L. J., & Schnute, J. T. (1986). An experimental and statistical approach to the question: is CPUE an index of abundance? *Canadian Journal of Fisheries and Aquatic Sciences*, *43*(6), 1214-1227.

Silver, N. C., & Dunlap, W. P. (1987). Averaging correlation coefficients: should Fisher's z transformation be used? *Journal of Applied Psychology*, *72*(1), 146.

Smith, R. J. (2009). Use and misuse of the reduced major axis for line-fitting. *American Journal of Physical Anthropology*, *140*(3), 476-486.

Stillman, R. A., Goss-Custard, J. D., & Caldow, R. W. G. (1997). Modelling interference from basic foraging behaviour. *Journal of Animal Ecology*, 692-703.

Sokal, R. R., & Rohlf, F. J. (1995). Biometry: the principles and practice of statistics in biological research. *WH. Freeman& Co., San Francisco. SokalBiometry: the principles and practice of statistics in biological research, 1995*.

Solomon, M. E. (1949). The natural control of animal populations. *The Journal of Animal Ecology*, 1-35.

Stuster, J. (1978). Where 'Mabel' May Mean 'Sea Bass'. *Natural History, 87*(9), 65-71.

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, *36*(3), 1-48.

Worm, B., & Myers, R. A. (2003). Meta-analysis of cod-shrimp interactions reveals top down control in oceanic food webs. *Ecology*, *84*(1), 162-173.

Chapter Three

The connection between meta-analysis and the application of GLM and GLMM to catch standardization: using nominal effort in the case of fixed and mobile gears.

Introduction

The meta-analysis used in chapter two combined the results of time series data from various studies. This provided a general idea about the relationship between catch and effort. The data used in the meta-analysis were combined across studies, though there were various fishery types, such as artisanal, commercial, and research. The effort measure used, such as hours, sets, trips, or hooks, was different for each fishery. Different gear was also used by each fishery, such as gillnet, trawl, electrofishing, purse seine, pole and line, long line, or a mix of various types. The species analyzed were tuna, such as bigeye tuna (*Thunnus obesus*), shark, such as blue shark (*Prionace glauca*), shrimp, such as brown tiger prawns (*Penaeus esculentus*), redfish (*Sebastes mentella*), billfish, such as swordfish (*Xiphias gladius*), salmon, such as chinook salmon (*Oncorhynchus tshawytsch*), haddock (*Melanogrammus aeglefinus*), herring, such as Atlantic herring (*Clupea harengus*), cod *(Gadus morhua)*, and Nile perch *(Lates niloticus)*. The general results from the meta-analysis may differ from the results of one fishery. A specific fishery can give a close look at the relationship between catch and effort. This will reveal any differences between the slope produced by meta-analysis and the slope produced with more highly resolved data from a specific fishery. The results from meta-analysis using a mixed-effects model with RMA favored proportionately higher catches with increasing effort. This may be because of a number of reasons, such as facilitation (Healey et al. 1990) and numerical responses (Millington 1984). The forest plot (Figure 2.3) illustrates that each fishery had a different slope, but these slopes varied significantly (indicated by measures of heterogeneity) around a value close to 1. This suggests that we must look individually at specific fisheries whose relationships between catch and effort may differ from what I have generally found with meta-analysis. If the specific fishery results of slope are greater than 1, it will be similar to the one obtained via meta-analysis.

This would suggest that facilitation and numerical responses maybe behind this disproportionality. If the results show that the slope is 1, then this would agree with the relationship assumed when using catch per unit effort (CPUE) and seen in results like those of Bishop et al. (2004). Catch would be proportional to nominal effort and, as effort increases, catch should increase in the same proportion. However, if the results show that the slope is less than 1, then this would suggest interference competition (Poos and Rijnsdorp 2007), gear saturation (Dauk and Schwarz 2001), or reasons related to other aspects of fishermen's behaviour, such as intending to fill the net (Salas and Gaertner 2004).

The meta-analysis was done using both ordinary least squares (OLS) and reduced major axis (RMA) regression. Linear regression is commonly used in fisheries and ecological studies. However, the parameters estimated using linear regression assume that model error is solely associated with fishery data, due to its having many problems, such as outliers, zeros, etc. The problem with OLS regression is that it may bias the results due to any errors in the observed values of the predictor variables. If there is heterogeneity in the error variance, errors not normally distributed, outliers are present, or errors exist in the predictor variables, then the estimated results from the model using ordinary least square regression must be biased (Hilborn and Walters 1992, Chen and Jackson 2000). Some of these problems can be solved by using a nonlinear model (Chen and Jackson 2000, Motulsky and Christopoulos 2004). The use of a GLM and a GLMM will fit the heteroscedastic and non-linear data more accurately than the use of simple linear regression. The resulting parameter estimates performed with a GLM and a GLMM should be more accurate. A GLMM can resolve issues such as autocorrelation, and nonlinearity in the relationship (Breslow and Clayton 1993, Carl al. 2008). A number of the analyses conducted on fisheries have either ignored vessels or treated them as fixed effects by using simple statistical techniques, such as linear regression. However, each vessel operates uniquely and each fisherman has his own ideas about and goals in fishing. This should be taken into consideration in the analysis (Salas and Gaertner 2004). To account for this, vessels will be treated as random effects, using a generalized linear mixed-effects model. This will account for variation in the parameter estimates caused by individual variation among vessels.

References

Bishop, J., Venables, W. N., & Wang, Y. G. (2004). Analysing commercial catch and effort data from a penaeid trawl fishery: a comparison of linear models, mixed models, and generalised estimating equations approaches. *Fisheries Research*, *70*(2), 179-193.

Breslow, N. E., & Clayton, D. G. (1993). Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association*, *88*(421), 9-25.

Carl, G., Dormann, C. F., & Kühn, I. (2008). A wavelet-based method to remove spatial autocorrelation in the analysis of species distributional data. *Web Ecology*, *8*(1), 22-29.

Chen, Y., & Jackson, D. A. (2000). An empirical study on estimators for linear regression analyses in fisheries and ecology. *Fisheries research*, *49*(2), 193-206.

Dauk, P. C., & Schwarz, C. J. (2001). Catch estimation in the presence of declining catch rate due to gear saturation. *Biometrics*, 287-293.

Healey, M. C., Thomson, R. E., & Morris, J. F. (1990). Distribution of commercial troll fishing vessels off southwest Vancouver Island in relation to fishing success and oceanic water properties and circulation. *Canadian Journal of Fisheries and Aquatic Sciences*, *47*(10), 1846-1864.

Hilborn, R., & Walters, C. J. (1992). Quantitative fisheries stock assessment: choice, dynamics and uncertainty. *Reviews in Fish Biology and Fisheries*, *2*(2), 177-178.

Millington, P. (1984). The fishermen as predator: numerical responses of British Columbia gillnet fishermen to salmon abundance.

Motulsky, H., & Christopoulos, A. (2004). *Fitting models to biological data using linear and nonlinear regression: a practical guide to curve fitting*. Oxford University Press.

Poos, J. J., & Rijnsdorp, A. D. (2007). An "experiment" on effort allocation of fishing vessels: the role of interference competition and area specialization. *Canadian Journal of Fisheries and Aquatic Sciences*, *64*(2), 304-313.

Salas, S., & Gaertner, D. (2004). The behavioural dynamics of fishers: management implications. *Fish and fisheries*, *5*(2), 153-167.

Chapter Four

Applying GLM and GLMM to catch standardization: using nominal effort in the case of deployments of fixed and mobile gear

Abstract

This chapter examines the relationship between catch and effort data in two different fisheries: the Gulf of St. Lawrence snow crab fishery (fixed gear) and the haddock/redfish trawl fishery of southwest Nova Scotia (mobile gear). The purpose of this analysis is to examine the proportionality between catch and nominal effort at the scale of individual sets (traps or trawls). This was conducted using a generalized linear mixed-effects model to account for variation among fishing vessels. Proportionality was explored through of the exponent of effort. The disproportionality indicated by this analysis, i.e. catches increasing less than proportionally to effort, was expected to be from interference competition, exploitation competition, and gear saturation in both fisheries examined. Data aggregated by month and year resulted in catch increasing disproportionately faster than effort, which could result from facilitation, numerical responses, or aggregation biases. This establishes that similar disproportional relationships can exist between nominal effort and the resulting catch in both fixed and mobile gear fisheries, contrary to the implicit assumptions underlying the use of CPUE as a fisheries indicator of abundance.

Key words: Generalized linear mixed-effects model (GLMM), catch, effort, proportionality, aggregation.

Introduction

Catch per unit effort (CPUE) data from commercial fisheries are commonly standardized to deal with spatial and temporal patterns or vessel characteristics that could confound its interpretation as an index of fish abundance (Goñi et al. 1999, Punt et al. 2000, Ye et al. 2001). However, standardization alone will not solve all of the problems present in the data. A GLM (Maunder and Punt 2004) can treat the covariates, such as year, month, gear type, etc., as fixed-effects but does not consider unique differences among vessels that are not captured in general measures such as size or horsepower. The "skipper effect" is a known example of such a difference (Marchal et al. 2006). A GLMM would treat the fixed covariates in the same manner as a GLM, it could also treat vessels as random effects (Helser et al. 2004), allowing some parameters' values to vary among vessels when specific covariates cannot be easily identified. A general overview of GLMM use can be found in McCulloch and Neuhaus (2001), Breslow and Clayton (1993), and Bolker et al. (2009).

The use of standardized CPUE as an index of abundance implies that if fishing effort increases, the catch size will increase in proportion. Thus, we can expect that, if fishermen double their effort, they will get twice the catch: a simple linear relationship. However, this may not occur for several reasons, including interference competition, exploitation competition, gear saturation, facilitation, and numerical responses. There can be competition for good fishing spots, which can produce aggregation of vessels and create interference between them. If the relationship between catch and effort is disproportional, where catch rate decreases in relation to increased effort, it may be a result of interference competition among these vessels. Poos and Rijnsdorp (2007) did an experiment on effort allocation of fishing vessels to examine interference competition and found that, even when effort was doubled, there was a decrease of

14% in the catch rate because of interference competition. A numbers of studies have discussed or shown interference competition among vessels (Healey and Morris 1992, Gillis and Peterman 1998, Rijnsdorp et al. 2000). Catch rate may decrease through exploitation competition because of the resultant reduction of the fisheries' resources (Healey et al. 1990). Rijnsdorp (2000) studied exploitation competition in an area heavily fished. The catch rate was reduced by 10% within 48 hours. Vessels with weaker engines had a higher decline in catch rate than did vessels with powerful ones. This is evidence of interference competition among vessels. Catch rate may increase rapidly with an increase in effort because of fishing facilitation, where fishermen share information on where to locate fish using various methods, such as radio, radar, and word of mouth (Healey et al. 1990). Facilitation may occur by communication in response to catch availability or in an attempt to minimize hazardous situations by communication about local risks associated with fishing (Salas and Gaertner 2004). Numerical responses may occur in response to local fish aggregation, attracting vessels to aggregate in the same area. Millington (1984) examined vessel aggregation in response to fish density in the British Columbia salmon gillnet fleet from 1979 to 1981. He found that there is a strong relationship between catch value per week per boat and the numbers of gillnet boats fishing in the following week. Gear saturation occurs when the gear used to fish becomes full, which is one of the reasons for a decline in catch rate (Dauk and Schwarz 2001). The presence of multiple species in an area can exacerbate gear saturation when the net fills up with species other than the one targeted, which, in turn, decreases the catch rate of the target species (Walters 2003). Longlines reach gear saturation quickly because of the limit on the number of hooks (Skud 1978). Trap saturation also causes a decrease in catch rate (Miller 1979). Fishermen's behaviour may also affect catch rate, for example, if the goal of a fisherman is to fill his net, or in the case of a small vessel that may be fishing to cover its travel costs (Salas and Gaertner 2004). If a fisherman intends to fill a number of nets from one area, the time required to fill each net will increase as abundance decreases gradually in the area.

In this chapter, I will focus on catch standardization using generalized linear mixed-effects models (GLMM). I will apply these models to fixed-gear (the Gulf of St. Lawrence snow crab fleet trap fishery data) and mobile-gear data (from a trawl fishery in southwest Nova Scotia). This will be done in an attempt to answer the fundamental question "is catch proportional to nominal effort?" in the context of a detailed analysis (set-by-set) of catches from specific commercial fisheries (Maravelias et al. 1996, Helser et al. 2004).

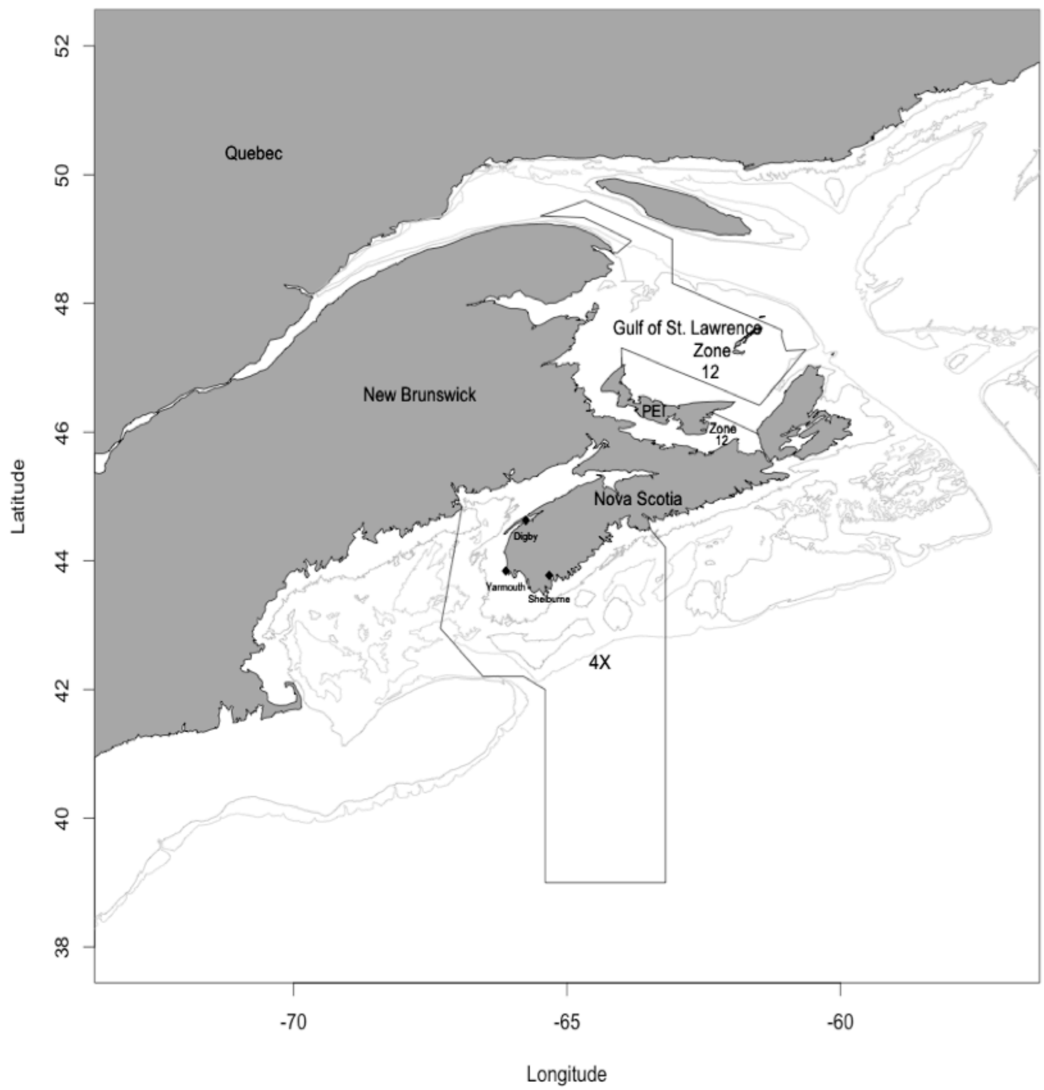
Methods

Two fisheries provided data for this study, one representing fixed and the other mobile fishing gear. Data on mixed species catches from the trawl fishery of southwest Nova Scotia (mobile gear) were provided by colleagues at the Bedford institution of Oceanography (Bedford Institute of Oceanography, Dartmouth, Nova Scotia). I focused on the otter trawl, the most common type of gear used by this fishery. This gear works by being dragged along the seafloor, where it captures most of the fish in its way (Engel and Kvitek 1998). The trawl data analyzed were limited to NAFO, Division 4X for the years of 2008 to 2013 in the Northwest Atlantic Fisheries, the southwestern Scotian Shelf, and the Bay of Fundy (Figure 4.1). Species landed in these areas included haddock (*Melanogrammus aeglefinus*), cod (*Gadus morhua*), halibut (*Reinhardtius hippoglossoides*), redfish (*Sebastes mentella*), pollock (*Pollachius virens L.*), red hake (*Urophycis chuss*), silver hake (*Merluccius bilinearis*), white hake (*Urophycis tenuis*), and winter flounder (*Pleuronectes americanus*).

In contrast, the data provided by my colleagues at the Gulf Fisheries Centre (Moncton, New Brunswick) from the snow crab trap fishery of the Gulf of St. Lawrence pertains to the fishing of a single species. The snow crab (*Chionoecetes opilio*) fishery is one of the dominant and most valuable fisheries in Atlantic Canada (Conan and Comeau 1986, Lovrich and Sainte-Marie 1997, Hébert et al. 2001). The fishery operates in the Gulf of St. Lawrence, around Cape Breton Island, on the Scotian Shelf, on the Labrador Shelf, and in the bays of Newfoundland (Miller 1976). The fishing season is between April and July. Only males with a carapace *(cephalothorax)* of a minimum size of 95 mm wide are allowed to be harvest in the fishery (Hébert et al. 2001). Watson (1970) examined the maturity of male snow crabs from the Gulf of St. Lawrence and found that, out of a sample size of 194, 50% were mature already with a size of 57 mm and all males with a size above 72 mm were mature. It is prohibited for fisheries to harvest female and soft-shelled crabs (Miller, 1976). This prohibition is meant to protect the species from exploitation and has been in enforcement since 1973 (Conan and Comeau 1986, Elner and Beninger 1992). Traps are the only type of fishing gear allowed in this fishery. Any boat entering the fishery is obligated to have a license for crab fishing and to support the fishery with data collection (Miller 1976).

The response variables examined in the trawl data analysis included the total value of fish landed by the trawl and the weight of haddock (29.27% of total landed weight), cod (6.32%), redfish (33.12%), and pollock (20.84%), which together account for 89.6% of the landed weight and 81.97% of the landed value. Values were calculated using price data publically available from the Prince Edward Island Department of Agriculture and Fisheries (http://www.gov.pe.ca/fard/index.php3?number=1024862). The analysis began by creating monthly values from the weekly values for each species. At first, the weekly values (price) of

Figure 4.1: The map illustrates the fishing area of the trawl fishery in Southwest Nova Scotia, NAFO, division 4X. The ports of Digby, Yarmouth, and Shelburne are identified on the west coast of Nova Scotia. It illustrates the fishing area for snow crab (*Chionoecetes opilio*). The fishery is in the Gulf of St. Lawrence, in Zone 12.



dressed fishes (fish whose fin, tail and head have been removed) were converted to the values of rounded fishes (whole fish) by multiplying by the conversion factor. The conversion factor used was defined by species and area (FAO 1980). For each species, the prices used were from 4X ports only (Yarmouth, Digby and Shelburne (Figure 4.1)). If there was more than one relevant port for a species, the average price was calculated. Prices were then aggregated to calculate a monthly price by averaging weekly values. Missing monthly values were imputed using data from the previous month. These same steps were taken for all six years and the results were combined to calculate monthly prices for the whole period. The monthly values created for each species were then matched to the data by species name, year, and month, in order to create values for each row in the data. Species that did not have values (in the weekly fish price reports) were removed from the analysis. The percentage of removed species was less than 1% (0.009%) of the total weight.

The next step was to aggregate the data, set-by-set, by month, and by year. A set is defined as a trawl in the 4X data and a trap in the snow crab data. The weight for each species was aggregated based on the set ID to create set-by-set data. Monthly data was aggregated to contain the monthly effort and catch of each vessel. The greatest level of aggregation was by year for each vessel. I removed sets greater than 6 hours, which are less than 0.02% of total recorded hours and choose to work with active vessels that fished more than 100 sets per year. The most extreme 5% of the catch weight or values were removed to reduce the impact of inaccuracies and atypical records from the overall trends studied in subsequent analysis.

After data selection and aggregation, I conducted data exploration. I followed Zuur's 2010 protocol (Zuur et al. 2010) by examining the response variables (catch and value of each species), and the predictor variables: length of vessel, tonnage, hours, year, month and week. I explored outliers, normality, excess zeros, and autocorrelation in the responses. I examined outliers, co-linearity, and possible interactions in the predictor variables. Finally, graphically I examined potential patterns in the relationships between responses and predictors. The same steps were taken before the analysis of data aggregated by year and by month. Preliminary analysis with a simple generalized linear model (GLM) produced a quantile-quantile plot that was not linear; it was skewed on high and small values. Potential individual variation in vessel performance led to the use of generalized linear mixed models (GLMM). The initial GLMM displayed severe autocorrelation in its residuals, so I attempted to reduce this by analyzing a subset of the full data (Dormann et al. 2007, Beguería and Pueyo 2009). The data were ordered and a random selection of one set per vessel per day was used to reduce the autocorrelation.

The examined data from the Gulf of St. Lawrence Snow Crab fishery were restricted to zone 12 (Figure 4.1) (Hébert et al. 2001) from 2006 to 2009. The covariates available included landing date, soak time (how long the trap was in the water), the location of the traps (longitude and latitude), effort (the number of traps on a line), and day of the year fished. The analysis focused on zone 12 because it covers the majority of the Gulf of St. Lawrence and the fishery based in New Brunswick and Prince Edward Island. The response variable examined in the snow crab data analysis was the weight of the snow crab. Zero-catches were dropped, which was five records out of the 14569 (0.0003%). The records greater than one week of soak time were then removed. These were 338 records out of 14564 records (0.02% of the total soak time). To keep the numbers of traps at 150 or less, 0.004% of the total traps, 55 records out of 14226, were removed. I then created a subset of the most active fishermen, whose minimum number of landing records for each year was 10 because their performance would more accurately reflect the interannual changes in abundance rather than less active fishers who were not evenly represented throughout the data. I then ordered the data by vessel, year, month, and day. In this, I followed Zuur's general 2010 protocol for data exploration (Zuur et al. 2010). I aggregated the data set-by-set, by month, and by year and randomly selected one record per fisherman per day similar to the 4X trawl data). The final subset contained 46.18% of the total landed weight.

I used a GLMM with a gamma distributed response for my analyses because the distribution of the catch data was strictly positive, skewed, and distinctly non-normal. A GLMM contains the same three components as a GLM: (1) a randomly distributed response; (2) the linear predictor; and (3) the link function. The observed value of the response (C in equation (2.1)) has an independent distribution, which is a member of the exponential family of distributions. This family includes the normal, binomial, and Poisson, or in my case the gamma distribution (McCullagh and Nelder 1989, Bishop et al. 2004). Gamma distributions are commonly used to represent natural variables that must be greater than zero and display skew similar to that observed in log-normal distributions. I modeled catch weight (C_{ij}) using the gamma distribution; i is the catch observation and j is the vessel I.D. The variance of catch weight is $(\phi \cdot \mu_{ij})$, where μ_{ij} is the mean of the catch weight and ϕ (also parameterized as ν^{-1}) is the dispersion parameter of the gamma distribution (Zuur et al 2009). The second component is the linear predictor, as shown in equation (4.1):

$$\eta_{ij} = log_e\beta_0 + \beta_1 log_e f_i + a_j \tag{4.1}$$

$$a_j \sim N(0, \sigma_j^2)$$

where η is the linear predictor, β_0 is the intercept, β_1 is the effort coefficient, and a_j is the random effect for vessel j. The third component is the link function. In general, the link function works by defining the relationship between the linear predictor η and the mean of the distribution μ_{ij} (Venables and Dichmont 2004). I used the log link function to represent the nonlinear relationships in the model. Thus, logging μ_{ij} will equal the value of the linear predictor shown in equation (4.2), which can also be expressed as (4.3).

$$log_e\ \mu_{ij} = \eta \tag{4.2}$$

$$\mu_{ij} = e^{\eta} \tag{4.3}$$

In my study, I raised the effort to a power to allow for nonlinearity. I used the log link in equation (4.2) to transform the predictor variables, which will express μ_i as:

$$\mu_{ij} = e^{log_e \beta_0 + \beta_1\, log_e\, f_i + a_j} \tag{4.4}$$

Using the algebraic properties of logarithms I can convert equation (4.4) to a multiplicative expression and simplify, which will give us equation (4.5).

$$\mu_{ij} = \beta_0 \cdot f^{\beta_1} \cdot e^{a_j} \tag{4.5}$$

This equation allows us to represent nonlinear relationships by raising the explanatory variables to powers.

The GLMM model consists of two parts: a fixed part, and a random part. The first part of equation (4.1), which containsβ_0, the intercept, and β, the coefficient of the explanatory variable, comprises the fixed part, while a_j is the random part. a_j changes the intercept by randomly changing the vessels fished. I used the R statistical programming language (R Core Team 2014) to fit the data to a generalized linear mixed-effects model (GLMM) by using the lme4 package. Model selection was performed using likelihood ratio tests, which compares two nested models (Vuong 1989).

Results

The exploratory analysis in both fisheries showed that catch value and weight were skewed to the right, with some extreme values. Variance appeared similar for both responses at all levels of data aggregation examined. The histograms of 4X species' were not normally distributed. The tests for autocorrelation were significant for both the response variables. Covariates were also examined. Checking for the existence of a correlation between covariates showed that the length of vessel overall and the gross registered tonnage (between 10 to 665 tons) were highly correlated. I choose to work with length of vessel to represent vessel size. Month and week were obviously correlated so I choose to work with month as a measure of time aggregated within years as it is often used in fisheries reports. Graphically examining patterns among covariates showed no obvious interactions.

The GLMM of the 4X trawl fishery in southwest Nova Scotia (haddock, mobile gear) is presented in Table 4.1. There are 25 vessels, which are represented by the random factor, and 2109 observations of haddock. The random effects (vessels' I.D.) are assumed to be normally distributed, with a mean of 0 and a variance of 1.837. This value indicates that there is little variability among the intercepts of the vessels in comparison to the overall residual variance (6.10, Table 4.1). The coefficient of the log of hours fished, representing the exponent of effort in equation 4.5, was significantly less than 1 (H_0: slope = 1, $t = -14.42$, $df = 4600$, $p < 0.0001$). This indicates disproportionality in the relationship between catch and effort. The coefficients for year and month mere mostly significant showing how catch changes among years and within a year (Table 4.1). The outcomes of the Gulf of St. Lawrence model (snow crab, fixed gear) shows that coefficient of log traps was significant and less than 1, showing disproportionality in the relationship between catch and effort similar to the trawl fishery. Log soak, year, and week of the year were all significant showing that catch changes depending on how long the trap was in the water, among years, and within a year. The 69 snow crab vessels are represented by the random factor. There are 4029 observations of snow crab catch. The random effects (vessels' I.D.) are assumed to be normally distributed, with a mean of 0 and a variance of 0.0181. This variance suggests that the lines representing the linear predictor are close together (small differences in the intercept) suggesting that there is little variation among vessels in their ability to catch snow crab (Table 4.2). Using a GLMM with a random selection of one set per vessel per day both the trawl and trap fishery helped to reduce the autocorrelation in the data (Figure 4.2).

The relationship between catch and effort varied among species and the level of data aggregation. The set-by-set analysis for the 4X data using GLMM (Table 4.3) showed that catch decelerated as effort increased in the case of haddock, cod, and 4X species' value. The slope is barely different than 0. On the other hand, catch decreased when effort increased in the case of redfish and pollock, as shown by a negative slope. The set-by-set analysis of the snow crab data shows catch increasing closer to one, but still significantly lower. The data aggregation has a significant impact on the slope as compared to analyzing the data set-by-set. As Table 4.3 illustrates, the aggregation by month of all the data sets for each vessel shows that the relationship between catch and effort changes. The slope under these conditions moved closer to 1 in the case of haddock, pollock, cod, and 4X species' value, but was greater than 1 for redfish. The snow crab data analysis continues to show catch increasing almost as much as effort; the slope is almost indistinguishable from 1 (upper confidence interval = 0.9932). The aggregation by year of all the data sets for each vessel results in a slope is greater than the set-by-set, but not much different from the analysis of monthly data. The slope from the snow crab data analysis shows the greatest deviation from the monthly analysis; however this is suspect because it is based on only the four years of observations. Most importantly for snow crab, the confidence intervals for the coefficient of nominal effort effect did not include one for all set-by-set analyses and for the monthly aggregations of data as well. This shows that the relationship between catch and effort can easily be disproportionate, and especially when analyses are based on disaggregated data.

Table 4.1: The outcomes of the 4X trawl fishery's haddock model using a generalized linear mixed model fit using data from individual trawls (sets). January 2008 is the reference month and year of the model contrasts. The coefficients are on the scale of the linear predictor. The response variable is catch weight in kilograms. n = 4620 for 25 vessels.

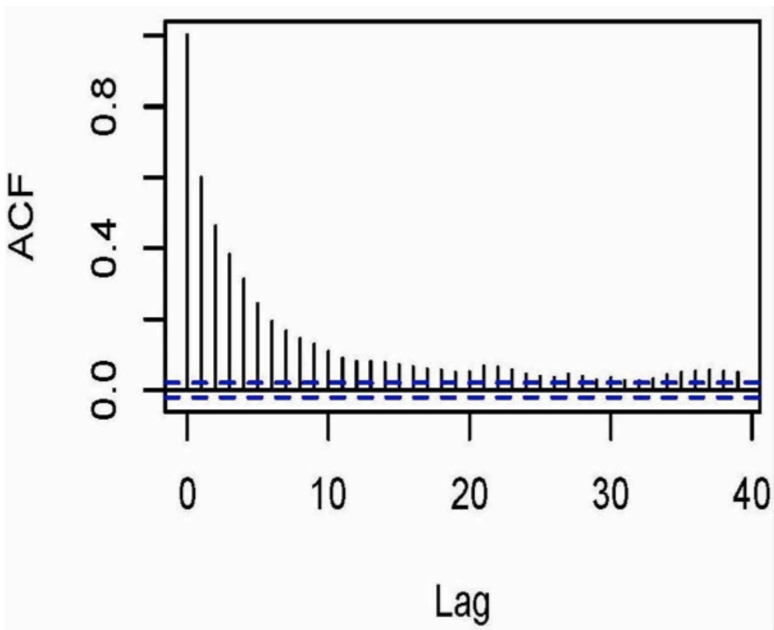
4X trawl fishery in southwest Nova Scotia, Haddock (mobile gear)				
Random effects				
Groups	Name	Variance		
Vessels' I.D.	Intercept	1.837		
Residual		6.100		
Fixed effects				
Covariates	Estimate	Standard error	t-value	p-value
Intercept	6.2035	0.1900	32.64	< 0.0001
Log hours	0.1130	0.0615	1.84	0.0663
Year 2009	0.3370	0.1002	3.36	0.0008
Year 2010	0.0437	0.0981	0.45	0.6558
Year 2011	-0.2803	0.0983	-2.85	0.0044
Year 2012	-0.5284	0.0911	-5.8	< 0.0001
Year 2013	-0.4795	0.0966	-4.96	< 0.0001
February	0.0817	0.1519	0.54	0.5906
March	-0.6397	0.1418	-4.51	< 0.0001
April	-1.9294	0.1449	-13.31	< 0.0001
May	-2.1067	0.1443	-14.6	< 0.0001
June	-1.4433	0.1507	-9.57	< 0.0001
July	-1.1022	0.1530	-7.2	< 0.0001
August	-1.2137	0.1572	-7.72	< 0.0001
September	-0.6841	0.1627	-4.2	< 0.0001
October	-0.6031	0.1610	-3.75	0.0002
November	-0.4481	0.1528	-2.93	0.0034
December	-0.4858	0.1651	-2.94	0.0033

Table 4.2: The outcomes of the Gulf of St. Lawrence snow crab data using a generalized linear mixed model fit using maximum-likelihood and gamma with log link function and set-by-set data. The coefficient estimates are provided on the scale of the linear predictor. The response variable is catch weight in kg. The AIC = 67592.3 and n = 4029.

Gulf of St. Lawrence snow crab (fixed gear)				
Random effects				
Groups	Name	Variance		
Vessels' I.D.	Intercept	0.0181		
Residual		0.2206		
Fixed effects				
Covariates	Estimate	Standard error	t-value	p-value
Intercept	5.2805	0.0766	68.91	< 0.0001
Log traps	0.7721	0.0166	46.51	< 0.0001
Log soak	0.0381	0.0071	5.37	< 0.0001
Year	-0.0479	0.0072	-6.64	< 0.0001
Week of year	-0.0448	0.0030	-14.82	< 0.0001

Figure 4.2: Comparison of the autocorrelation between GLMM models (haddock) fit with (a) all data and (b) a subset of single randomly selected sets (one per vessel per day)

(a)



(b)

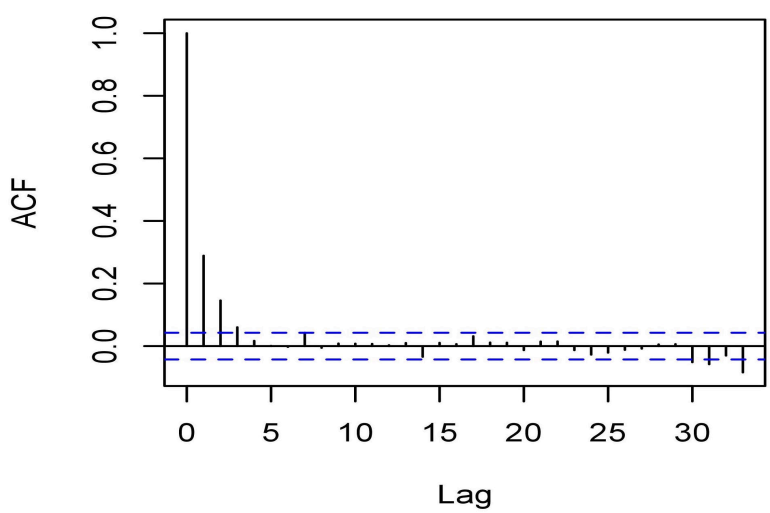


Table 4.3: The outcomes of models using set-by-set, and data aggregated by month and by year, using a GLMM. The coefficient estimates are provided on the scale of the linear predictor. The 95% confidence intervals for the estimates are provided in parentheses. The covariates used are $\log_{10}$ effort (hours or traps), year, and month as fixed effects, and vessel as a random effect. For aggregation by year month and week were not covariates. The response variables are catch weights or total landed value.

Species	Criteria	n	Intercept, upper and lower CI	Log_e hours (slope), upper and lower CI
Cod	Set	4442	4.9090, (4.63275.1853)	0.2057, (0.1183, 0.2931)
	Monthly	1115	4.7639, (4.3036, 5.2243)	0.7419, (0.6719, 0.8117)
	Yearly	149	5.7623, (4.6706, 6.8540)	0.6779, (0.5093, 0.8465)
Haddock	Set	4620	6.2036, (5.8311, 6.5760)	0.1130, (-0.0076, 0.2337)
	Monthly	1127	6.0528, (5.5750, 6.5305)	0.7255, (0.6613, 0.7898)
	Yearly	149	6.8148, (5.6319, 7.9975)	0.6775, (0.5047, 0.8502)
Pollock	Set	4045	5.6562, (5.2245, 6.0879)	-0.0632, (-0.1813,0.0549)
	Monthly	1095	4.7866, (4.1667, 5.4064)	0.8669, (0.7864, 0.9473)
	Yearly	149	6.5867, (5.4285, 7.7449)	0.6685, (0.4894, 0.8477)
Redfish	Set	3479	6.5073, (6.0043, 7.0104)	-0.7708, (-0.8819, -0.6597)
	Monthly	986	3.8048, (3.0769, 4.5326)	1.2059, (1.1170, 1.2949)
	Yearly	150	2.9233, (2.3027, 3.5439)	1.2532, (1.2356, 1.2707)
Value	Set	3434	7.6958, (7.4831, 7.9086)	0.1223, (0.0482, 0.1964)
	Monthly	1151	7.8754, (7.4773, 8.2735)	0.5222, (0.4694, 0.5749)
	Yearly	150	6.9341, (6.1071, 7.7612)	0.7011, (0.6027, 0.7993)

Table 4.3 (continued)

Species	Criteria	n	Intercept, upper and lower CI	Log_e traps, upper and lower CI
Snow crab	Set	4029	5.2805, (5.1303, 5.4307)	0.7721, (0.7396, 0.8046)
	Monthly	689	4.3328, (4.0941, 4.5716)	0.9558, (0.9185, 0.9932)
	Yearly	276	8.1719, (7.6696, 8.6741)	0.4504, (0.3822, 0.5186)

Discussion

The relationship between catch and nominal effort is strongly affected by the level of data aggregation prior to analyses. It is well known that spatial data aggregation has its issues (Openshaw 1977, Holt et al. 1996). Temporal data aggregation may have similar issues, causing the aggregated data to magnify a model's result. In my case, the temporal data aggregation may have caused such a situation, as the slope was less than 1 using the data set-by-set but greatly enlarged to values approaching and greater than 1 when using aggregated data. The impact of data aggregation was visible in all cases where the slope from data using aggregation by month and year was greater than the slope produced from set-by-set data. Data aggregation may also cause the loss of important information, which may bias the results or obscure existing patterns (Orcutt et al. 1968, Clark and Avery 1976, Chang et al. 2014). Taylor and Iwanek (1980) suggest disaggregating data to more accurately represent existing patterns. It is also possible that the data included in the aggregation are not enough to estimate the necessary parameters. Increasing the number of observations as well as the number of groups aggregated will help produce unbiased parameter estimates (Orcutt et al. 1968, Clark and Avery 1976). For example, the use of data aggregated by year for the snow crab fishery yields a biased coefficient estimate because the number of parameters estimated is five while the number of observations is four, one for each year, which is less than the parameters estimated. My aggregated data calculates mean catch per vessel on a monthly or yearly basis instead of examining catches from individual sets and thus has less variation. The variability of aggregated catches (based on area, time, etc.) can result in biased estimates (Holt et al. 1996). Though the error in the models can be reduced through data aggregation, this leads to uncertainty about the estimated parameters and any decisions to be made from the results (Chang et al. 2014).

The statistical phenomena of "errors-in-variables" (Hilborn and Walters 1992) may also bias the observed proportionality in our analyses. In this case, the attribution of all variation to the response variable (catch) tends to bias the results towards slopes closer to zero (randomness). In our case, it would favour slopes less than one even if the true underlying relationship was proportional. However, it would not account for slopes great than one, or exclude the effect of other mechanisms.

Fisherman's behaviour may also impact the slope values in temporally aggregated data. The slope can be less than 1 (using data set-by-set) when fishermen fish with the intention of filling their nets or traps rather than for a fixed time interval, which will pull the slope down to be less than 1. This could be the case when, for example, a fisherman intends to fill 15 trawls from one area. Here the first net may be full within one hour, while the second takes perhaps two hours and the third takes four hours. The overall relationship between catch and effort will decelerate because the catch is decreasing gradually in the area, likely due to a decline in abundance or dispersal of fish. This effect may be behind the extreme (negative) coefficient of effort observed for redfish, where trawls of long duration and poor success could fail to fill the net but dominate the overall relationship. Another reason for the slope to be less than one is gear saturation by other species. For example, each trap can hold only a certain number of crabs and may fill up with species other than crabs, which then of course decreases the catch size. Gear saturation also occurs within species. Here, the traps may also fill with crabs within a short period of time. If the traps are left in place fishing becomes unproductive, which also decreases the catch rate over time (Dauk and Schwarz 2001). A relationship between catch and effort could emerge in this case, but it would be a weak relationship.

The slope may be less than 1 in the case of competition even when fish are abundant.

Vessels may interfere with each other by direct interactions or disturbing the fish (interference competition, Goss-Custard 1980) or by locally depleting the fish (exploitation competition) (Forman 1967, Gillis and Peterman 1998). Previous studies have found a role for both forms of competition in the success of fishing activities (Rijnsdorp et al. 2000, Poos and Rijnsdorp 2007). In the case of interference, increasing effort will be less effective at fish capture and may result in a nonlinear relationship between catch and effort, producing lower catches than would be expected from proportionality. Exploitation itself may lead to disproportionality between catch and effort due to continued fishing while local availability declines. For example, if the level of fishing effort takes 10% of the local population weekly, then after the first week they will leave 90% of the original abundance. The 10% harvest of the following week will be lower for the same amount of effort. The longer that fishing occurs the more the catch will be depressed. This is the standard conceptual model of fishing mortality (Ricker 1975), which may not be taken into account when reporting fisheries statistics. This can be even more extreme in cases where increased fishing effort can impact communities and collapse local fishing success (Laë 1997).

Facilitation (Lapointe 1989) would tend to bias proportionality opposite to the effects of competition. Technology, such as GPS, radiotelephone, and radar, increase the capacity for sharing information within fleets and can support facilitation that increases catch rates (Deng et al. 2005, Stevenson et al. 2011). Another form of facilitation is word-of-mouth transfer of information, where fishermen share tips and techniques amongst each other to locate fish (Healey et al. 1990).

In this chapter, we analyzed fixed and mobile gear data using a generalized mixed-effects model to examine the proportionality between catch and nominal effort. In the aggregated data snow cab was closest to a proportional relationship, though still less than one. The disaggregated trawl fishery data displayed slopes closer to, or in several cases indistinguishable from, zero. This may be due to the manner in which effort was measured. In the snow crab fishery the nominal effort measure is number of traps, which are set for similar time periods. Effort increases are associated with more traps. In the trawl fishery, nominal effort is measured by hours fished. "More effort" could be due to more vessels or longer trawls. Thus, fishing to fill a trawl through varying fish densities may reduce or eliminate any relationship between nominal effort and catch. My analysis compared data aggregated by individual set as well as data aggregated by month and year. I found that the relationship between catch and effort was often disproportionate which may be because such factors as interference, exploitation competition, gear saturation, numerical responses, and other fishermen's behaviours (e.g. whether or not fishermen intend to fill their nets). Data aggregated by month and year could produce a slope closer to or even greater than 1, which may be because of the ability of these time frames to capture the variation in vessels entering or leaving the active fishery (numerical responses) or responding to information available in the fleet (facilitation). My chief recommendation is to collect set-by-set (disaggregate) data or records whenever possible to avoid misleading trends and to retain the ability to examine patterns across temporal scales. This would increase the reliability in estimating the parameters and help avoid biased results (Taylor and Iwanek 1980). It is also recommended that the number of observations aggregated be increased, in order to produce an unbiased parameter estimate (Clark and Avery 1976, Orcutt et al. 1968). Distinguishing amongst the many potential behavioural and statistical influences discussed here is not possible without further directed studies. However, I can say that the *a priori* assumption of proportionality, implicit in the use of CPUE, is unwarranted. Instead, fitting models with catch as the response and effort as a predictor will allow the test of this hypothesis and lead to less potentially biased catch standardizations for stock assessments in the future.

References

Beguería, S., & Pueyo, Y. (2009). A comparison of simultaneous autoregressive and generalized least squares models for dealing with spatial autocorrelation. *Global ecology and biogeography*, *18*(3), 273-279.

Bishop, J., Venables, W. N., & Wang, Y. G. (2004). Analysing commercial catch and effort data from a Penaeid trawl fishery: a comparison of linear models, mixed models, and generalized estimating equations approaches. *Fisheries Research*, *70*(2), 179-193.

Bolker, B. M., Brooks, M. E., Clark, C. J., Geange, S. W., Poulsen, J. R., Stevens, M. H. H., & White, J. S. S. (2009). Generalized linear mixed models: a practical guide for ecology and evolution. *Trends in ecology & evolution*, *24*(3), 127-135.

Breslow, N. E., & Clayton, D. G. (1993). Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association*, *88*(421), 9-25.

Chang, W., Haran, M., Olson, R., & Keller, K. (2014). Fast dimension-reduced climate model calibration and the effect of data aggregation. *The Annals of Applied Statistics*, *8*(2), 649-673.

Clark, W. A., & Avery, K. L. (1976). The effects of data aggregation in statistical analysis. *Geographical Analysis*, *8*(4), 428-438.

Conan, G. Y., & Comeau, M. (1986). Functional maturity and terminal molt of male snow crab, Chionoecetes opilio. *Canadian Journal of Fisheries and Aquatic Sciences*, *43*(9), 1710-1719.

Dauk, P. C., & Schwarz, C. J. (2001). Catch estimation in the presence of declining catch rate due to gear saturation. *Biometrics*, 287-293.

Deng, R., Dichmont, C., Milton, D., Haywood, M., Vance, D., Hall, N., & Die, D. (2005). Can vessel monitoring system data also be used to study trawling intensity and population depletion? The example of Australia's northern prawn fishery. *Canadian Journal of Fisheries and Aquatic Sciences*, *62*(3), 611-622.

Dormann, C., M. McPherson, J., B. Araújo, M., Bivand, R., Bolliger, J., Carl, G. & Kühn, I. (2007). Methods to account for spatial autocorrelation in the analysis of species distributional data: a review. *Ecography*, *30*(5), 609-628.

Elner, R. W., & Beninger, P. G. (1992). The reproductive biology of snow crab, *Chionoecetes opilio*: a synthesis of recent contributions. *American Zoologist*, *32*(3), 524-533.

Engel, J., & Kvitek, R. (1998). Effects of otter trawling on a benthic community in Monterey Bay National Marine Sanctuary. *Conservation Biology*, *12*(6), 1204-1214.

FAO 1980. Quantity conversion factors: Atlantic fish species -landed or product weight to live weight. FAO Fish. Circ., No. 725, 217 p.

Forman, S. (1967). Cognition and the catch: the location of fishing spots in a Brazilian coastal village. *Ethnology*, 417-426.

Gillis, D. M., & Peterman, R. M. (1998). Implications of interference among fishing vessels and the ideal free distribution to the interpretation of CPUE. *Canadian Journal of Fisheries and Aquatic Sciences*, *55*(1), 37-46.

Goñi, R., Alvarez, F., & Adlerstein, S. (1999). Application of generalized linear modeling to catch rate analysis of Western Mediterranean fisheries: the Castellón trawl fleet as a case study. *Fisheries Research*, *42*(3), 291-302.

Goss-Custard, J. D. (1980). Competition for food and interference among waders. *Ardea*, *68*, 31-52.

Healey, M. C., & Morris, J. F. T. (1992). The relationship between the dispersion of salmon fishing vessels and their catch. *Fisheries research*, *15*(1), 135-145.

Healey, M. C., Thomson, R. E., & Morris, J. F. (1990). Distribution of commercial troll fishing vessels off southwest Vancouver Island in relation to fishing success and oceanic water properties and circulation. *Canadian Journal of Fisheries and Aquatic Sciences*, *47*(10), 1846-1864.

Hébert, M., Miron, G., Moriyasu, M., Vienneau, R., & DeGrâce, P. (2001). Efficiency and ghost fishing of snow crab (Chionoecetes opilio) traps in the Gulf of St. Lawrence. *Fisheries Research*, *52*(3), 143-153.

Helser, T. E., Punt, A. E., & Methot, R. D. (2004). A generalized linear mixed model analysis of a multi-vessel fishery resource survey. *Fisheries Research*, *70*(2), 251-264.

Hilborn, R., & Walters, C. J. (1992). Quantitative fisheries stock assessment: choice, dynamics and uncertainty. *Reviews in Fish Biology and Fisheries*, *2*(2), 177-178.

Holt, D., Steel, D. G., Tranmer, M., & Wrigley, N. (1996). Aggregation and ecological effects in geographically based data. *Geographical analysis*, *28*(3), 244-261.

Laë, R. (1997). Does overfishing lead to a decrease in catches and yields? An example of two West African coastal lagoons. *Fisheries Management and Ecology*, *4*(2), 149-164.

Lapointe, M. F. (1989). *Numerical and functional responses of British Columbia trawler fishermen* (Doctoral dissertation, M.Sc. thesis, University of British Columbia, Vancouver).

Lovrich, G. A., & Sainte-Marie, B. (1997). Cannibalism in the snow crab, Chionoecetes opilio (O. Fabricius)(Brachyura: Majidae), and its potential importance to recruitment. *Journal of Experimental Marine Biology and Ecology*, *211*(2), 225-245.

Maravelias, C. D., Reid, D. G., Simmonds, E. J., & Haralabous, J. (1996). Spatial analysis and mapping of acoustic survey data in the presence of high local variability: geostatistical application to North Sea herring (Clupea harengus). *Canadian Journal of Fisheries and Aquatic Sciences*, *53*(7), 1497-1505.

Marchal, P., Andersen, B., Bromley, D., Iriondo, A., Mahévas, S., Quirijns, F., & Ulrich, C. (2006). Improving the definition of fishing effort for important European fleets by accounting for the skipper effect. *Canadian Journal of Fisheries and Aquatic Sciences*, *63*(3), 510-533.

Maunder, M. N., & Punt, A. E. (2004). Standardizing catch and effort data: a review of recent approaches. *Fisheries Research*, *70*(2), 141-159.

McCullagh, P., & Nelder, J. A. (1989). Generalized linear models, Chapman & Hall/CRC monographs on statistics & applied probability.

McCulloch, C. E., & Neuhaus, J. M. (2001). *Generalized linear mixed models*. John Wiley & Sons, Ltd.

Miller, R. J. (1976). North American crab fisheries: regulations and their rationales. *Fish. Bull*, *74*(3), 623-633.

Miller, R. J. (1979). Saturation of crab traps: reduced entry and escapement. *Journal du conseil*, *38*(3), 338-345.

Millington, P. (1984). The fishermen as predator: numerical responses of British Columbia gillnet fishermen to salmon abundance.

Openshaw, S. (1977). A geographical solution to scale and aggregation problems in region-building, partitioning and spatial modelling. *Transactions of the institute of british geographers*, 459-472.

Orcutt, G. H., Watts, H. W., & Edwards, J. B. (1968). Data aggregation and information loss. *The American Economic Review*, 773-787.

Poos, J. J., & Rijnsdorp, A. D. (2007). An "experiment" on effort allocation of fishing vessels: the role of interference competition and area specialization. *Canadian Journal of Fisheries and Aquatic Sciences*, *64*(2), 304-313.

Punt, A. E., Walker, T. I., Taylor, B. L., & Pribac, F. (2000). Standardization of catch and effort data in a spatially-structured shark fishery. *Fisheries Research*, *45*(2), 129-145.

R Core Team (2014). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org.

Ricker, W.E. (1975). Computation and interpretation of biological statistics of fish populations. *Bull. Fish. Res. Board Can.* No. 191.

Rijnsdorp, A. D., Dol, W., Hoyer, M., & Pastoors, M. A. (2000). Effects of fishing power and competitive interactions among vessels on the effort allocation on the trip level of the Dutch beam trawl fleet. *ICES Journal of Marine Science: Journal du Conseil*, *57*(4), 927-937.

Salas, S., & Gaertner, D. (2004). The behavioural dynamics of fishers: management implications. *Fish and fisheries*, *5*(2), 153-167.

Skud, B. E. (1978). Factors affecting longline catch and effort: III. Bait loss and competition. *Int. Pac. Halibut Comm. Sci. Rep*, *64*, 25-50.

Stevenson, T. C., Tissot, B. N., & Dierking, J. (2011). Fisher behaviour influences catch productivity and selectivity in West Hawaii's aquarium fishery. *ICES Journal of Marine Science: Journal du Conseil*, *68*(5), 813-822.

Taylor, P. B., & Iwanek, R. (1980). Intuitive versus optimal solutions to some problems in distribution. *Omega*, *8*(2), 183-192.

Venables, W. N., & Dichmont, C. M. (2004). GLMs, GAMs and GLMMs: an overview of theory for applications in fisheries research. *Fisheries research*, *70*(2), 319-337.

Vuong, Q. H. (1989). Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica: Journal of the Econometric Society*, 307-333.

Walters, C. (2003). Folly and fantasy in the analysis of spatial catch rate data. *Canadian Journal of Fisheries and Aquatic Sciences*, *60*(12), 1433-1436.

Watson, J. (1970). Maturity, mating, and egg laying in the spider crab, *Chionoecetes opilio*. *Journal of the Fisheries Board of Canada*, *27*(9), 1607-1616.

Ye, Y., Al-Husaini, M., & Al-Baz, A. (2001). Use of generalized linear models to analyze catch rates having zero values: the Kuwait driftnet fishery. *Fisheries Research*, *53*(2), 151-168.

Zuur, A. F., Ieno, E. N., & Elphick, C. S. (2010). A protocol for data exploration to avoid common statistical problems. *Methods in Ecology and Evolution*, *1*(1), 3-14.

Zuur, A. F., Ieno, E. N., Walker, N. J., Saveliev, A. A., & Smith, G. M. (2009). Mixed effects models and extensions in ecology with R; Gail M, Krickeberg K, Samet JM, Tsiatis A, Wong W, editors. *New York, NY: Spring Science and Business Media*.

Chapter Five

Conclusion

Conclusion

This research used various methods to examine the proportionality between catch and nominal effort. I used meta-analysis and a generalized linear mixed-effects model on two different fisheries, one using fixed and the other mobile gear. I used fixed-effects with meta-analysis and I incorporated the random effects using vessels' I.D. in the GLMM. This research has shown that, catch is not universally proportional to nominal effort in fisheries data. It is not the case that increasing fishing effort, such as increasing hours spent on fishing, or increasing the number of traps will necessarily increase catch by the same factor. Effort does not correlate directly with catch in this case; there is no single possible explanation for the variability of catch. There are a number of factors that act collectively. In order to find an accurate explanation, we need to consider all of these when collecting data.

Fishermen's behaviour should always be considered when interpreting catch and effort statistics. Fishermen are different from each other in their goals in fishing, their knowledge, and their ideas. Catch may reach its peak in one fishery, while competing crews may not catch very much, even though both of them are in the same area, and this may be explainable through the fact that each of them has different goals. Fishermen's behaviour may decrease the catch when the goal is to fill the net even when other tactics may yield higher catch rates. Another factor that may decrease catch is that, when there is a high catch and a high number of vessels aggregated in the same area (numerical responses), the fishermen may face interference competition. In addition, exploitation competition has a huge effect on a number of fish species, such as sharks, large pelagic fish, etc. Some areas have been depleted because of over-fishing, which, in turn, affects catch and may create disproportionality in the relationship between catch and effort. When fishermen intended to catch a given species, disproportionality in the form of a decelerated catch rate could appear for a number of reasons. One of them is gear saturation by species other than the intended target filling nets and traps and preventing the capture of the target. If gear fills up with debris or if high numbers of fish fill up a net quickly, further potential catches may be lost and catch rate will decrease as a result. Another factor one may encounter is species competition, where species compete for the bait. When fish populations increase in an area, high numbers of vessels will aggregate creating numerical responses. If these vessels cooperate with each other, facilitation will result, and catch may increase for all vessels.

Statistical influences on the relationship between catch and effort must also be considered. As in so many models, “error-in-variables” bias may be present when there is uncertainty in both catch and effort measures. In addition, aggregated data may provide false information about the relationship between catch and effort. It may show that catch has increased, but, in reality, catch may have decelerated with effort. This was shown to be the case in the set-by-set analysis, which more accurately portrayed the relationship on the scale of fishing activities than the analyses of data aggregated by month and year.

Thus, although a many reasons could explain the disproportionality between catch and effort, both behavioural and statistical, the use of CPUE in estimating abundance, as C/f (C=catch, f=nominal effort) assumes that catch will be proportional to effort. This analysis has shown that this is not the case, which leads to the conclusion that CPUE may be misleading and should not be automatically assumed to accurately reflect fish abundance trends in fisheries data.

Appendix A: R Code for Meta-analysis.

The following samples of R code represent typical instructions to perform the analyses of Chapter 2. To conserve space, it is not an exhaustive listing of all analyses used in the chapter.

```
# load library(ies) and read data
require(metafor)
metaData=read.table("~/Desktop/final analysis /Meta-sheet with p values L to S.csv",
                  sep=",",
                  header=T,
                  stringsAsFactors=F)
table(metaData$FisheryType)
keep=c("artisanal","commercial", "research")
metaData=metaData[metaData$FisheryType%in%keep,]
table(metaData$FisheryType)
# Define meta-analysis variables
yi=metaData$RMA_slope   # effect size as RMA slope
sei=metaData$OLS_SE     # sample variability as standard errors of the slopes
#--------------------------------------------------------------------------
# Meta-analysis and summaries for all data...
# For interpretation see: Viechtbauer, W. (2010). Conducting meta-analyses in
#                   R with the metafor package. Journal of Statistical
#                   Software, 36(3), 1-48.
metaRMA=rma(yi,sei)
summary(metaRMA)
confint(metaRMA)
par(mfrow=c(1,2))
forest(metaRMA, refline=1, order="obs", slab=metaData$StudyID)
funnel(metaRMA)
# test for funnel plot asymmetry
regtest(metaRMA, model="rma", predictor="sei")
#--------------------------------------------------------------------------
# Define moderator: FisheryType
FisheryType=factor(metaData$FisheryType,
          levels=c("artisanal","commercial", "research"))
nLevels= length(levels(FisheryType))
contrasts(FisheryType)=contr.treatment(n=nLevels,base=1)
table(FisheryType)
# Mixed-model meta-analysis (RMA slope as RE, FisheryType as moderator variable)
metaRMAFisheryType=rma(yi, sei, mods = ~ FisheryType)
summary(metaRMAFisheryType)
confint(metaRMAFisheryType)
print(levels(FisheryType))
par(mfrow=c(1,2))
forest(metaRMAFisheryType, refline=1, order="obs", slab=FisheryType)
```

```
funnel(metaRMAFisheryType,pch=as.character(FisheryType),cex=0.5)
#------------------------------------------------------------------------------
# Mixed-model meta-analysis (RMA slope as RE, FisheryType as moderator variable)
metaRMAFisheryType=rma(yi, sei, mods = ~ FisheryType)
summary(metaRMAFisheryType)
confint(metaRMAFisheryType)
print(levels(FisheryType))
par(mfrow=c(1,2))
forest(metaRMAFisheryType, refline=1, order="obs", slab=FisheryType)
funnel(metaRMAFisheryType,pch=as.character(FisheryType),cex=0.5)
```

Appendix B: R code for 4X analysis using set-by-set data.

The following samples of R code represent typical instructions to perform the analyses of Chapter 4. To conserve space, it is not an exhaustive listing of all analyses used in the chapter.

```
rawDataVal = read.table("rawDataVal1.csv",
                        header = TRUE,
                        sep = ",",
                        stringsAsFactors = FALSE)
head(rawDataVal)
names(rawDataVal)
#How many species that have Na prices in the data frame
sum(is.na(rawDataVal$price))
dim(rawDataVal)[1]
# Let's explore a bit...
X=rawDataVal[is.na(rawDataVal$price),]
head(X)
unique(X$SPECIES)
NA_SPECIES=unique(X$SPECIES)
NA_SPECIES_WT=sum(rawDataVal$RND_WEIGHT_KGS[
  rawDataVal$SPECIES%in%NA_SPECIES])
TOT_WT=sum(rawDataVal$RND_WEIGHT_KGS)
NA_SPECIES_PCT=NA_SPECIES_WT/TOT_WT
NA_SPECIES_PCT
# Remove species that does not have prices in the beginning
x=rawDataVal[rawDataVal$SPECIES%in%NA_SPECIES,]
x
str(x)
rawDataVal1 <- rawDataVal[!rawDataVal$SPECIES%in%NA_SPECIES,]
dim(rawDataVal1)
str(rawDataVal1)
allData=rawDataVal1
head(allData)
names(allData)
str(allData)
dim(allData)
# How many sets from each year?
table(allData$YEAR_FISHED)
# Change integers into numeric
allData$TRIP_ID=as.numeric(allData$TRIP_ID)
allData$SETID=as.numeric(allData$SETID)
allData$YEAR_FISHED=as.numeric(allData$YEAR_FISHED)
allData$MONTH_FISHED=as.numeric(allData$MONTH_FISHED)
allData$DAY_FISHED=as.numeric(allData$DAY_FISHED)
allData$TONNAGE_CLASS=as.numeric(allData$TONNAGE_CLASS)
allData$GROSS_TONNAGE=as.numeric(allData$GROSS_TONNAGE)
```

```
allData$HOURS_FISHED=as.numeric(allData$HOURS_FISHED)
str(allData)
# Change what needs to be a factor into a factor
allData$VCLASS=factor(allData$VCLASS)
allData$TONNAGE_CLASS=factor(allData$TONNAGE_CLASS)
allData$PSEUDO_VESSEL_ID=factor(paste("V",allData$PSEUDO_VESSEL_ID,
                                sep=""))
allData$GEAR_CODE_1=factor(allData$GEARCODE1)
allData$GEAR_TYPE_ID=factor(allData$GEAR_TYPE_ID)
allData$GEAR_TYPE_DESC=factor(allData$GEAR_TYPE_DESC)
allData$NAFO_DIV=factor(allData$NAFO_DIV)
allData$UNIT_AREA=factor(allData$UNIT_AREA)
allData$GEAR_CODE_2=factor(allData$GEARCODE2)
allData$GEAR_DESC=factor(allData$GEAR_DESC)
allData$EFFORT_LEVEL=factor(allData$EFFORT_LEVEL)
str(allData)
# Next let's deal with the issue of date...
# Change LANDED_DATE and DATE_FISHED into posix time
# Strip quotes from the string
wstr = substr(allData$LANDED_DATE,1,11)
# Convert the string to a date
allData$LANDED_DATE1=as.Date(wstr, format="%Y-%m-%d")
# Strip quotes from the string
wstr2 = substr(allData$DATE_FISHED,1,11)
# convert the string to a date
allData$DATE_FISHED1=as.Date(wstr2, format="%Y-%m-%d")
# create day of year
allData$FISHED.doy=as.numeric(format(allData$DATE_FISHED1,
                              format = "%j",
                              tz="GMT"))
# creat week
allData$FISHED.week=as.numeric(format(allData$DATE_FISHED1,
                               format = "%W",
                               tz="GMT"))
# creat weekday
allData$FISHED.weekday=as.numeric(format(allData$DATE_FISHED1,
                                  format = "%w",
                                  tz="GMT"))
str(allData)

# Compare YEAR_FISHED, ...  with date fields
allData[1400:1430,c(5:7,17)] # spotcheck look OK
# General data set characteristics
# Size:
dim(allData)
# Number records by year
```

```
table(allData$YEAR_FISHED)
# Number of trips and trips by year
x=allData[!duplicated(allData[,"TRIP_ID"]),]
dim(x)
table(x$YEAR_FISHED)
rm(x)
# Number of sets and sets per year
x=allData[!duplicated(allData[,"SETID"]),]
dim(x)
table(x$YEAR_FISHED)
rm(x)
head(allData)
# Create data frame with important variables
setData=as.data.frame(list(TRIP_ID = allData$TRIP_ID,
                           SETID = allData$SETID,
                           value = allData$value,
                           RND_WEIGHT_KGS = allData$RND_WEIGHT_KGS,
                           UNIT_AREA = allData$UNIT_AREA,
                           PSEUDO_VESSEL_ID = allData$PSEUDO_VESSEL_ID,
                           HOURS_FISHED = allData$HOURS_FISHED,
                           LOA = allData$LOA,
                           SPECIES = allData$SPECIES,
                           SPECIES_ABBREV= allData$SPECIES_ABBREV,
                           GEAR_DESC = allData$GEAR_DESC,
                           GROSS_TONNAGE = allData$GROSS_TONNAGE,
                           YEAR_FISHED = allData$YEAR_FISHED,
                           MONTH_FISHED = allData$MONTH_FISHED,
                           DAY_FISHED= allData$DAY_FISHED,
                           FISHED.doy = allData$FISHED.doy,
                           EFFORT_LEVEL=allData$EFFORT_LEVEL,
                           NAFO_DIV = allData$NAFO_DIV,
                           DATE_FISHED= allData$DATE_FISHED,
                           LAT = allData$LAT,
                           LON = allData$LON))

str(setData)
head(setData)
dim(setData)
#Remove non-positive values
setData = setData[setData$value>0,]
setData = setData[setData$RND_WEIGHT_KGS>0,]
setData = setData[setData$HOURS_FISHED>0,]
setData = setData[setData$YEAR_FISHED>0,]
head(setData)
# Remove sets greater than 6 hours
setData = setData[setData$HOURS_FISHED<7,]
```

```
table(setData$HOURS_FISHED)
head(setData)
# Distribution of gear types
table(setData$GEAR_DESC)
levels(setData$GEAR_DESC)
# Keep "OTTER TRAWL, STERN" and remove 4 of SCOTTISH SEINE
setData = setData[setData$GEAR_DESC=="OTTER TRAWL, STERN",]
table(setData$GEAR_DESC)
head(setData)
# Occurrences of sizes and vessels
table(setData$GROSS_TONNAGE)
table(setData$LOA)
table(setData$PSEUDO_VESSEL_ID)
# Remove missing values
dim(setData)[1]
setData <- na.omit(setData)
dim(setData)[1]
# Choose vessels fished more than 100 sets per year
# Convert factor to character
setData$PSEUDO_VESSEL_ID <- as.character(setData$PSEUDO_VESSEL_ID)
str(setData)
table(setData$YEAR_FISHED)
table(setData$PSEUDO_VESSEL_ID,setData$YEAR_FISHED)
x=table(setData$PSEUDO_VESSEL_ID,setData$YEAR_FISHED)
x=x>99
x
x=x*1
x
is.matrix(x)
xx=rowSums(x)
xx
xxx=xx[xx==6]
xxx
names(xxx)
keepVessels=names(xxx)
keepVessels
setData=setData[setData$PSEUDO_VESSEL_ID%in%keepVessels,]
head(setData)
dim(setData)
table(setData$PSEUDO_VESSEL_ID,setData$YEAR_FISHED)
str(setData)
# We will convert year, month, and loa to factors
setData$fYEAR_FISHED <- factor(setData$YEAR_FISHED)
setData$fMONTH_FISHED <- factor(setData$MONTH_FISHED)
setData$fLOA <- cut(setData$LOA, 3)
str(setData)
```

```
# Order data by vessel, year, month, day
o=order(setData$PSEUDO_VESSEL_ID,
        setData$YEAR_FISHED,
        setData$MONTH_FISHED ,
        setData$DAY_FISHED,
        setData$SETID)
x=setData[o,]
head(x)
dim(setData)
summary(setData$RND_WEIGHT_KGS)
setData=setData[setData$RND_WEIGHT_KGS>0.45,]
summary(setData$RND_WEIGHT_KGS)
qqnorm(log(setData$RND_WEIGHT_KGS)); qqline(log(setData$RND_WEIGHT_KGS))
dim(setData)
# QUESTION: What is the relationship between weight and set characteristics for each species
# VARIABLES: Response: weight
#            Predictors:  log(HOURS_FISHED), LOA,
#                         YEAR_FISHED, MONTH_FISHED
#                          (VID as random effect?)
# Compare GLM, GLMM, and GAMM approaches for each species
#Create data frame for each species
#---------------------------------------------------------------------------------------------------HADDOCK
df1 = setData[setData$SPECIES == "HADDOCK", ]
head(df1)
dim(df1)
dim(setData)
df1 <- subset(df1, !duplicated(df1[,1:2]))
sum(duplicated(df1[,1:2])) #OK!
table(df1$YEAR_FISHED, df1$SPECIES)
#Remove Nas
df1[!is.na(df1$HOURS_FISHED),]
# How many sets from each year?
table(df1$YEAR_FISHED)
dim(df1)
# Number records by year
table(df1$YEAR_FISHED)
# Number of trips and trips by year
x=df1[!duplicated(df1[,"TRIP_ID"]),]
dim(x)
table(x$YEAR_FISHED)
rm(x)
# Number of sets and sets per year
x=df1[!duplicated(df1[,"SETID"]),]
dim(x)
table(x$YEAR_FISHED)
rm(x)
```

```
table(df1$HOURS_FISHED)
# Distribution of gear types
table(df1$GEAR_DESC)
levels(df1$GEAR_DESC)
# Occurrences of sizes and vessels
table(df1$GROSS_TONNAGE)
table(df1$LOA)
table(df1$PSEUDO_VESSEL_ID)
# How many vessels fishing for haddock
table(df1$PSEUDO_VESSEL_ID,df1$YEAR_FISHED)
# >>> Data Exploration (use Zuur as a guide) <<<
# Zuur's general 2010 protocol can be summarized as examining:
#  A Outliers in Y & X
#  B Homogeniety in Y
#  C Normality in Y
#  D Zero trouble in Y
#  E Collinearity X
#  F Relationships Y vs X
#  G Interactions in X
#  H Independence in Y (spatial & temporal)
# QUESTION: What is the relationship between catch and effort in the trawl of specific species
# VARIABLES: Response: Catch
#           Predictors:  HOURS_FISHED, LOA, YEAR_FISHED, MONTH_FISHED,
 FISHED.week
o=order(df1$PSEUDO_VESSEL_ID,df1$YEAR_FISHED,df1$MONTH_FISHED
       ,df1$DAY_FISHED,df1$SETID)
x=df1[o,]
head(x)
#----------------------------------------------------------------------------------------------------
# QUESTION: What is the relationship between catch and effort for each species
#         #Find if the relationship is proportional or disproportional between catch and effort for
 each species
#
# VARIABLES: Response: Catch
#           Predictors:  log(HOURS_FISHED), LOA,
#                        YEAR_FISHED, MONTH_FISHED
#                         (VID as random effect?)
#
# Compare GLM, GLMM, and GAMM approaches
#----------------------------------------------------------------------------------------------------
# GLM full model
glm.01a= glm(RND_WEIGHT_KGS ~ log(HOURS_FISHED) +
             fLOA +
             fYEAR_FISHED+
             fMONTH_FISHED,
           data=df1,
```

```
            family=Gamma(link="log"))
summary(glm.01a)
par(mfrow=c(2,2))
plot(glm.01a)
par(mfrow=c(2,2))
Y_hat = predict(glm.01a)
E = residuals(glm.01a, type="pearson")
E_lo = lowess(as.numeric(Y_hat), as.numeric(E))
plot(Y_hat,E); lines(E_lo, col="red")
hist(E)
qqnorm(E); qqline(E)
acf(E)
# Compare models after dropping one predictor
drop1(glm.01a)
# Create table for word from glm
str(summary(glm.01a))
out=summary(glm.01a)
table=as.data.frame(out$coefficients[,1:2])
table$lowerCI = table[,1] - table[,2]*1.96
table$upperCI = table[,1] + table[,2]*1.96
table
#---------------------------------------------------------------------------------------------------
#GLMM
# Generalized linear mixed model
library(lme4)
# Dealing with autocorrelation
# 0) order by vessel, year, month, day, setID
o=order(df1$PSEUDO_VESSEL_ID,
      df1$YEAR_FISHED,
      df1$MONTH_FISHED ,
      df1$SETID)
df1=df1[o,]
head(df1)
#Typically how many sets per day?
x=aggregate(df1$RND_WEIGHT_KGS,
          by=list(df1$PSEUDO_VESSEL_ID,df1$FISHED.doy),
          FUN=length)
summary(x)
# Randomly select one set per vessel per day
# 1) define function to randomly pick a single
#    Value from a numeric vector
pickem=function(x) {
  sample(x,1)
}
# 2) Make the numeric vector of sets to keep
keepSets=as.numeric(aggregate(df1$SETID,
```

```
                        by=list(df1$PSEUDO_VESSEL_ID,
                              df1$FISHED.doy),
                        FUN=pickem)[,3])
length(keepSets)
# 3) select set records for analysis
df1=df1[df1$SETID%in%keepSets,]
dim(df1)
# Simplify model? Use intercept random effect model with data subset
glmm4=  glmer(RND_WEIGHT_KGS ~ log(HOURS_FISHED) +
             fYEAR_FISHED+
             fMONTH_FISHED +
             (1 | PSEUDO_VESSEL_ID),
           data=df1,
           family=Gamma(link="log"))
drop1(glmm4,test="Chisq")
# keep it all
print(summary(glmm4), corr=F)
# Diagnostic plots
par(mfrow=c(2,2))
Y_hat = predict(glmm4)
E = residuals(glmm4, type="deviance")
E_lo = lowess(as.numeric(Y_hat), as.numeric(E))
plot(Y_hat,E); lines(E_lo, col="red")
hist(E)
qqnorm(E); qqline(E)
acf(E)
#
str(summary(glmm4))
out=summary(glmm4)
table=as.data.frame(out$coefficients[,1:2])
table$lowerCI = table[,1] - table[,2]*1.96
table$upperCI = table[,1] + table[,2]*1.96
table
#----------------------------------------------------------------------------------------------------COD
#create data frame with COD species
str(setData)
dim(setData)
head(setData)
df1 = setData[setData$SPECIES == "COD", ]
head(df1)
dim(df1)
str(df1)
df1 <- subset(df1, !duplicated(df1[,1:2]))
sum(duplicated(df1[,1:2])) #OK!
# How many COD each year
table(df1$YEAR_FISHED, df1$SPECIES)
```

```
#Remove Nas
df1[!is.na(df1$HOURS_FISHED),]
dim(df1)
#How many sets from each year?
table(df1$YEAR_FISHED)
# Number records by year
table(df1$YEAR_FISHED)
#Number of trips and trips by year
x=df1[!duplicated(df1[,"TRIP_ID"]),]
dim(x)
table(x$YEAR_FISHED)
rm(x)
#Number of sets and sets per year
x=df1[!duplicated(df1[,"SETID"]),]
dim(x)
table(x$YEAR_FISHED)
rm(x)
#Distribution of gear types
table(df1$GEAR_DESC)
levels(df1$GEAR_DESC)
#Occurrences of sizes and vessels
table(df1$GROSS_TONNAGE)
table(df1$LOA)
table(df1$PSEUDO_VESSEL_ID)
table(df1$PSEUDO_VESSEL_ID,df1$YEAR_FISHED)
summary(df1$RND_WEIGHT_KGS)
qqnorm(log(df1$RND_WEIGHT_KGS)); qqline(log(df1$RND_WEIGHT_KGS))
# We will convert year, month, and loa to factors
df1$fYEAR_FISHED <- factor(df1$YEAR_FISHED)
df1$fMONTH_FISHED <- factor(df1$MONTH_FISHED)
df1$fLOA <- cut(df1$LOA, 3)
str(df1)
# Order data by vessel, year, month, day
o=order(df1$PSEUDO_VESSEL_ID,df1$YEAR_FISHED,df1$MONTH_FISHED
 ,df1$DAY_FISHED,df1$SETID)
x=df1[o,]
head(x)
# GLM full model
glm.01= glm(RND_WEIGHT_KGS ~ log(HOURS_FISHED) +
            fLOA +
            fYEAR_FISHED+
            fMONTH_FISHED,
          data=df1,
          family=Gamma(link="log"))
summary(glm.01)
par(mfrow=c(2,2))
```

```
plot(glm.01)
par(mfrow=c(2,2))
Y_hat = predict(glm.01)
E = residuals(glm.01, type="pearson")
E_lo = lowess(as.numeric(Y_hat), as.numeric(E))
plot(Y_hat,E); lines(E_lo, col="red")
hist(E)
qqnorm(E); qqline(E)
acf(E)
#Other potential issues?
glm.01a= glm(RND_WEIGHT_KGS ~ log(HOURS_FISHED) +
           fLOA +
           fYEAR_FISHED+
           fMONTH_FISHED,
          data=df1,
          family=Gamma(link="log"))
print(summary(glm.01a), corr=F)
par(mfrow=c(2,2))
plot(glm.01a)
par(mfrow=c(2,2))
Y_hat = predict(glm.01a)
E = residuals(glm.01a, type="pearson")
E_lo = lowess(as.numeric(Y_hat), as.numeric(E))
plot(Y_hat,E); lines(E_lo, col="red")
hist(E)
qqnorm(E); qqline(E)
acf(E)
# Compare models after dropping one predictor
drop1(glm.01a)
str(summary(glm.01a))
out=summary(glm.01a)
table=as.data.frame(out$coefficients[,1:2])
table$lowerCI = table[,1] - table[,2]*1.96
table$upperCI = table[,1] + table[,2]*1.96
table
#---------------------------------------------------------------------------------------------------
#GLMM
library(lme4)
# Dealing with autocorrelation
# 0) order by vessel, year, month, day, setID
o=order(df1$PSEUDO_VESSEL_ID,
     df1$YEAR_FISHED,
     df1$MONTH_FISHED ,
     df1$SETID)
df1=df1[o,]
head(df1)
```

```
# Typically how many sets per day?
x=aggregate(df1$RND_WEIGHT_KGS,
          by=list(df1$PSEUDO_VESSEL_ID,df1$FISHED.doy),
          FUN=length)
summary(x)
# Median of 4 sets per day
#
# Randomly select one set per vessel per day
# 1) define function to randomly pick a single
#    value from a numeric vector
pickem=function(x) {
  sample(x,1)
}
# 2) make the numeric vector of sets to keep
keepSets=as.numeric(aggregate(df1$SETID,
                    by=list(df1$PSEUDO_VESSEL_ID,
                            df1$FISHED.doy),
                    FUN=pickem)[,3])
length(keepSets)
# 3) select set records for analysis
df2=df1[df1$SETID%in%keepSets,]
dim(df2)
#-----------------------------------------------------------------------------------------------
glmm4=  glmer(RND_WEIGHT_KGS ~ log(HOURS_FISHED) +
              fYEAR_FISHED+
              fMONTH_FISHED +
              (1 | PSEUDO_VESSEL_ID),
             data=df2,
             family=Gamma(link="log"))
drop1(glmm4,test="Chisq")
# remove fLOA
print(summary(glmm4), corr=F)
# diagnostic plots
par(mfrow=c(2,2))
Y_hat = predict(glmm4)
E = residuals(glmm4, type="deviance")
E_lo = lowess(as.numeric(Y_hat), as.numeric(E))
plot(Y_hat,E); lines(E_lo, col="red")
hist(E)
qqnorm(E); qqline(E)
acf(E)
str(summary(glmm4))
out=summary(glmm4)
table=as.data.frame(out$coefficients[,1:2])
table$lowerCI = table[,1] - table[,2]*1.96
table$upperCI = table[,1] + table[,2]*1.96
```

```
table
#----------------------------------------------------------------------------------------------------
# FYI look back at GLM equilivant with same subset
glm.4=  glm(RND_WEIGHT_KGS ~ log(HOURS_FISHED) +
         fYEAR_FISHED+
         fMONTH_FISHED,
        data=df1a2,
        family=Gamma(link="log"))
drop1(glm.4,test="Chisq")
print(summary(glm.4), corr=F)
# Diagnostic plots
par(mfrow=c(2,2))
Y_hat = predict(glm.4)
E = residuals(glm.4, type="deviance")
E_lo = lowess(as.numeric(Y_hat), as.numeric(E))
plot(Y_hat,E); lines(E_lo, col="red")
hist(E)
qqnorm(E); qqline(E)
acf(E)
glmm5a =  glmer(RND_WEIGHT_KGS ~ log(HOURS_FISHED) +
            (1 | PSEUDO_VESSEL_ID),
           data=df1a2,
           family=Gamma(link="log"))
print(summary(glmm5a), corr=F)
# What about random effects?
glm5a =  glm(RND_WEIGHT_KGS ~ log(HOURS_FISHED),
         data=df1a2,
         family=Gamma(link="log"))
print(summary(glm5a), corr=F)
table(df1a2$PSEUDO_VESSEL_ID,df1a2$YEAR_FISHED)
head(df1a2)
dim(df1a2)
str(df1a2)
#---------------------------------------------------------------------------------------------------- POLLOCK
df1 = setData[setData$SPECIES == "POLLOCK", ]
head(df1)
dim(df1)
dim(setData)
df1 <- subset(df1, !duplicated(df1[,1:2]))
sum(duplicated(df1[,1:2])) #OK!
table(df1$YEAR_FISHED, df1$SPECIES)
#remove Nas
df1[!is.na(df1$HOURS_FISHED),]
# How many sets from each year?
table(df1$YEAR_FISHED)
# Number records by year
```

```
table(df1$YEAR_FISHED)
# Number of trips and trips by year
x=df1[!duplicated(df1[,"TRIP_ID"]),]
dim(x)
table(x$YEAR_FISHED)
rm(x)
# Number of sets and sets per year
x=df1[!duplicated(df1[,"SETID"]),]
dim(x)
table(x$YEAR_FISHED)
rm(x)
# Distribution of gear types
table(df1$GEAR_DESC)
levels(df1$GEAR_DESC)
# Occurrences of sizes and vessels
table(df1$GROSS_TONNAGE)
table(df1$LOA)
table(df1$PSEUDO_VESSEL_ID)
table(df1$PSEUDO_VESSEL_ID,df1$YEAR_FISHED)
# Order data by vessel, year, month, day
o=order(df1$PSEUDO_VESSEL_ID,df1$YEAR_FISHED,df1$MONTH_FISHED
 ,df1$DAY_FISHED,df1$SETID)
x=df1[o,]
head(x)
# GLM full model
glm.01= glm(RND_WEIGHT_KGS ~ log(HOURS_FISHED) +
            fLOA +
            fYEAR_FISHED+
            fMONTH_FISHED,
          data=df1,
          family=Gamma(link="log"))
summary(glm.01)
par(mfrow=c(2,2))
plot(glm.01)
par(mfrow=c(2,2))
Y_hat = predict(glm.01)
E = residuals(glm.01, type="pearson")
E_lo = lowess(as.numeric(Y_hat), as.numeric(E))
plot(Y_hat,E); lines(E_lo, col="red")
hist(E)
qqnorm(E); qqline(E)
acf(E)
str(summary(glm.01))
out=summary(glm.01)
table=as.data.frame(out$coefficients[,1:2])
table$lowerCI = table[,1] - table[,2]*1.96
```

```
table$upperCI = table[,1] + table[,2]*1.96
table
#-----------------------------------------------------------------------------------------------
#GLMM
# Generalized linear mixed model
library(lme4)
# Dealing with autocorrelation
# 0) order by vessel, year, month, day, setID
o=order(df1$PSEUDO_VESSEL_ID,
       df1$YEAR_FISHED,
       df1$MONTH_FISHED ,
       df1$SETID)
df1=df1[o,]
head(df1)
# Typically how many sets per day?
x=aggregate(df1$RND_WEIGHT_KGS,
          by=list(df1$PSEUDO_VESSEL_ID,df1$FISHED.doy),
          FUN=length)
summary(x)
# Randomly select one set per vessel per day
# 1) define function to randomly pick a single
#    value from a numeric vector
pickem=function(x) {
  sample(x,1)
}
# 2) make the numeric vector of sets to keep
keepSets=as.numeric(aggregate(df1$SETID,
                        by=list(df1$PSEUDO_VESSEL_ID,
                              df1$FISHED.doy),
                        FUN=pickem)[,3])
length(keepSets)
# 3) select set records for analysis
df1=df1[df1$SETID%in%keepSets,]
dim(df1)
#-----------------------------------------------------------------------------------------------
# Simplify model? Use intercept random effect model with data subset
glmm4=  glmer(RND_WEIGHT_KGS ~ log(HOURS_FISHED) +
              fYEAR_FISHED+
              fMONTH_FISHED +
              (1 | PSEUDO_VESSEL_ID),
            data=df1,
            family=Gamma(link="log"))
# no warning
drop1(glmm4,test="Chisq")
# remove fLOA
print(summary(glmm4), corr=F)
```

```
# diagnostic plots
par(mfrow=c(2,2))
Y_hat = predict(glmm4)
E = residuals(glmm4, type="deviance")
E_lo = lowess(as.numeric(Y_hat), as.numeric(E))
plot(Y_hat,E); lines(E_lo, col="red")
hist(E)
qqnorm(E); qqline(E)
acf(E)
# FYI look back at GLM equilivant with same subset
glm.4=  glm(RND_WEIGHT_KGS ~ log(HOURS_FISHED) +
         fYEAR_FISHED+
         fMONTH_FISHED+
         fLOA,
        data=df1a2,
        family=Gamma(link="log"))
drop1(glm.4,test="Chisq")
print(summary(glm.4), corr=F)
# Diagnostic plots
par(mfrow=c(2,2))
Y_hat = predict(glm.4)
E = residuals(glm.4, type="deviance")
E_lo = lowess(as.numeric(Y_hat), as.numeric(E))
plot(Y_hat,E); lines(E_lo, col="red")
hist(E)
qqnorm(E); qqline(E)
acf(E)
#------------------------------------------------------------------------------------------------
# Create table for word from glmm4
str(summary(glmm4))
out=summary(glmm4)
table=as.data.frame(out$coefficients[,1:2])
table$lowerCI = table[,1] - table[,2]*1.96
table$upperCI = table[,1] + table[,2]*1.96
table
#------------------------------------------------------------------------------------------------ REDFISH
df1 = setData[setData$SPECIES == "REDFISH", ]
head(df1)
dim(df1)
dim(setData)
df1 <- subset(df1, !duplicated(df1[,1:2]))
sum(duplicated(df1[,1:2])) #OK!
table(df1$YEAR_FISHED, df1$SPECIES)
#remove Nas
df1[!is.na(df1$HOURS_FISHED),]
# How many sets from each year?
```

```
table(df1$YEAR_FISHED)
# Number records by year
table(df1$YEAR_FISHED)
# Number of trips and trips by year
x=df1[!duplicated(df1[,"TRIP_ID"]),]
dim(x)
table(x$YEAR_FISHED)
rm(x)
# Number of sets and sets per year
x=df1[!duplicated(df1[,"SETID"]),]
dim(x)
table(x$YEAR_FISHED)
rm(x)
# Distribution of gear types
table(df1$GEAR_DESC)
levels(df1$GEAR_DESC)
# Occurrences of sizes and vessels
table(df1$GROSS_TONNAGE)
table(df1$LOA)
table(df1$PSEUDO_VESSEL_ID)
table(df1$PSEUDO_VESSEL_ID,df1$YEAR_FISHED)
# Order data by vessel, year, month, day
o=order(df1$PSEUDO_VESSEL_ID,df1$YEAR_FISHED,df1$MONTH_FISHED
 ,df1$DAY_FISHED,df1$SETID)
x=df1[o,]
head(x)
# GLM full model
glm.01= glm(RND_WEIGHT_KGS ~ log(HOURS_FISHED) +
            fLOA +
            fYEAR_FISHED+
            fMONTH_FISHED,
          data=df1,
          family=Gamma(link="log"))
summary(glm.01)
par(mfrow=c(2,2))
plot(glm.01)
par(mfrow=c(2,2))
Y_hat = predict(glm.01)
E = residuals(glm.01, type="pearson")
E_lo = lowess(as.numeric(Y_hat), as.numeric(E))
plot(Y_hat,E); lines(E_lo, col="red")
hist(E)
qqnorm(E); qqline(E)
acf(E)
glm.01a= glm(RND_WEIGHT_KGS ~ log(HOURS_FISHED) +
             fLOA +
```

```
        fYEAR_FISHED+
        fMONTH_FISHED,
       data=df1a1,
       family=Gamma(link="log"))
summary(glm.01a)
par(mfrow=c(2,2))
plot(glm.01a)
par(mfrow=c(2,2))
Y_hat = predict(glm.01a)
E = residuals(glm.01a, type="pearson")
E_lo = lowess(as.numeric(Y_hat), as.numeric(E))
plot(Y_hat,E); lines(E_lo, col="red")
hist(E)
qqnorm(E); qqline(E)
acf(E)
# Compare models after dropping one predictor
drop1(glm.01a)
str(summary(glm.01a))
out=summary(glm.01a)
table=as.data.frame(out$coefficients[,1:2])
table$lowerCI = table[,1] - table[,2]*1.96
table$upperCI = table[,1] + table[,2]*1.96
table
#-------------------------------------------------------------------------------------------
#GLMM
library(lme4)
# We assume that the random effect
# changes ONLY for the intercept value for eta (liner predictor)
# Dealing with autocorrelation
# 0) order by vessel, year, month, day, setID
o=order(df1$PSEUDO_VESSEL_ID,
    df1$YEAR_FISHED,
    df1$MONTH_FISHED ,
    df1$SETID)
df1=df1[o,]
head(df1)
# Typically how many sets per day?
x=aggregate(df1$RND_WEIGHT_KGS,
      by=list(df1$PSEUDO_VESSEL_ID,df1$FISHED.doy),
      FUN=length)
summary(x)
# Median of 5 sets per day
# Randomly select one set per vessel per day
# 1) define function to randomly pick a single
#    value from a numeric vector
pickem=function(x) {
```

```
  sample(x,1)
}
# 2) make the numeric vector of sets to keep
keepSets=as.numeric(aggregate(df1$SETID,
                        by=list(df1$PSEUDO_VESSEL_ID,
                              df1$FISHED.doy),
                        FUN=pickem)[,3])
length(keepSets)
# 3) select set records for analysis
df1=df1[df1$SETID%in%keepSets,]
dim(df1)
#------------------------------------------------------------------------------------------
# Simplify model? Use intercept random effect model with data subset
glmm5=  glmer(RND_WEIGHT_KGS ~ log(HOURS_FISHED) +
              fYEAR_FISHED+
              fMONTH_FISHED+
              (1 | PSEUDO_VESSEL_ID),
             data=df1,
             family=Gamma(link="log"))
drop1(glmm5,test="Chisq")
print(summary(glmm5), corr=F)
# Diagnostic plots
par(mfrow=c(2,2))
Y_hat = predict(glmm5)
E = residuals(glmm5, type="deviance")
E_lo = lowess(as.numeric(Y_hat), as.numeric(E))
plot(Y_hat,E); lines(E_lo, col="red")
hist(E)
qqnorm(E); qqline(E)
acf(E)
#------------------------------------------------------------------------------------------
# Create table for word from glmm5
str(summary(glmm5))
out=summary(glmm5)
table=as.data.frame(out$coefficients[,1:2])
table$lowerCI = table[,1] - table[,2]*1.96
table$upperCI = table[,1] + table[,2]*1.96
table
#------------------------------------------------------------------------------------------Snow crab
rawData=read.table("Crab Logbook 2006-2009 Gillis 2015.csv",
               sep=",",
               header=T,
               stringsAsFactors=F)
head(rawData)
str(rawData)
Data = rawData
```

```
# Variable definitions and type conversions
Data$Status[Data$Status=="Autochtones"]="First Nations"
Data$Status=factor(Data$Status)
Data$Year=factor(Data$year)
Data$Fisher=factor(Data$Fisher)
Data$Zone=factor(Data$Zone)
Data$Status=factor(Data$Status)
Data$WofY=factor(Data$weekOfYear, ordered=T)
Data$WofF=factor(Data$WeekOfFishing, ordered=T)
Data$DofY=factor(Data$dayOfYearFished, ordered=T)
Data$Catch=Data$SlipQtyDayKg
Data$logCatch=log10(Data$Catch)
Data$Traps=Data$EffortDay
Data$logTraps=log(Data$EffortDay)
Data$Soak=Data$soaktimeHrs
# Record with 0 soak time are assigned 1 hour
Data$Soak[Data$Soak==0] = 1
Data$logSoak=log(Data$Soak)
str(Data)
# Choose only Area 12 Sets
Data=Data[Data$Zone=="12",]
# Drop Zero Catch records
sum(Data$Catch==0)
Data = Data[Data$Catch>0,]
# Choose Soak Times one week or less
sum(Data$Soak>=168)
Data=Data[Data$Soak<=168,]
# Choose records with 150 traps or less (license limit)
sum(Data$Traps>=150)
Data=Data[Data$Traps<=150,]
dim(Data)
# Data available to use
dim(rawData)[1]-dim(Data)[1]
names(Data)
x=table(Data$Fisher, Data$Year)
# Minimum number of landing records in every year to include?
yearlySetMin = 10
x=x[x[,"2006"] >= yearlySetMin &
   x[,"2007"] >= yearlySetMin &
   x[,"2008"] >= yearlySetMin &
   x[,"2009"] >= yearlySetMin, ]
selectFishers=row.names(x)
length(selectFishers)
DataFishers=Data[Data$Fisher %in% selectFishers,]
dim(DataFishers)
table(DataFishers$Year)
```

```
table(Data$Year)
rm(x)  # housekeeping
str(DataFishers)
# Convert character to POSIX
DataFishers$pDateLanded = strptime(DataFishers$DateLanded,
                          format = "%m/%d/%Y",
                          tz="GMT")
DataFishers$pDateSailed = strptime(DataFishers$DateSailed,
                          format = "%m/%d/%Y",
                          tz="GMT")
DataFishers$pDateCaught = strptime(DataFishers$DateCaught,
                          format = "%m/%d/%Y",
                          tz="GMT")
DataFishers$TripDur = DataFishers$pDateLanded - DataFishers$pDateSailed
table(DataFishers$Year)
setData=DataFishers
#Remove non-positive values
setData = setData[setData$Catch>0,]
dim(setData)
summary(setData$Catch)
setData=setData[setData$Catch>470,]
dim(setData)
summary(setData$Catch)
qqnorm(log(setData$Catch)); qqline(log(setData$Catch))
table(setData$Year)
hist(setData$logSoak)
hist(setData[setData$logSoak>=0,"logSoak"])
hist(setData$logTraps)
hist(setData[setData$logTraps>=0,"logTraps"])
# Will log Traps to get multiplicative equation
setData$logCatch <- log(setData$Catch)
setData$logTraps <- log(setData$Traps)
# convert factor to numeric
setData$WofY=as.numeric(setData$WofY)
setData$Status=as.numeric(setData$Status)
setData$TripDur=as.numeric(setData$TripDur)
setData$Year=as.numeric(setData$Year)
# Order data by vessel, year, month, day
o=order(setData$Fisher,setData$Year,setData$WofY,setData$DofY,setData$Traps)
x=setData[o,]
head(x)
acf(log(x$Catch), main = "catch") #auto-correlation functions
library(lme4)
# Assume that the random effect
# changes ONLY for the intercept catch for eta (liner predictor)
# Dealing with autocorrelation
```

```
# 0) order by Fisher, year, WofY, DofY, Traps
o=order(setData$Fisher,
        setData$Year,
        setData$WofY ,
        setData$DofY,
        setData$Traps)
setData1=setData[o,]
head(setData1)
# Typically how many Traps per day?
x=aggregate(setData1$Catch,
            by=list(setData1$Fisher,setData1$DofY),
            FUN=length)
summary(x)
# Median of 2 traps per day
#
# Randomly select one trap per Fisher per day
# 1) define function to randomly pick a single
#    value from a numeric vector
pickem=function(x) {
  sample(x,1)
}
# 2) make the numeric vector of traps to keep
keepTraps=as.numeric(aggregate(setData1$Traps,
                               by=list(setData1$Fisher,
                                       setData1$DofY),
                               FUN=pickem)[,2])
length(keepTraps)
# 3) select set records for analysis
setData2=setData1[setData1$Traps%in%keepTraps,]
dim(setData2)
glmm4=  glmer(Catch~ logTraps
                                    + logSoak
                                    + Year
                                    + WofY
                +(1 | Fisher),
              data=setData2,
              family=Gamma(link="log"))
#good no warning
drop1(glmm4,test="Chisq")
print(summary(glmm4), corr=F)
# diagnostic plots
par(mfrow=c(2,2))
Y_hat = predict(glmm4)
E = residuals(glmm4, type="deviance")
E_lo = lowess(as.numeric(Y_hat), as.numeric(E))
plot(Y_hat,E); lines(E_lo, col="red")
```

```
hist(E)
qqnorm(E); qqline(E)
acf(E)
str(summary(glmm4))
out=summary(glmm4)
table=as.data.frame(out$coefficients[,1:2])
table$lowerCI = table[,1] - table[,2]*1.96
table$upperCI = table[,1] + table[,2]*1.96
table
```