Universidade de Évora
ESCOLA DE CIÊNCIAS E TECNOLOGIA

Mestrado em Modelação Estatística e Análise de Dados
Especialização em Modelação Estatística e Análise de Dados

Dissertação

Modelling catch and mortality rates of blue shark captured by the Portuguese longline fleet in the Atlantic Ocean

"Modelação de taxas de captura e mortalidade de tintureira capturada pela frota Portuguesa de palangre de superfície no Oceano Atlântico"

Rui Pedro Andrade Coelho

Orientador:
Professor Doutor Paulo Jesus Infante dos Santos, Univ. Évora

Co-Orientador:
Doutor Miguel Neves dos Santos, IPMA, I.P.

2013

Mestrado em Modelação Estatística e Análise de Dados
Especialização em Modelação Estatística e Análise de Dados

Dissertação

Modelling catch and mortality rates of blue shark captured by the Portuguese longline fleet in the Atlantic Ocean

"Modelação de taxas de captura e mortalidade de tintureira capturada pela frota Portuguesa de palangre de superfície no Oceano Atlântico"

Rui Pedro Andrade Coelho

Orientador:
Professor Doutor Paulo Jesus Infante dos Santos, Univ. Évora

Co-Orientador:
Doutor Miguel Neves dos Santos, IPMA, I.P.

Modelação de taxas de captura e mortalidade de tintureira capturada pela frota Portuguesa de palangre de superfície no Oceano Atlântico

A tintureira (*Prionace glauca*) é um tubarão pelágico relativamente abundante e frequentemente capturado como espécie acessória em pescarias de palangre de superfície. Apesar dos parâmetros biológicos terem já sido relativamente bem estudados, os impactos das pescarias nestas populações são ainda bastante incertos. Assim, o presente estudo pretendeu criar e apresentar modelos para melhor avaliar os impactos da pescaria Portuguesa de palangre de superfície dirigida ao espadarte nas populações de tintureira. Especificamente, o trabalho apresenta modelos relativos à mortalidade durante a operação de pesca utilizando modelos binomiais, recorrendo a abordagens com modelos lineares generalizados e equações de estimação generalizadas; e modelos relativos às taxas de captura usando modelos lineares generalizados e modelos mistos generalizados. Os resultados apresentados podem agora ser usados para prever as taxas de captura e de mortalidade da tintureira em diferentes cenários de pesca, contribuindo assim para uma melhor compreensão dos impactos desta pescaria nesta espécie.

Modelling catch and mortality rates of blue shark captured by the Portuguese longline fleet in the Atlantic Ocean

The blue shark (*Prionace glauca*) is a relatively abundant and wide ranging pelagic shark, commonly captured as bycatch in pelagic longline fisheries. While it is a species with relatively known biological parameters, the impacts of the fisheries in their populations is still largely unknown. Therefore, the present study aimed to create and present models for understanding the impacts of the Portuguese pelagic longline fishery targeting swordfish, in this shark species. Specifically, the work focused on modeling two different fisheries aspects, namely the at-haulback mortality using binomial models with generalized linear models and generalized estimation equations; and the catch rates using generalized linear models and generalized mixed models. The results presented can now be used to predict the catch and mortality rates under various fishing scenarios, and contribute to a better understanding of the impacts of the fishery in this shark species.

AGRADECIMENTOS / ACKNOWLEDGMENTS

Ao Professor Doutor Paulo Infante (Universidade de Évora), orientador desta tese de Mestrado, pela grande ajuda ao longo não só da tese mas também ao longo de todo este curso de Mestrado. Tenho a certeza de ter encontrado um excelente matemático e de ter feito um bom amigo, e de que certamente continuaremos a trabalhar em conjunto e parceria no futuro.

Ao Doutor Miguel Neves dos Santos (IPMA, I.P), orientador desta tese de Mestrado, pelas fundamentais contribuições ao nível do entendimento do modo de operação e especificidades da pesca de palangre de superfície. Esta contribuição foi fundamental não só para esta tese de Mestrado, mas também em períodos anteriores ao mestrado, desde que em 2008 comecei a focar a minha investigação na biologia e pescarias dos grandes migradores pelágicos.

Ao Instituto Português do Mar e da Atmosfera (IPMA, I.P.), na pessoa do seu Presidente, Professor Doutor Miguel Miranda, pela oportunidade de realizar este trabalho.

Aos restantes professores do curso de Mestrado em Modelação Estatística e Análise de Dados, por terem preparado e leccionado um curso de Mestrado muito interessante, útil, e com bastantes aplicações práticas em diversas áreas do conhecimento.

Aos meus colegas de curso de Mestrado pelo permanente companheirismo, ajuda e amizade ao longo deste período.

A minha família, e sobretudo aos meus pais, por sempre terem acreditado e apoiado os meus projetos.

À minha esposa Joana, por sempre ter aturado e apoiado a 100% mais este meu projeto, da mesma forma e com o mesmo entusiasmo com que sempre me tem apoiado em todos os outros projetos, por mais extravagantes e despropositados que pareçam.

CONTENTS

CHAPTER I.   GENERAL INTRODUCTION

I.1. General introduction to the Chondrichthyan fishes

Chondrichthyan fishes (sharks, rays, skates and chimeras) are an old animal group that first appeared during the Devonian period, with the earliest evidence in the fossil record dating from 409-363 million years (Ma) ago (Compagno, 2005). They survived several major mass extinction episodes, including, for example, the Cretaceous–Paleogene mass extinction event 65.5 Ma that caused the extinction of the dinosaurs. The modern Chondrichthyans living today in the world Oceans derived from the forms that were present during the Mesozoic period, 245-65 Ma (Grogan and Lund, 2004).

Chondrichthyans are characterized by an internal skeleton formed by flexible cartilage, without the formation of true bone in their skeletons, fins or scales. Other characteristic that further separate the Chondrichthyans from other fishes are the presence of claspers in males (sexual organs used to inseminate females) that are formed by the mineralization of the endoskeleton tissue along the pelvic fins (Grogan and Lund, 2004). It is accepted that the class Chondrichthyes is a monophyletic group (Compagno *et al.*, 2005) that is divided into two sister taxa: the subclass Elasmobranchii that groups sharks, rays and skates and the subclass Holocephali that groups the chimaeras (Table I.1). Within this group, the Elasmobranchs are recognized from their multiple (5 to 7) paired gill openings on the sides of the head, while the Holocephalans have a soft gill cover with just a single opening on each side of the head that protects the 4 pairs of gill openings (Compagno *et al.*, 2005). There are currently *circa* 1180 Chondrichthyan species described worldwide (White and Last, 2012), including approximately 480 species of sharks, 650 batoids and 50 chimaeras.

Chondrichthyan fishes occupy a wide range of habitat types, including freshwater rivers and lake systems, inshore estuaries and lagoons, coastal waters, the open sea, and the deep ocean. Although sharks are generally thought of being wide-ranging, only a few (including some commercially important species) make oceanic migrations. Overall, some 5% of Chondrichthyan species are oceanic (found offshore and migrating across ocean basins), 50% occur in shelf waters down to 200 m depth, 35% are found in

deeper waters from 200 to 2000 m, 5% occur in fresh water, and 5% have been recorded in several of these habitats (Camhi *et al.*, 1998).

Table I.1: Extant orders of the class Chondrichthyes, according to Compagno (2001) and Compagno *et al.* (2005).

Subclass	Superorder	Order	Common name
Holocephali		Chimaeriformes	Chimaeras
Elasmobranchii	Squalomorphii	Hexanchiformes	Cow and frilled sharks
		Squaliformes	Dogfish sharks
		Squatiniformes	Angel sharks
		Pristiophoriformes	Saw sharks
		Rajiformes	Batoids
	Galeomorphii	Heterodontiformes	Bullhead sharks
		Orectolobiformes	Carpet sharks
		Lamniformes	Mackerel sharks
		Carcharhiniformes	Ground sharks

I.2. The exploitation of Chondrichthyans with emphasis on the pelagic sharks

In recent years elasmobranch fishes have become relatively important fisheries resources, with a substantial increase in fishing effort worldwide (Vannuccini, 1999; Barker and Schluessel, 2005). However, elasmobranchs have not traditionally been highly priced products, with the exception of the fins of some species that are marketed at very high prices in oriental markets for shark fin soup (Bonfil, 1994; Clarke *et al.*, 2007). The exploitation of elasmobranch resources has been attributed in part to fisheries specifically targeting elasmobranchs (e.g. Campbell *et al.*, 1992; Castillo-Geniz *et al.*, 1998; Francis, 1998; Hurley, 1998; McVean *et al.*, 2006; Cartamil *et al.*, 2011) but perhaps more importantly to the bycatch of fisheries targeting other species (e.g. Stevens, 1992; Buencuerpo *et al.*, 1998; McKinnell and Seki, 1998; Francis *et al.*, 2001; Beerkircher *et al.*, 2003; Coelho *et al.*, 2003; Megalofonou *et al.*, 2005; Coelho and Erzini, 2008; Belcher and Jennings, 2011; Coelho *et al.*, 2012a). Game fishing also has some impact on elasmobranch fishes, especially on the large pelagic species (e.g. Stevens, 1984; Pepperell, 1992; Campana *et al.*, 2006, Lynch *et al.*, 2010).

Even though elasmobranchs are currently impacted by commercial and recreational fisheries, there is still limited information about these species life cycles, biological parameters, movement patterns and habitat utilization, and in the general impact of fisheries in their populations. Elasmobranch fishes have typically K-strategy life cycles, characterized by slow growth rates and reduced progeny, with maturity occurring late in their life cycle (Smith *et al.*, 1998; Stevens *et al.*, 2000; Cortés, 2000; Cortés, 2007). This low fecundity and relatively high survival rate of newborns suggests that there is a strong relationship between the number of mature females in the population and the new recruits for the next cohort, meaning that the success of the future generation is mainly dependant on the present mature population abundance (Ellis *et al.*, 2005).

While the total worldwide marine fishes landings seem to have reached a plateau in the late 1980's, elasmobranch catches increased progressively since the 1950's until the early 2000's, followed by a decreasing trend for the more recent years (Figure I.1). However, and even though the marine fish catches seem to have remained relatively stable since the late 1980's, the fisheries have shifted in these last decades from catching mainly long lived high trophic level fishes, towards catching more short lived, low trophic level invertebrates and small planktivorous pelagic fishes (Pauly *et al.*, 1998; Pauly and Palomares, 2005). This effect, originally called "fishing down the marine food web" by Pauly *et al.* (1998) shows that the marine ecosystems top predators (such as the sharks) are the first ones to suffer from overfishing and population declines. Indeed, most elasmobranchs are predators at, or near the top of the marine food webs (Cortés, 1999), and are extremely important for the entire ecosystems balance, by regulating not only their direct main preys, but also second and third degree non-prey species through the trophic linkages (Schindler *et al.*, 2002). The effects of the removal of such predators from the marine ecosystems are difficult to foresee, but may be ecologically and economically significant, and may persist over long time periods (Stevens *et al.*, 2000).
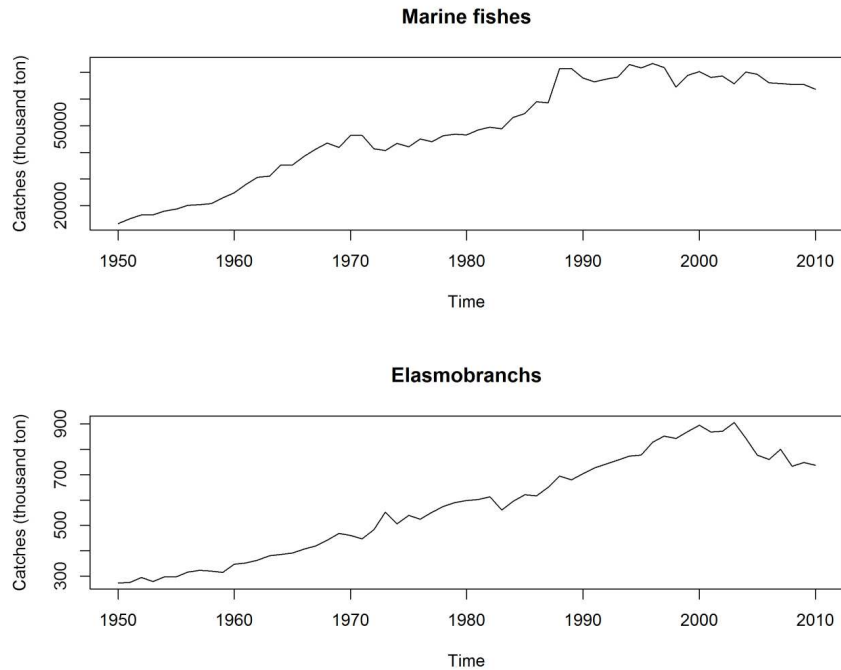
Figure I.1: Global capture of marine fishes (top) and elasmobranchs (bottom) from 1950 to 2010. Data from FAO FIGIS data collection (FAO, 2012)

Up until the 1980's, elasmobranch fisheries were generally unimportant small fisheries, with generally a low commercial value. Traditionally, these elasmobranch fisheries of the past were multi-specific fisheries that caught several species of elasmobranchs depending on the region and season of the year. There was little interest in these fisheries, mainly due to their relatively small scale and low commercial value. Bonfil (1994) reported that cartilaginous fishes were a minor group which contributed with an average of 0.8% of the total world fishery landings between 1947 and 1985, while bony fishes such as clupeoids, gadoids and scombroids, accounted for 24.6%, 13.9% and 6.5%, respectively. In the last decades, however, the declining catches per unit effort (CPUE) and rising prices of traditional food fishes, along with the growing market for shark fins for the oriental markets, have made the previously underutilized elasmobranchs increasingly important resources (Castro *et al.*, 1999).

The history of elasmobranch fisheries worldwide indicates, however, that these resources are usually not sustainable. Most elasmobranch targeted fisheries have been characterized by "*boom and burst*" scenarios, where an initial rapid increase of the exploitation and catches is followed by a rapid decline in catch rates and eventually a complete collapse of the fishery (Stevens *et al.*, 2000). Bonfil (1994) and Shotton

(1999) provided reviews of world elasmobranch fisheries and included examples of situations where commercial catches have been declining, such as in the northeast Atlantic and Japan, and examples of situations of high concern such as in India. Baum *et al.* (2003) stated that the northwest populations of large pelagic sharks including the scalloped hammerhead, *Sphyrna lewini*, and the threshers *Alopias vulpinus* and *A. superciliosus*, have declined by more than 75% over the last 15 years, and even though the values presented in Baum *et al.* (2003) seem to have been severely overestimated (Burgess *et al.*, 2005), there is consensus that there are currently causes for concern.

However, and even though overexploitation and population collapses is the most common scenario in elasmobranch fisheries, Walker (1998) demonstrated that elasmobranch stocks can be harvested sustainably and provide for stable fisheries when carefully managed. Some species such as the tope shark, *Galeorhinus galeus*, the sandbar shark, *Carcharhinus plumbeus*, the great white shark, *Carcharodon carcharias* and several species of dogfishes (order Squaliformes) have very low productivity and cannot withstand high levels of fishing, whereas other species such as the gummy shark, *Mustelus antarcticus*, the Atlantic sharpnose shark, *Rhizoprionodon terraenovae*, the bonnethead, *Sphyrna tiburo* and the blue shark, *Prionace glauca* have higher productivity and can support higher levels of fishing mortality (Walker, 1998).

Within the industrial oceanic fisheries such as longlines, driftnets and purse seines, the pelagic longlines are responsible for most of the captures of oceanic sharks at a global level, which are usually captured during the fishing operations that target swordfish and tunas (Aires-da-Silva *et al.*, 2008). Several pelagic shark species are frequently caught in those oceanic longline fisheries, but the two most important and abundant are the blue shark, *Prionace glauca*, and the shortfin mako, *Isurus oxyrhynchus*. In the case of the Portuguese fishery, those two species together can account for more than 50% of the total oceanic longline fishery catch, and can represent more than 95% of the total elasmobranch catch (Coelho *et al.*, 2012a).

I.3. The studied species, blue shark (*Prionace glauca*)

The blue shark (*Prionace glauca*) (Figure I.2) is one of the most wide ranging of all sharks, found throughout tropical and temperate seas from latitudes of about 60°N to

50°S (Last and Stevens, 2009) (Figure I.3). It is a pelagic species mainly distributed from the sea surface to depths of about 350 m, even though deeper dives of up to 1000m have been recorded (Campana, *et al.*, 2011). The blue shark is an oceanic species capable of large scale migrations (Queiroz *et al.*, 2005; Silva *et al.*, 2010; Campana *et al.*, 2011), but it can also occasionally occur closer to inshore waters, especially in areas where the continental shelf is narrow (Last and Stevens, 2009).
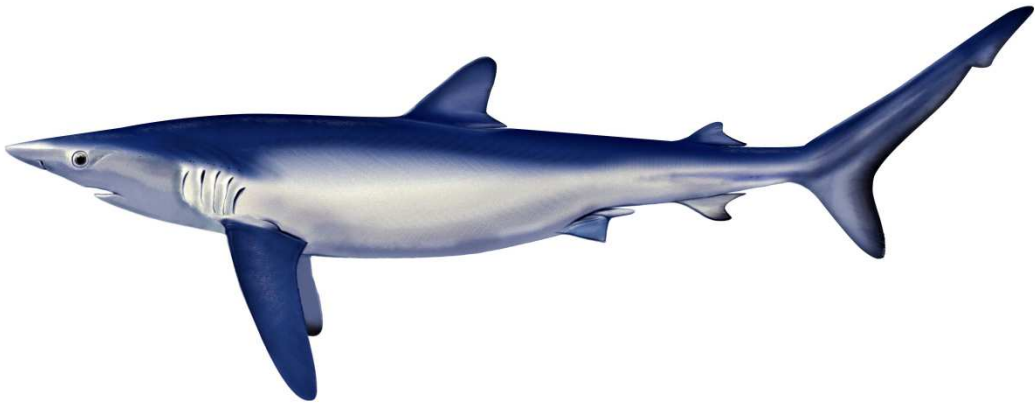
Figure I.2: The blue shark, *Prionace glauca* (Drawing by: João T. Tavares/Gobius).
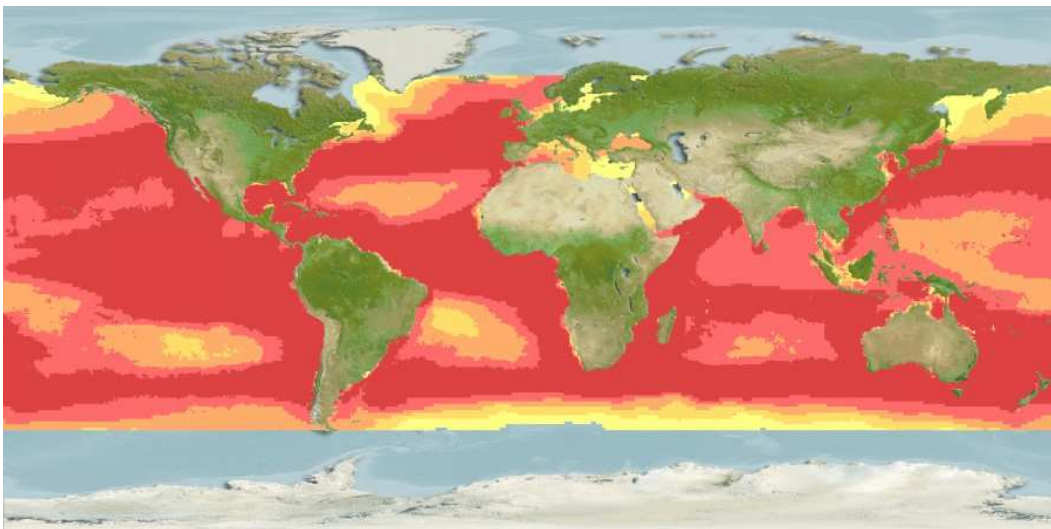
Figure I.3: Global distribution map for the blue shark, *Prionace glauca*. The color scale represents the relative probabilities of occurrence, with red and yellow representing higher and lower probabilities of occurrence, respectively. Map generated from Fishbase (Froese and Pauly, 2012) using AquaMaps, a presence-only species distribution model (Ready *et al.*, 2010).

The blue shark reaches a maximum size of about 380 cm total length (TL), and size at 50% maturity for the Atlantic has been estimated at 218 cm TL for males and

221 cm TL for females (Pratt, 1979). The blue shark is a placental viviparous shark, and shows a relatively high fecundity within the elasmobranchs, producing an average of 35 pups per litter (Zhu *et al.*, 2011), with the maximum litter size recorded being 135 pups, after a gestation period of 9-12 months (Compagno, 1984; Castro and Mejuto, 1995; Snelson *et al.*, 2008). The pups are born at 35-50 cm TL, and the reproductive cycle has been reported as seasonal in most areas, with the young being born usually in the spring and summer (Pratt, 1979; Stevens, 1984; Nakano, 1994; Hazin *et al.*, 1994). Age and growth studies have suggested that longevity is of about 20 years, with the males maturing at 4-6 and females at 5-7 years of age (Stevens, 1975; Cailliet *et al.*, 1983; Nakano, 1994; Skomal and Natanson, 2003; Lessa *et al.*, 2004; Blanco-Parra *et al.*, 2008; Megalofonou *et al.*, 2009a). The diet of the blue shark consists mainly of small pelagic fishes and cephalopods, particularly squid (Vaske Jr. *et al.*, 2009; Markaida and Sosa-Nishizaki, 2010; Preti *et al.*, 2012). However, invertebrates such as pelagic crustaceans, small sharks, and seabirds have also been reported to be taken as food (Compagno, 1984).

Blue sharks are a highly migratory oceanic species, with complex movement patterns and spatial structure probably related to the reproduction cycles and prey distribution (Montealegre-Quijano and Vooren, 2010; Tavares *et al.*, 2012). Some tagging studies have shown extensive movements of blue sharks in the Atlantic, with numerous trans-Atlantic migrations probably accomplished by using the major oceanic current systems (Stevens, 1976; Stevens 1990; Queiroz *et al.*, 2005; Silva *et al.*, 2010; Campana *et al.*, 2011). At least in the north Atlantic, data on the distribution, movements and reproductive behavior seems to suggest a complex reproductive cycle, involving major oceanic migrations associated with mating areas in the north-western Atlantic and pupping areas in the north-eastern Atlantic (Pratt, 1979; Stevens, 1990).

The blue shark is possibly the most abundant of all pelagic shark species, and even though it can be captured by a variety of fishing gears, most captures take place as bycatch in pelagic longlines targeting tunas and swordfish (Aires-da-Silva *et al.*, 2008; Stevens, 2009). In the Atlantic Ocean, the management of the oceanic tuna and tuna-like species (including pelagic sharks) is a mandate of ICCAT, the *International Commission for the Conservation of Atlantic Tunas*. ICCAT maintains the catch records from those fisheries (Figure I.4) and carries out stock assessments and other research initiatives for determining their vulnerability status.
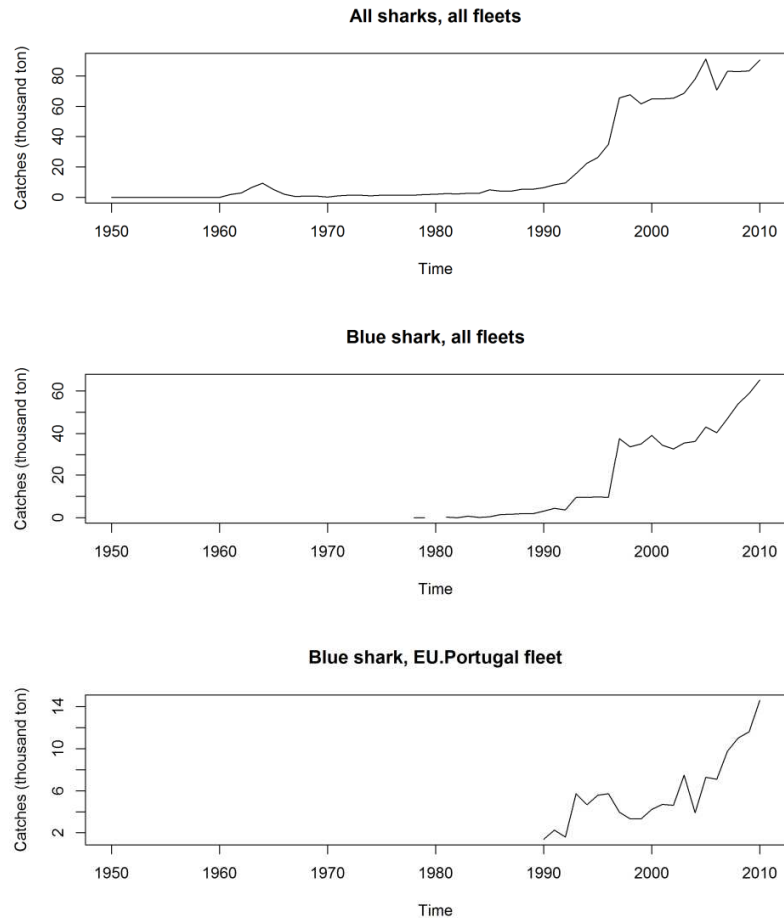
Figure I.4: Nominal catches of all pelagic shark species by all oceanic fleets in the Atlantic Ocean (above), blue sharks captured by all fleets (center) and blue shark captured by the Portuguese fleet (below). Data from ICCAT Task1 (nominal catch information) database (ICCAT, 2012a).

Within the ICCAT scientific work, an Ecological Risk Assessment (ERA) was carried out for priority species of pelagic sharks in the Atlantic in 2010 (Cortés *et al.*, 2010), with that analysis currently being updated with more recent information (Cortés *et al.*, 2012). With both analyses it was demonstrated that most pelagic sharks have exceptionally limited biological productivity and, as such, can be overfished even at very low levels of fishing mortality, with the blue shark in particular shown to have an intermediate vulnerability. More recently, and for the Indian Ocean (managed by IOTC, the *Indian Ocean Tuna Commission*), an ERA analysis was also conducted for pelagic and some coastal shark species (Murua *et al.*, 2012) and similar results were obtained for the blue shark, also characterized for having a relatively higher productivity but also a high susceptibility to longline fisheries, making it a species with an overall

intermediate level of vulnerability. The last blue shark stock assessment for the Atlantic Ocean was carried out by ICCAT in 2008 (ICCAT, 2009), and although a high level of uncertainty was reported in the models, the results showed that the current biomass was believed to be above the biomass that would support Maximum Sustainable Yield (MSY), and the harvest levels were believed to be below the Maximum Fishing Mortality (F) at MSY.

I.4. Challenges in modeling Chondrichthyans bycatch

The main goal of fisheries science and stock assessments are to inform decision makers on the potential consequences of different management actions, using the best available scientific information and data (Ludwig *et al.*, 1993; Hilborn and Walters, 1992; McAllister *et al.*, 1999; Quinn and Deriso 1999; Hilborn, 2006). The increasing concerns on the vulnerability of elasmobranch species to fisheries has lead, in recent years, to an increased interest on assessing the conservation status and carrying out stock assessments for those populations (McAllister *et al.*, 2008). In general and when compared to other fishes, the current available information for assessing the status of elasmobranch populations is usually very poor, and as such most elasmobranchs are today in what is called data-poor situations. This is a situation characterized by little available information in terms of their biology (e.g. age and growth, reproduction, ecology, migratory movements), but also in terms of reliable time-series of their historical abundance and fisheries catches.

One commonly used analytical method that has been applied to some shark populations are demographic methods, which are useful particularly because they rely primarily on biological aspects (Cortés, 2002; Mollet and Cailliet, 2002), rather than on the historical catches or indexes of abundance. The inputs required are basic population dynamics parameters, such as the rate of survivorship at each age/stage, the duration of each life stage (in case of stage-based approaches), and the fecundity or number of newly born offspring produced per female at each age/stage. One important limitation on those methods is that they assume that there is no density-dependence, and that the estimated parameters are those of theoretical populations under stable conditions. Typical approaches for studying species demography include life table analysis (e.g.

Cailliet, 1992; Cortés, 1995) and matrix algebra analysis (e.g. Aires-da-Silva and Gallucci, 2007; Smith *et al.,* 2008). The most important output of those methods for fisheries management is the estimation of *r*, the intrinsic rate of population increase under a stable condition and assuming density-independence, as this parameter provides an indication of the population resilience to exploitation.

For more elaborate and data-intensive stock assessment methods, one common approach used for some shark species are surplus production methods (e.g. shortfin mako assessment carried out by ICCAT, 2012b), that uses information from total catches and relative indexes of abundance of the stocks over time. Ideally, these indexes of abundance should be based on fishery-independent datasets, collected for example during scientific surveys using statistically adequate protocols (e.g. random sampling over predetermined strata such as area, season, year, etc). However, these type of data are very difficult to obtain and costly in the pelagic realm, as the sampling collection would have to occur in the high seas and cover very wide geographical areas. Therefore, and particularly when dealing with pelagic bycatch species such as sharks, the data available is usually based on fishery-dependent datasets, collected by commercial fishing vessels while operating during their normal fishing operations. Because of this, for calculating time series with the relative indexes of abundance useful for stock assessment, it is first necessary to adjust the data for the impacts of factors other than the changing abundances of the species over time. There are several methods for achieving this, but a recent common approach is to use statistical models such as Generalized Linear Models (GLM) to build the time series of the species abundance over time that only reflects the changes in the abundance, and where other effects inherent to the fishery-dependence itself have been removed. A good revision on the use of GLM for standardizing fishery-dependant datasets for stock assessment purposes was presented by Maunder and Punt (2004). For addressing the lack of independence in the data, alternative approaches such as Generalized Linear Mixed Models (GLMM) that use random effects on some variables allowing the introduction of variability (McCulloch and Searle, 2001; Bolker *et al.*, 2009), and Generalized Estimating Equations (GEE) that introduce a dependence structure in the data (Zeger and Liang, 1986; Zeger *et al.,* 1988), can be used.

Another potential issue and challenge when modeling data from shark populations is that the datasets of bycatch species often have some (sometimes many) fishing sets

with zero catches. Those represent the fishing sets that existed (have an associated effort), but resulted in zero catches for the species of concern, and this poses a special mathematical problem in terms of modeling. For example, one possible and common way of modeling catch rate data is to use GLM with a *log* link and some continuous distribution (e.g. Gaussian, Gamma), but in datasets with zeros adjustments need to be made for accommodating those observations, given that the *log* of zero is undefined. Possible solutions for those observations have ranged from simple solutions like adding a small constant to the observed data, to more complex approaches like zero-inflated models. Adding a small constant to the data was a common approach in the past, but as mentioned by Campbell (2004) the value of the constant to be added can be somewhat arbitrary and that constitutes a problem as bias are introduced in the analysis. Still, when the proportion of zeros in the datasets are low (<5-10% of the data), this approach is still commonly used in fisheries science. Besides this strategy, Maunder and Punt (2004) summarized other three classes of methods that can handle zero observations, specified as: 1) statistical distributions that allow for zero observations (e.g. Poisson, Negative Binomial, Tweedie); 2) methods that inflate the expected numbers of zeros (zero-inflated models); and 3) the delta-lognormal approach (Lo *et al.*, 1992) that combines two separate models, usually one binomial model for modeling the proportion of positives and one continuous distribution model for modeling the predicted values conditional to the positive observations.

I.5. General objectives of the study with a note on the dissertation style

Given the general lack of information on the fisheries of the blue shark captured as bycatch in pelagic longline fisheries, and the increasingly importance of this species as a marine fisheries resource, there was a need to carry out a study focusing this species and its impacts in pelagic longline fisheries. The specific objectives of the present study were to:

1) Provide a general introduction to the Chondrichthyan fishes, their biology and susceptibility to fishing mortality, with a particular emphasis on the oceanic sharks and especially the blue shark (Chapter I);

2) Model the hooking mortality of the blue shark captured in the Portuguese longline fishery in the Atlantic Ocean (Chapter II);

3) Model the catch rates of the blue shark captured in the Portuguese longline fishery in the South Atlantic Ocean (Chapter III);

Each of the following chapters (specifically chapters II and III) of this thesis has been written in a paper-style format, suitable and appropriate to be published in a scientific journal. Each of those chapters constitutes a complete study and can be read independently of the others. At the beginning of each chapter information regarding that particular chapter publication status is given. Tables and figures appear in the text inside each chapter, but all acknowledgements have been compiled at the beginning of the thesis and all references have been compiled in a final section. A final Annex section is provided with a compilation of the R-language code that was produced and used in this thesis.

CHAPTER II.   MODELING AT-HAULBACK MORTALITY OF BLUE SHARKS CAPTURED IN A PELAGIC LONGLINE FISHERY IN THE ATLANTIC OCEAN.[1]

II.1. Introduction

In the Atlantic Ocean several pelagic shark species are commonly bycatch on pelagic longline fisheries (e.g. Buencuerpo *et al.*, 1998; Petersen *et al.*, 2009; Simpfendorfer *et al.*, 2002) but still, information on their life history, population parameters and the effects of fisheries on these populations is limited. Generally, elasmobranchs have K-strategy life cycles, characterized by slow growth rates and long lives, and reduced reproductive potential with few offspring and late maturity. The natural mortality rates are usually low, and increased fishing mortality may have severe consequences on these populations, with population declines occurring even at relatively low levels of fishing mortality (Smith *et al.*, 1998; Stevens *et al.*, 2000). Of the several elasmobranch species caught in surface pelagic longline fisheries, the blue shark, *Prionace glauca*, is the most frequently caught species (e.g. Coelho *et al.*, 2012a).

Previous studies have focused on elasmobranch mortality during fishing operations, but most were carried out for coastal species caught in trawl fisheries. Those include the studies by Mandelman and Farrington (2007) for the spurdog (*Squalus acanthias*) and Rodríguez-Cabello *et al.* (2005) for the small-spotted catshark (*Scyliorhinus canicula*). For pelagic elasmobranchs caught in pelagic fisheries in the NW Atlantic Ocean, Campana *et al.* (2009) analyzed blue sharks captured by the Canadian fleet and studied both the short term mortality (recorded at-haulback) and the longer term mortality (recorded with satellite telemetry). Also for the NW Atlantic, Diaz and Serafy (2005) worked with data from the U.S. pelagic fishery observer program and analyzed factors affecting the live release of blue sharks.

Knowledge on the at-haulback mortality can be used to evaluate conservation and management measures that include the prohibition to retain particular vulnerable

---

species, such as those recently implemented by some tuna Regional Fisheries Management Organizations (tRFMOs). In particular and for the Atlantic Ocean, the *International Commission for the Conservation of Atlantic Tunas* (ICCAT) has recently implemented mandatory discards for the bigeye thresher (ICCAT Rec. 09-07), the oceanic whitetip (ICCAT Rec. 10-07), hammerheads (ICCAT Rec. 10-08) and silky sharks (ICCAT Rec. 11-08). However, important parameters, such as the at-haulback fishing mortality (recorded at time of fishing gear retrieval), remain largely unknown and therefore the efficiency of such measures also remains unknown. Even considering that all specimens of these particular species are now being discarded, fishing mortality is still occurring due to at-haulback mortality, as part of the catch is already dead at time of fishing gear retrieval and is therefore being discarded dead.

At-haulback mortality studies are also important as they can be incorporated into stock assessments, such as the study by Cortés *et al.* (2010), which used an ecological risk assessment analysis for eleven species of elasmobranchs captured in pelagic longlines in the Atlantic Ocean. With this analysis, both the susceptibility and the productivity of each species are analyzed in order to rank and compare their vulnerability to the fishery. One of the parameters that can be included in the susceptibility component is the probability of survival after capture, which can in part be inferred from the mortality at-haulback.

This study had two main objectives:

1) to compare the use of Generalized Linear Models (GLM) and Generalized Estimation Equations (GEE) for predicting the at-haulback mortality of blue sharks captured in the Portuguese pelagic longline fishery in the Atlantic Ocean targeting swordfish and,

2) to identify variables that are significant and influence the blue shark at-haulback mortality rates.

II.2. Material and Methods

II.2.1. Data collection

Data for this study was collected by fishery observers from the *Portuguese Institute for Sea and Atmospheric Research* (IPMA, I.P.) that were placed onboard Portuguese longliners targeting swordfish along the Atlantic Ocean. Data was collected between August 2008 and December 2011. During that period, information from a total of 762 longline sets corresponding to 1,005,486 hooks was collected. The study covered a wide geographical area (from both hemispheres) of the Atlantic Ocean (Figure II.1).
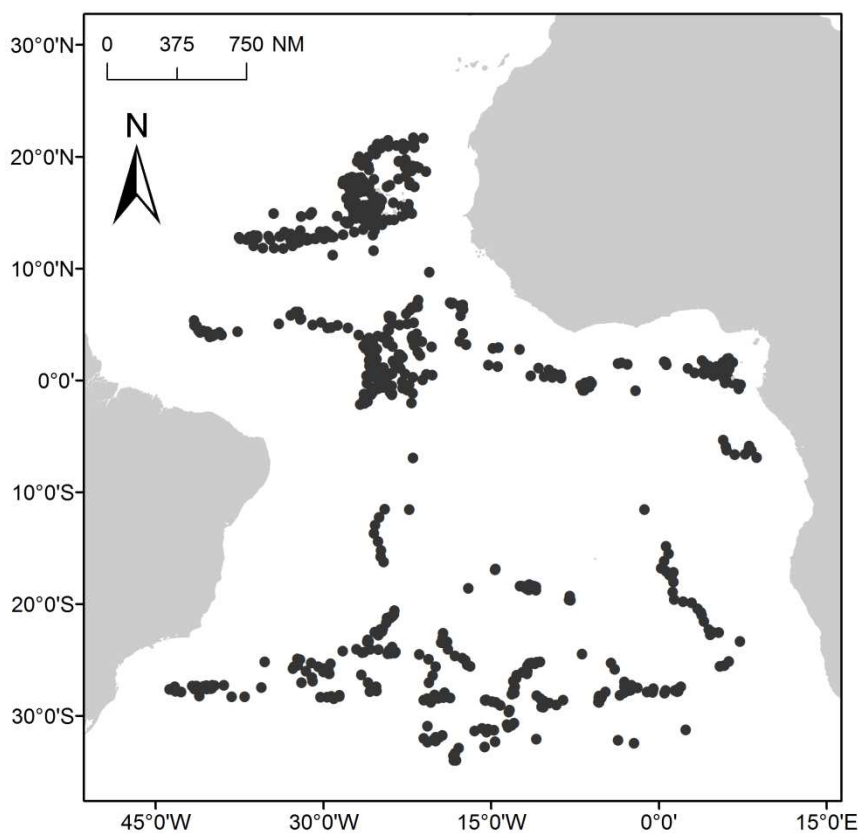
Figure II.1: Location of the longline fishing sets analyzed in this study along the Atlantic Ocean. The scale bar is represented in nautical miles (NM).

For every specimen that was caught, onboard fishery observers recorded the species, specimen size (FL, fork length measured to the nearest lower cm), sex, at-haulback condition (alive or dead at time of fishing gear retrieval), fate (retained or discarded), and the condition if discarded (alive or dead at time of discarding). For each

longline set carried out some additional information was recorded, including date, geographic location (coordinates: latitude and longitude), number of hooks deployed in the set, and branch line material used (monofilament or wire). Additional variables relative to the fishing sets that were calculated *a posteriori* included the Sea Surface Temperature (SST), which was interpolated from satellite data using the known date and location of each fishing set. The algorithm used to interpolate SST data followed the methods described by Kilpatrick *et al.* (2001), and was applied using the Marine Geospatial Ecology Tools (MGET) developed by Roberts *et al.* (2010).

II.2.2. *Preliminary data analysis*

The length frequency distribution of male and female blue sharks captured was analyzed, and compared with a 2-sample Kolmogorov–Smirnov test and a Mann Whitney rank sum test. Those non-parametric tests were chosen after calculating the skewness and kurtosis coefficients for the data, and confirming that the data was non-normal with a Lilliefors test. The proportions of dead and alive blue sharks were calculated for each level of each categorical covariate (trip, sex, year, quarter, vessel, branch line material), and the differences in the proportions were compared with contingency tables and Chi-square statistics (using Yates' continuity correction in the cases of 2x2 tables). For this preliminary analysis, the continuous variables FL, latitude, longitude and SST were categorized by their quartiles.

II.2.3. *Statistical Modeling*

Generalized Linear Models (GLM) and Generalized Estimation Equations (GEE) were used to model blue shark at-haulback mortality, and compare the odds of a shark being dead at-haulback given the various variables considered. The response variable was the condition of the specimens at time of haulback (Y_i: binominal variable, i.e., dead or alive), and for this study we considered that the event occurred if the shark died during the fishing operation. Therefore, the response variable was coded with 1 for sharks dead at-haulback and coded with 0 for sharks alive at-haulback.

Each captured shark (Y_i) follows a Bernoulli distribution with p_i (probability of success *versus* dying at-haulback = π_i), and can be specified as:

$$Y_i \sim B(1, \pi_i)$$

With the expected value and the variance defined by:

$$E(Y_i) = \pi_i$$

$$Var(Y_i) = \pi_i \times (1 - \pi_i)$$

The relationship (link function) between the mean value of Y_i and the model covariates considered for this model was the *logit*, and the model was therefore defined by:

$$logit(\pi_i) = log\left(\frac{\pi_i}{1 - \pi_i}\right) = \beta_0 + \beta_1 x_{1,i} + \beta_2 x_{2,i} + \cdots + \beta_k x_{k,i}$$

Where x_i are the model variables and β are the coefficients that were estimated by maximum likelihood.

The explanatory variables initially considered for the model were the specimen size (FL in cm), sex (male or female), fishing location (latitude and longitude in decimal degrees), year (2008 to 2011), quarter of the year (1 = January to March, 2 = April to June, 3 = July to September and 4 = October to December), vessel identity (two vessels involved in the study), branch line material (wire or monofilament) and SST (decimal degrees in ºC). Some potential additional variables were not considered due to being unbalanced or correlated with other variables, such as the month with quarter of the year, and fishing trip with vessel.

The first modeling approach was carried out with GLM. The univariate significance of each explanatory variable was determined by the Wald statistic and with likelihood ratio tests, comparing each univariate model with the null model. The significant variables were then used to construct a simple effect multivariate GLM, with the non-significant variables (at the 5% level) eliminated consecutively from the model. The significance of each variable was determined by the Wald statistic and by an analysis of deviance table. At this stage, the variables had been eliminated in the first step were further tested, in order to determine an eventual significance within the

framework of a multivariate model, as recommended by Hosmer and Lemeshow (2000). Once a final multivariate simple effects model using only significant variables was obtained, each pair of possible first degree interactions between variables was tested. The interactions were considered for inclusion in the final model if significant at the 1% level both with the Wald statistic, and with likelihood ratio tests comparing the models with and without the interaction.

The GLM assumptions in terms of both the continuous and categorical explanatory variables were assessed. Regarding the continuous variables, GLM have the assumption that those variables are linear with the linear predictor (in this case the *logit*) and such linearity was assessed with the method of discretizing the continuous variables by the quartiles as described by Hosmer and Lemeshow (2000), and by analyzing GAM plots. If transformations were required, then the best possible solution was estimated with multivariate fractional polynomials and the transformed variables were used in the models instead of the original values, following the method developed by Royston and Altman (1994) and recommended by Hosmer and Lemeshow (2000). Regarding the categorical variables, GLM assume that all levels of the categories have sufficient information in the binomial response to allow contrasts in the data and achieve model convergence. These assumptions follow the contingency tables and Chi-square tests assumptions, in which the contingency tables should not have cells with zero values, or more than 25% of the cells with predicted values lower than 5. These assumptions were validated by building contingency tables for all categorical variables that were considered.

Another assumption in the GLM modeling approach is that the data in the sample should be independent, in this case that the Y_i correspond to a succession of independent Bernoulli trials. Given that the data used in this study is fisheries-dependant data, it is plausible to consider that this assumption was not validated. Therefore an alternative modeling approach with Generalized Estimation Equations (GEE) was considered as this allows for a working correlation to be estimated within the data. Within this GEE model framework, the fishing set was considered as the grouping variable, meaning that the data could be considered to be clustered and not independent within each fishing set. This allowed for a model formulation in which the blue shark at-haulback mortality data recorded within each fishing set carried out by each particular vessel in each particular fishing trip did not require the assumption of independence. With this GEE model

formulation, the correlation structure of the data within each set was assumed to be of the type exchangeable, as this seems to be the most adequate correlation structure for clustered data (Halekoh *et al.*, 2006).

With the final model estimated, examples of model interpretation were presented. One parameters that is important to interpret in biological terms in the specimen size, and therefore the probabilities of a shark dying at-haulback with varying specimen sizes were calculated. Additionally, the odds-ratios for increasing specimen sizes by 10cm FL (also calculated along the range of shark sizes in the sample), were also calculated and presented. The probabilities were calculated as the *inverse-logit* function of the final equations considered, and the odds-ratios were calculated as the exponential values of the differences (in 10cm FL sizes) in the *logits*. For this specific example, the variables that were interacting with FL were considered to be on their baseline levels.

II.2.4. Diagnostics and goodness-of-fit

A residual analysis using Pearson and Deviance residuals was used to search for outliers, and the Cooks distances and DfBetas were used to identify eventual values with influence in the estimated parameters of the models. Model goodness-of-fit was assessed with the Hosmer and Lemeshow statistic that groups the observations into 10% quantiles (deciles) according to their predicted values, and uses a chi-square test for comparing the observed versus predicted values in each group (Hosmer and Lemeshow, 2000). Additionally, the Nagelkerke coefficient of determination (R^2) (Nagelkerke, 1991) was also calculated. The discriminative capacity of the models was determined by the Area Under the Curve (AUC) value of the Receiver Operating Characteristic (ROC) curves, with the determinations of the model sensitivity (capacity to correctly detect the event = mortality at-haulback) and model specificity (capacity to correctly exclude sharks not dead at-haulback).

Cross validation was carried out with a *k*-fold cross validation procedure (with *k*=10) to estimate the expected level of fit of the models to new data, and to assess eventual over-fitting problems. Because the models in this study are of the binomial type, the cross validation procedure was used to estimate the misclassification error rate, with the procedure randomly partitioning the original sample into *k*-subsamples, and

then retaining one subsample as the validation dataset and using the remaining k-1 subsamples as training datasets to build the models. The cross-validation procedure was repeated k times, with each of the k subsamples used one time as the validation dataset, and the use of k=10 was chosen as this seems to be an adequate value for models using large datasets (Fushiki, 2011). Finally, a bootstrapped cross validation procedure was also used to calculate new AUC values, that were compared to the original AUC calculated using the entire dataset.

All statistical analysis for this study was carried out with the R Project for Statistical Computing version 2.14.1 (R Development Core Team, 2012). Most functions are available in the core R Program, but some analysis required additional libraries, including library "gmodels" (Warnes, 2011a) for the contingency table analysis, library "gplots" (Warnes, 2011b) for some of the graphics produced, library "moments" (Komsta and Novomestky, 2012) for data summaries including the kurtosis and skewness coefficients, library "gam" (Hastie, 2011) for the GAM models and plots, library "mfp" (Ambler & Benner, 2010) for the multivariate fractional polynomials transformations, library "geepack" (Halekoh *et al.*, 2006) for the GEE models, library "Epi" (Carstensen *et al.*, 2011) for the ROC curve plots, and library "boot" (Canty and Ripley, 2011) for the cross validation procedure.

II.3. Results

II.3.1. Description of the catches

A total of 26,383 blue shark specimens were captured and recorded during the sampling period. Of those, complete capture information including at-haulback condition, size, sex, date and coordinates of the capture was available for 24,958 specimens (94.6% of the blue shark catch) and the analysis was therefore performed on those specimens. Of the specimens analyzed, 13,530 (54.2%) were females, while the remaining 11,428 (45.8%) were males. The females mean size in the sample was 199.5 cm FL (SD = 31.7) with the distribution ranging from 40 to 305 cm FL, while the males had a mean size of 194.5 (SD= 36.9) and the size distribution ranged from 69 to 295 cm FL (Figure II.2). The size distribution of males and females was considered significantly different, given that the null hypothesis that both sexes come from the

same continuous distribution was rejected (2-sample Kolmogorov-Smirnov test: D = 0.06, p-value < 0.001). Likewise, the ranks of the sizes of males and females was also significantly different (Mann-Whitney test: W = 7.9e+7, p-value = 0.002). The non-normality in the size data was confirmed with a Lilliefors test (D = 0.030, p-value < 0.001), with the data having a skewness coefficient of -0.41 (negatively asymmetrical) and a kurtosis coefficient of 4.99 (leptokurtic data). Note that the kurtosis coefficient used was calculated as the ratio between the 4th sample moment and the square of 2nd sample moment, and therefore the reference value for a mesokurtic sample would have been 3.
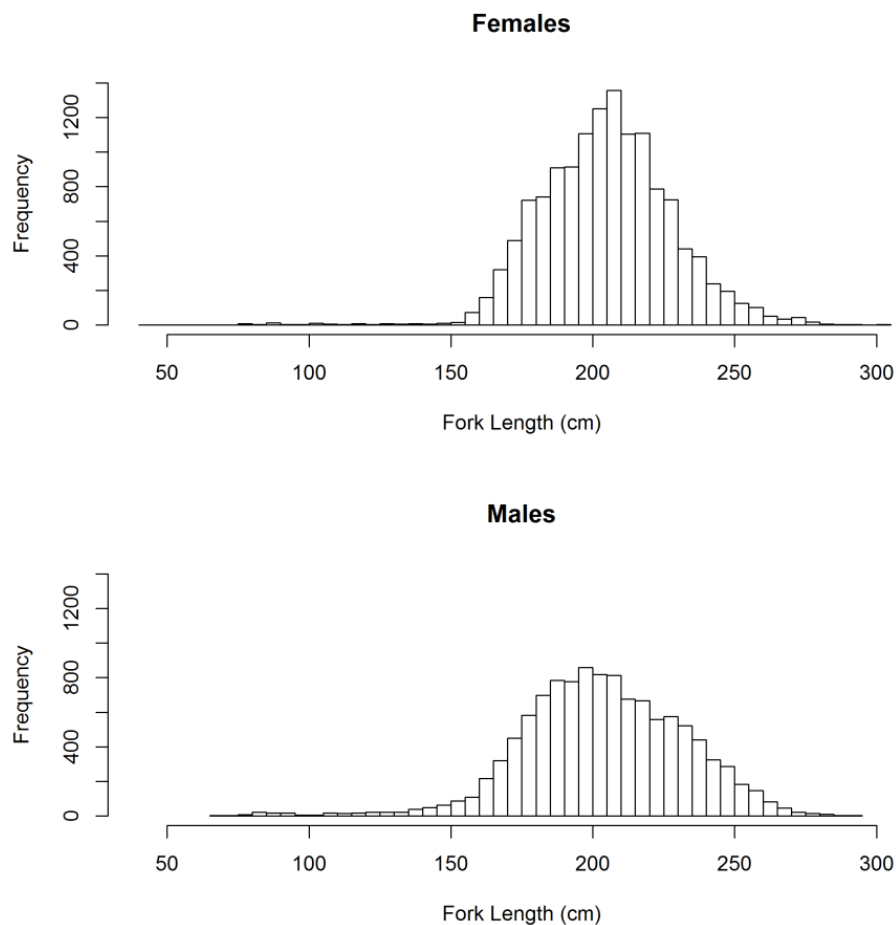
Figure II.2: Size frequency distribution of female and male blue sharks captured and analyzed during this study.

II.3.2. Proportions of hooking mortality

In general terms, 13.3% of the blue shark specimens that were captured during this study were dead at-haulback, while the remaining 86.7% were alive. In terms of the categorical variables, the proportions of alive:dead blue sharks were significantly different between all levels of the variables that were initially considered, specifically fishing trip (chi-square = 2092.5, df = 13, p-value < 0.001), sexes (chi-square = 94.4, df = 1, p-value < 0.001), year (chi-square = 1191.2, df = 3, p-value < 0.001), quarter (chi-square = 193.8, df = 3, p-value < 0.001), vessel identity (chi-square = 181.3, df = 1, p-value < 0.001) and branch line material (chi-square = 39.4, df = 1, p-value < 0.001) (Figure II.3).

Regarding the continuous variables, and considering the data grouped by the quartiles, the proportions of alive:dead sharks were different between sizes (chi-square = 833.5, df = 3, p-value < 0.001), latitude (chi-square = 643.2, df = 3, p-value < 0.001) and longitude (chi-square = 323.3, df = 3, p-value < 0.001), but not significantly different considering SST (chi-square = 2.8, df = 3, p-value = 0.419) (Figure II.3). Besides not being significant in the contingency table analysis, the SST was also found to be significantly correlated with latitude (Pearson correlation = 0.605, p-value < 0.001; Spearman correlation = 0.581, p-value < 0.001), and with longitude (Pearson correlation = -0.363, p-value = 0.001; Spearman correlation = -0.353, p-value < 0.001) which might create multicollinearity problems if both the SST and the geographical coordinates were used as explanatory variables in a multivariate model. Additionally, and because the geographical coordinates were available for all fishing sets, while SST was only available for part of the sets (specifically for 231 of the 762 sets carried out), the SST variable was discarded and not used in the final models.
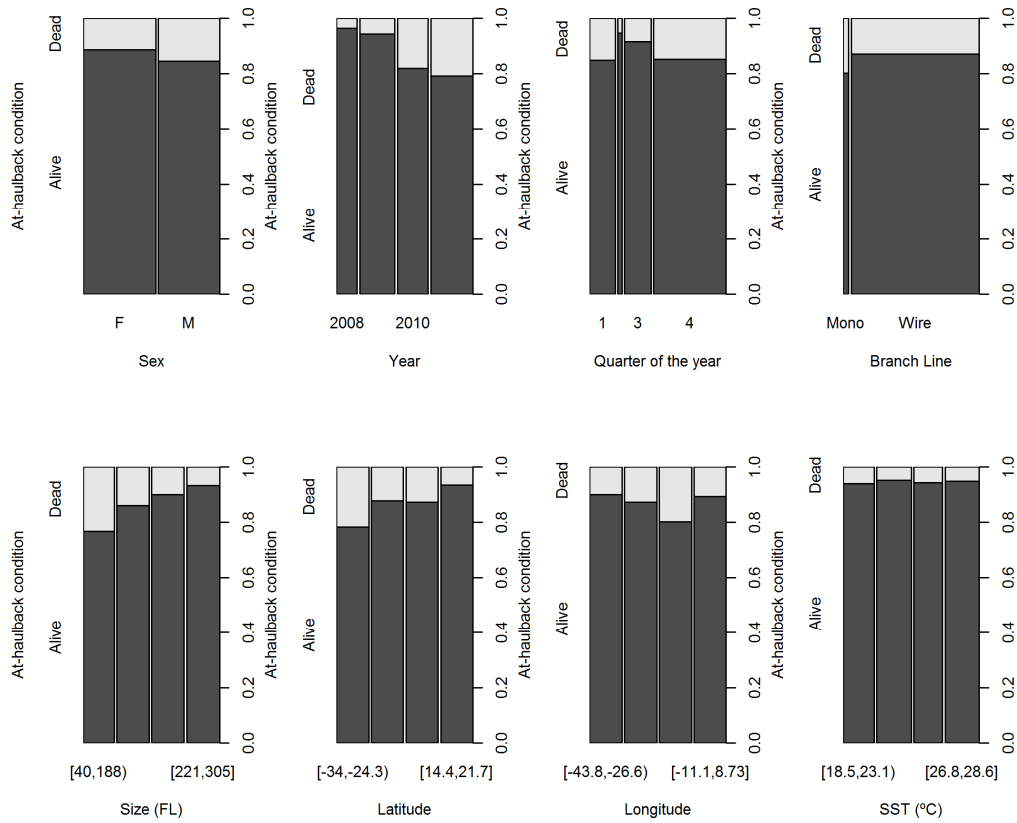
Figure II.3: Proportions of alive and dead blue sharks at-haulback with the various categorical and continuous explanatory variables considered for the analysis. The continuous variables are categorized by their quartiles.

II.3.3. Simple effects GLM and GEE models

The functional form of the continuous explanatory variables (FL, latitude and longitude) was assessed with GAM plots. The at-haulback mortality tended to decrease with increasing specimen size, towards northern latitudes and eastern longitudes (Figure II.4).
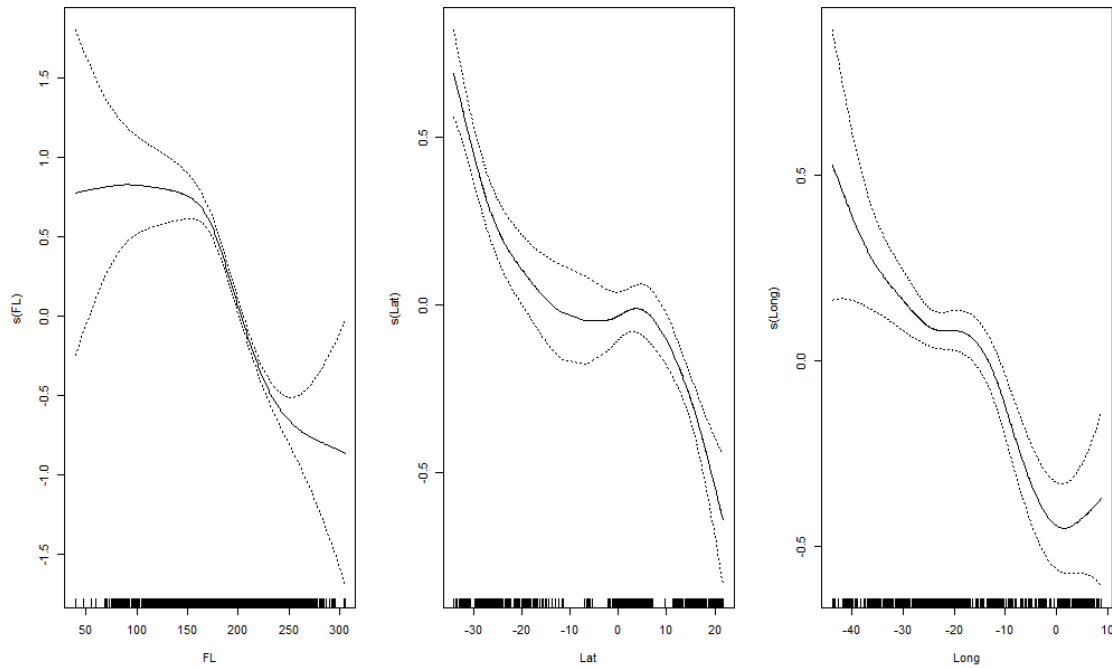
Figure II.4. Generalized Additive Model (GAM) plots with the shape of the continuous explanatory variables (FL, latitude and longitude) for modeling blue shark at-haulback mortality.

As verified with multivariate fractional polynomials models, only the longitude was significantly linear, while the specimen size and latitude were non-linear variables that needed to be transformed in order to be used within the assumptions of GLM. By applying the multivariate fractional polynomial transformations to those three continuous variables, the best candidate alternatives to the transformations of the functional form were:

Size (FL): $\left(\frac{FL}{100}\right)^{-0.5} + log\left(\frac{FL}{100}\right)$

Latitude: $log\left(\frac{Lat+34.1}{10}\right) + \left(\frac{Lat+34.1}{10}\right)^{3}$

Longitude: $\left(\frac{Long+43.8}{10}\right)$

The transformation regarding the longitude is a simple scale transformation, while the transformations for specimen size and latitude refer to transformations in the functional form. These transformed variables were used in the models instead of the original values.

24

In the simple effects multivariate model, all the variables that were initially considered were significant at the 5% level except the vessel effect. Regarding the quarter of the year the overall effect was significant, but no differences were found between quarters 1 and 2 (Wald statistic: $z = -0.323$; p-value $= 0.747$) and quarters 1 and 4 (Wald statistic: $z = 0.578$; p-value $= 0.563$). Therefore, this variable was simplified into a binomial variable (season), coded with: season 1 = quarter 3 and season 2 = quarters 1, 2 and 4.

The results of the simple effects GLM parameters in terms of significance are given in the analysis of deviance presented in Table II.1, where it is possible to see the contribution of each parameter for explaining part of the deviance observed in the blue shark at-haulback mortality. The parameters that are contributing more for the model deviance explanation are the effects of the year and specimen size, followed by the geographical location of the capture (latitude and longitude). Finally, the effects of season, branch line material and sex are contributing less for the blue shark at-haulback mortality deviance explanation, but are still significant variables in the model (Table II.1).

Table II.1. Deviance table for the simple effects GLM for the binomial response (alive or dead) status of blue sharks at-haulback. Resid.df are the residual degrees of freedom and Resid.dev is the residual deviance. Significance of the terms is given by the p-values of the chi-square test. The ".t" notations after the continuous variables (FL, Lat and Long) represent the utilization of the transformed variables in the models.

Parameter	Df	Deviance	Resid.df	Resid.dev	p-value
Null			24957	19561	
FL.t	1	645.24	24956	18915	< 0.001
Latitude.t	1	273.10	24955	18642	< 0.001
Longitude.t	1	251.79	24954	18390	< 0.001
Year	3	908.63	24951	17482	< 0.001
Season	1	11.06	24950	17471	< 0.001
Branch line	1	7.07	24949	17464	0.008
Sex	1	12.71	24948	17451	< 0.001

When applying a GEE model to those variables, and considering the fishing set as the grouping (cluster) variable, the estimated correlation value was low (alpha $= 0.058$, SE $= 0.019$), and the estimated parameters were very similar between the GLM and

GEE models, with only some minor differences (Table II.2). The overall parameter interpretation would be similar with both modeling approaches, given that the parameters were consistently positive or negative when comparing the models. The only major different in these multivariate simple effects models was that the effect of sex was significant in the GLM model but not significant (at the 5% level) within the GEE framework (Table II.2).

Table II.2. Multivariate simple effect GLM and GEE model parameters (coefficients and standard errors) for the binomial response (alive or dead) status of blue sharks at haulback. Significance of the explanatory variables is given by the Wald statistic with the respective p-values. The ".t" notations after the continuous variables (FL, Lat and Long) represent the utilization of the transformed variables in the models.

Variable	Generalized Linear Model				Generalized Estimating Eq.			
	Estimate	SE	Wald	p-value	Estimate	SE	Wald	p-value
Intercept	3.95	0.35	11.4	< 0.001	4.29	0.49	75.9	< 0.001
FL.t	-4.19	0.23	-18.5	< 0.001	-4.29	0.34	156.4	< 0.001
Lat.t	-0.01	0.00	-14.5	< 0.001	-0.01	0.01	60.0	< 0.001
Long.t	-0.25	0.02	-10.4	< 0.001	-0.21	0.05	19.5	< 0.001
Year2009	0.51	0.11	4.7	< 0.001	0.41	0.18	5.3	0.021
Year2010	1.60	0.09	16.8	< 0.001	1.34	0.18	58.6	< 0.001
Year2011	1.79	0.09	19.5	< 0.001	1.70	0.16	114.3	< 0.001
Season2	-0.19	0.07	-3.0	0.003	-0.23	0.10	5.2	0.023
BranchWire	-0.19	0.09	-2.3	0.022	-0.28	0.12	5.6	0.018
SexMale	0.15	0.04	3.6	< 0.001	0.06	0.05	1.7	0.197

II.3.4. Models with interactions

Several possible 1st degree interactions between the variables were significant at the 1% significance level and therefore a model with significant interactions was created. In this model, year and specimen size were still the most important explanatory variables, followed by the location, season, branch line material and sex (Table II.3). In terms of interactions, specimen size was significantly interacting with longitude and year; specimen sex was interacting with longitude and season; longitude was interacting with season; and branch line material was interacting with year (Table II.3). The interactions between longitude and season, and between year and branch line material seemed to be particular significant in this model, with relatively high values of deviance (Table II.3).

Table II.3. Deviance table for the GLM model with significant 1st degree interactions for the binomial response (alive or dead) status of blue sharks at-haulback. Resid.df are the residual degrees of freedom and Resid.dev is the residual deviance. Significance of the terms is given by the p-values. The ".t" notations after the continuous variables (FL, Lat and Long) represents the use of transformed variables in the models.

Parameter	Df	Deviance	Resid.df	Resid.dev	p-value
Null			24957	19561	
FL.t	1	645.24	24956	18915	< 0.001
Lat.t	1	273.1	24955	18642	< 0.001
Long.t	1	251.79	24954	18390	< 0.001
Year	3	908.63	24951	17482	< 0.001
Season	1	11.06	24950	17471	0.001
Branch line	1	7.07	24949	17464	0.008
Sex	1	12.71	24948	17451	< 0.001
FL.t:Long.t	1	13.62	24947	17437	< 0.001
FL.t:Year	3	41.96	24944	17395	< 0.001
Long.t:Season	1	71.25	24943	17324	< 0.001
Long.t:Sex	1	15.06	24942	17309	< 0.001
Year:Branchline	3	80.81	24939	17228	< 0.001
Season:Sex	1	8.71	24938	17220	0.003

Like with the simple effects model, a GEE model was also applied to this case (considering interactions), again considering the fishing set as the grouping (cluster) variable. Like in the simple effects model, the correlation within the fishing set was low (alpha = 0.051, SE = 0.022), and the parameters estimated with both the GLM and GEE models were similar, with consistently positive or negative parameters (Table II.4). In this case, the only major difference between using GLM or GEE was the loss of significance (at the 1% significance level) for the interaction between season and specimen sex (Table II.4).

Table II.4. Multivariate GLM and GEE parameters of the models with significant 1st degree interactions (coefficients and standard errors) for the binomial response (alive or dead) status of blue sharks at-haulback. Significance of the explanatory variables is given by the Wald statistic with the respective p-values. The ".t" after the continuous variables (FL, Lat and Long) represents the use of transformed variables in the models.

Variable	Generalized Linear Model				Generalized Estimating Eq.			
	Estimate	SE	Wald	p-value	Estimate	SE	Wald	p-value
Intercept	3.90	1.26	3.1	0.002	2.50	1.39	3.2	0.073
FL.t	-4.24	0.88	-4.8	< 0.001	-3.21	0.98	10.8	0.001
Lat.t	-0.01	0.00	-13.4	< 0.001	-0.01	0.00	52.8	< 0.001
Long.t	-0.96	0.29	-3.3	0.001	-0.50	0.42	1.4	0.231
Year2009	7.85	1.67	4.7	< 0.001	6.49	1.90	11.6	0.001
Year2010	2.32	1.34	1.7	0.083	2.61	1.58	2.7	0.100
Year2011	5.70	1.35	4.2	< 0.001	5.52	1.44	14.6	< 0.001
Season2	0.85	0.18	4.7	< 0.001	0.78	0.27	8.3	0.004
BranchWire	-1.26	0.21	-6.0	< 0.001	-1.25	0.27	22.3	< 0.001
SexMale	0.17	0.17	1.0	0.301	0.30	0.16	3.6	0.056
FL.t:Long.t	0.83	0.20	4.1	< 0.001	0.51	0.30	2.9	0.087
FL.t:Year2009	-5.47	1.19	-4.6	< 0.001	-4.51	1.37	10.8	0.001
FL.t:Year2010	-1.85	0.94	-2.0	0.050	-2.07	1.10	3.5	0.060
FL.t:Year2011	-3.61	0.96	-3.8	< 0.001	-3.52	1.03	11.8	0.001
Long.t:Season2	-0.49	0.05	-9.0	< 0.001	-0.44	0.09	21.9	< 0.001
Long.t:SexMale	-0.12	0.04	-3.0	0.003	-0.13	0.04	12.1	0.001
Year2009: BranchWire	0.04	0.29	0.1	0.894	0.01	0.37	0.0	0.983
Year2010: BranchWire	2.12	0.30	7.0	< 0.001	1.94	0.42	21.3	< 0.001
Year2011: BranchWire	1.42	0.24	6.0	< 0.001	1.41	0.30	22.1	< 0.001
Season2:SexMale	0.36	0.12	3.0	0.003	0.14	0.12	1.5	0.226

By using significant interactions, model interpretation gets more complex as the effects of the interacting variables need to be considered at the same time. Regarding the interaction between size and year, the at-haulback mortality for all size classes tended to increase along the years, but the relative increase was different between sizes, with the smaller specimens having a more sharp increase in mortality for the more recent years (Figure II.5). In terms of the relation between size and longitude, the at-haulback mortality remained at relatively low levels for the larger size classes throughout the entire longitude range, while a peak of at-haulback mortality was observed for the smaller size classes towards the eastern longitudes (Figure II.5).
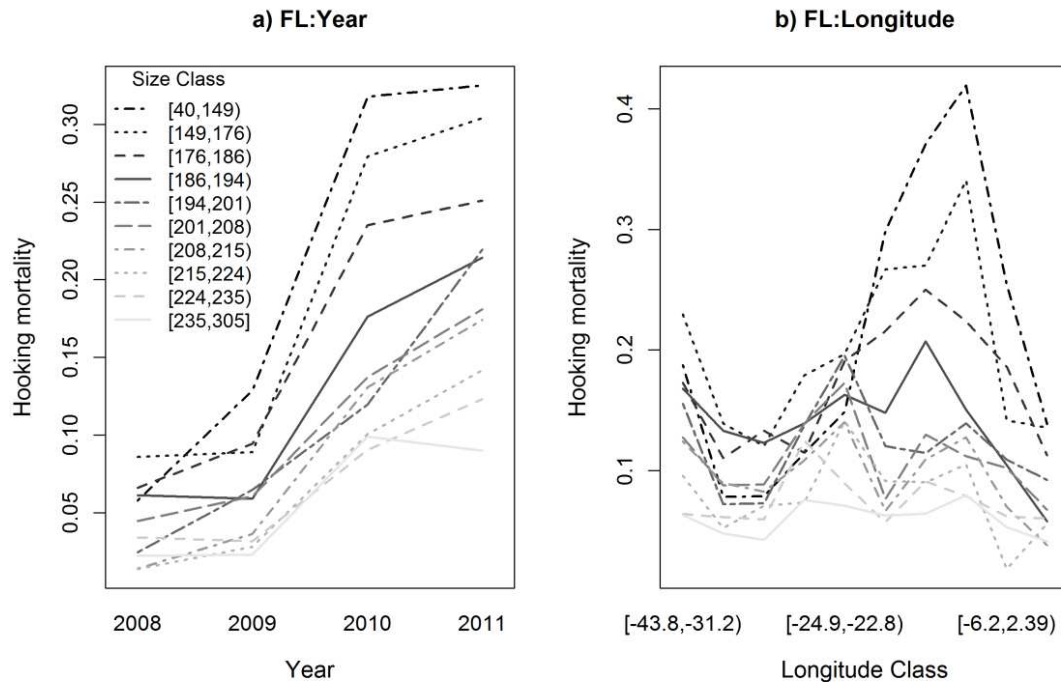
Figure II.5. Interactions between specimen size (FL) with year (a) and longitude (b). The classes of the continuous variables specimen size and longitude are categorized by the deciles.

The categorical variable sex was significantly interacting with both season and longitude (Figure II.6). On both cases the male mortality rates tended to be higher than that for females, but there were some small differences in the changing patterns. For the relation between sex and season there was an increased mortality during the combined winter (autumn to spring) season, but the increasing rate was higher for males than for females (Figure II.6).

The other significant 1st degree interaction considered in the model was between branch line material and year. In general terms, the at-haulback mortality when using monofilament branch lines remained relatively high between 2008 and 2011 (except for 2010, when a decrease was observed), while an increasing trend along the time period was observed for wire branch lines (Figure II.6).
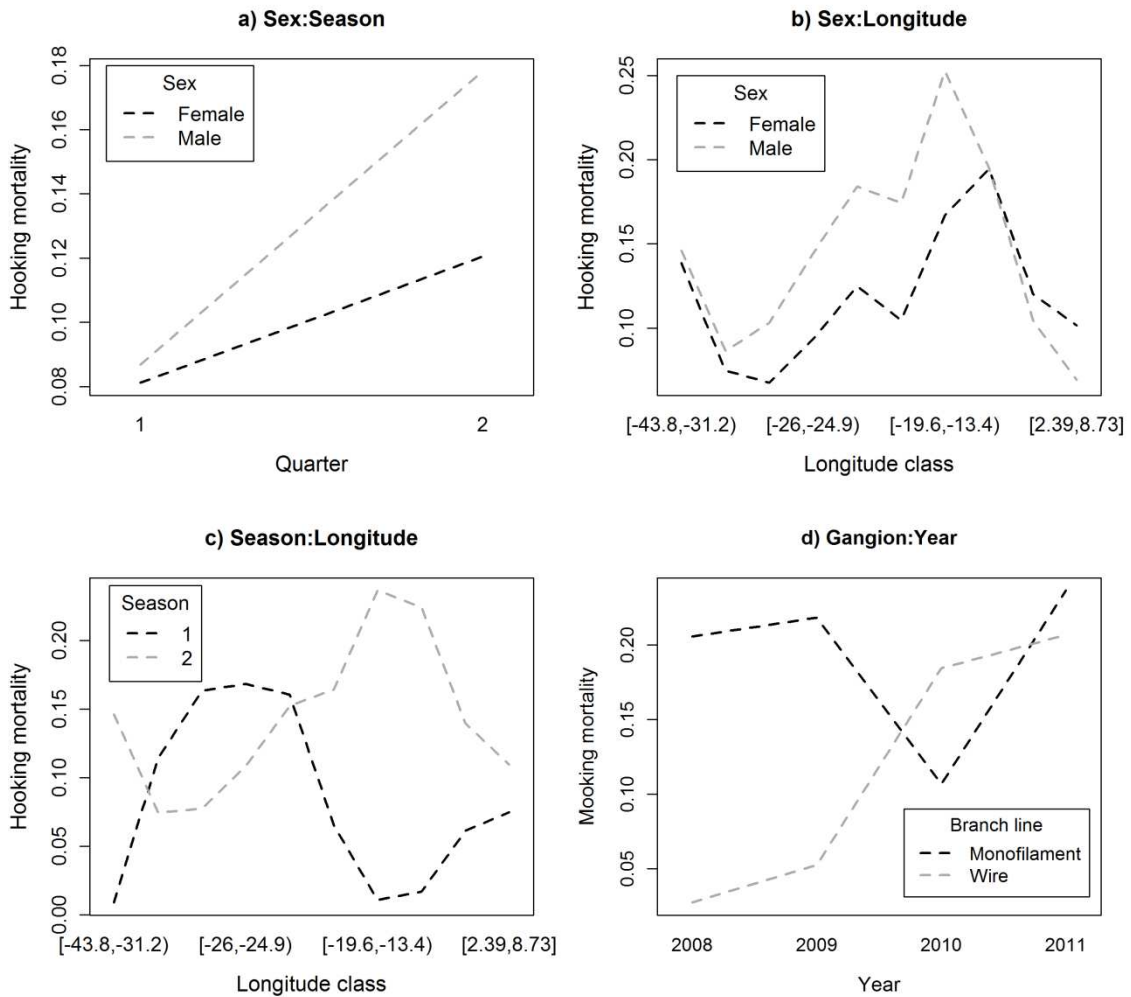
Figure II.6. Interactions plots between specimen sex with season (a), sex with longitude (b), season with longitude (c) and branch line material with year (d). The classes of the continuous variable longitude are categorized by the deciles.

II.3.5. Diagnostics and goodness-of-fit

For the final multivariate model, validation with the Pearson and Deviance residuals confirmed that there were no values that presented major and significant outliers (Figure II.7). For the Cooks distances two points presented values relatively higher than the remaining and those could possibly be values with influence in the estimated parameters (Figure II.7).
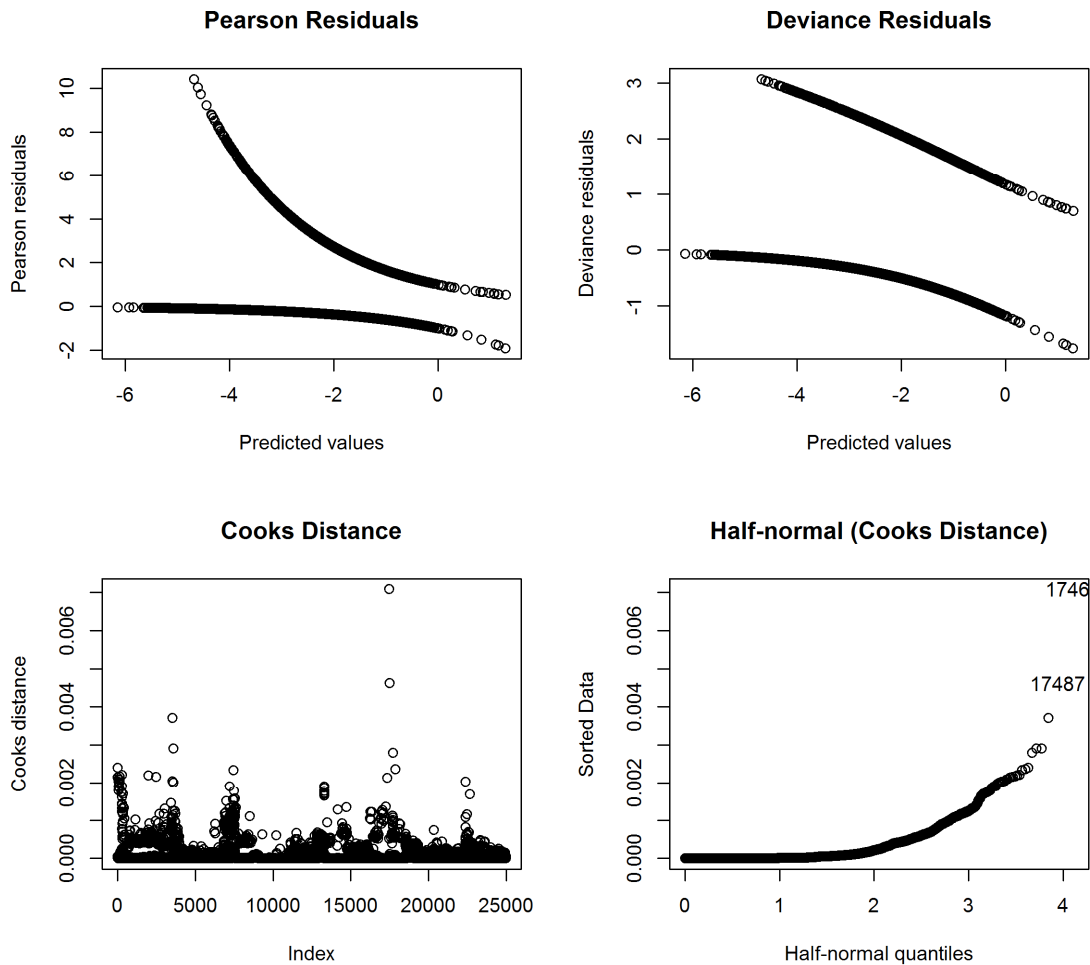
Figure II.7: Residual analysis (Pearson and Deviance residuals) and leverage values (Cooks distances) for the final GLM model including the main effects and the 1st degree interactions. The residuals are plotted in terms of the predicted values and the Cooks distances along the data index. A half-normal plot of the Cooks distances is presented to help identify the extreme values.

The DfBetas were also calculated, identifying possible observations that had more influence in the parameter estimation. Two observations seem to be possibly influential (Figure II.8), with those two observations corresponding to the values that had also been identified with the Cooks distances, specifically the data points 17,469 and 17,487 in the dataset used for the models.

Because of those two observations, two new models were created, with each new model excluding each of those data points identified. The results of the new models with the respective new estimated parameters and SE are presented in Table II.5. It is possible to see that for most of the parameters the differences in the estimations are

relatively small and lower than 20%. In terms of improvement of the explained deviance, by removing these possible influential values the differences were almost negligible, with the improvement in the R^2 of the two alternative models lower than 0.2% when compared to the original model using all data points. Because the differences in the estimated parameters were in general small, and the improvements in terms of the deviance explained are almost negligible, the remaining model diagnostics, goodness-of-fit and model discriminative capacity were tested for the original models using all data points in the dataset and without excluding any possible outliers or influential values.
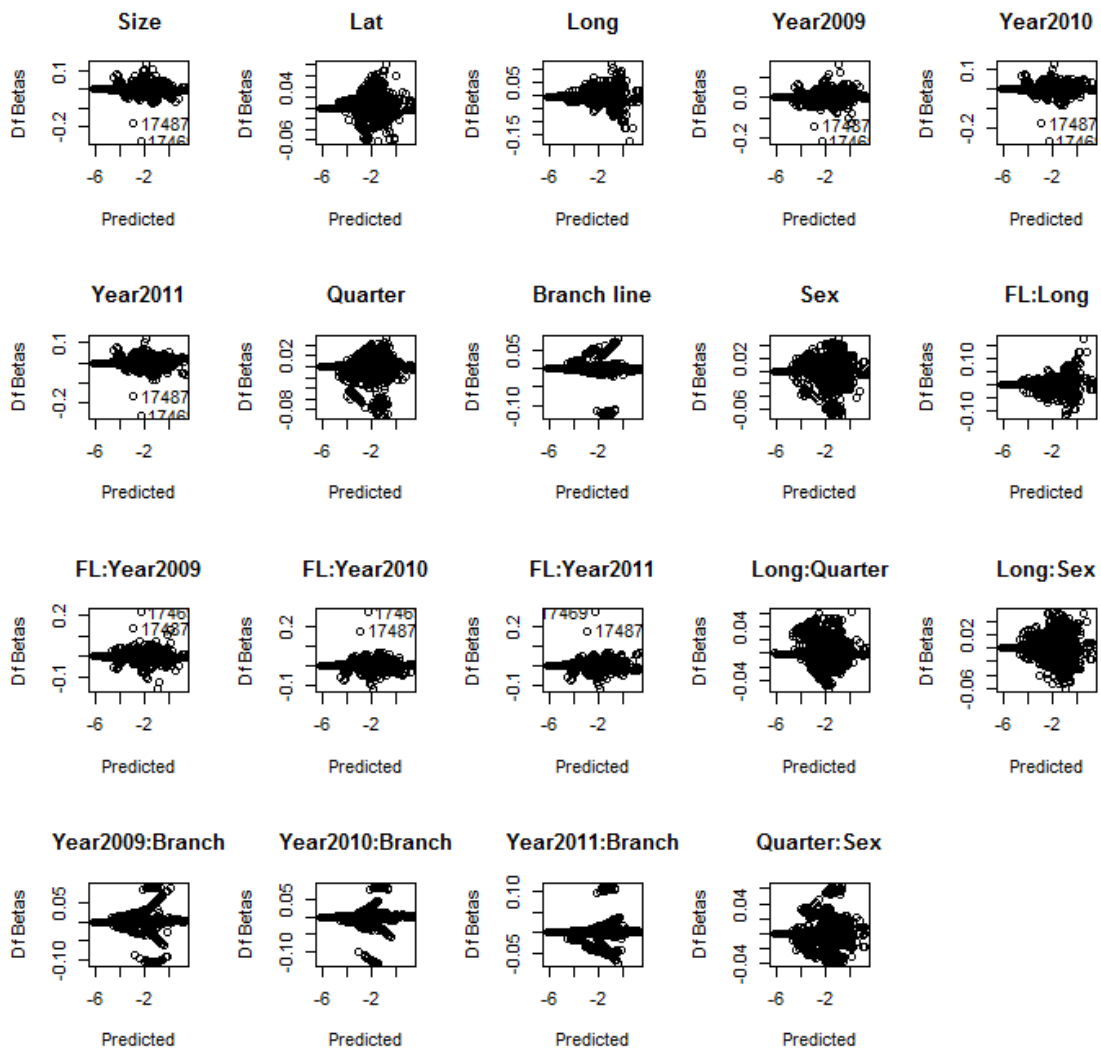
Figure II.8. Df Betas for the final GLM model including the main effects and interactions. The DfBetas are plotted along the predicted values, and the two observations that are possibly influential in some of the estimated parameters are identified.

Table II.5. Estimated parameters (with the respective SE) of the original model using all data points, and two alternative models each excluding one possible influential outlier. The differences in each of the estimated parameters (in percentage) are also included.

Parameter	All dataset		Excluding point 17,469			Excluding point 17,487		
	Estimate	**SE**	**Estimate**	**SE**	**Dif(%)**	**Estimate**	**SE**	**Dif(%)**
(Intercept)	3.90	1.26	3.48	1.30	10.70	3.57	1.29	8.51
FL.t	-4.24	0.88	-3.94	0.91	7.10	-4.00	0.90	5.62
Lat.t	-0.01	0.00	-0.01	0.00	-0.09	-0.01	0.00	-0.09
Long.t	-0.96	0.29	-0.96	0.29	-0.60	-0.96	0.29	-0.32
Year2009	7.85	1.67	8.28	1.70	-5.59	8.19	1.69	-4.38
Year2010	2.32	1.34	2.76	1.38	-18.66	2.66	1.37	-14.63
Year2011	5.70	1.35	6.14	1.39	-7.69	6.04	1.38	-6.03
Season2	0.85	0.18	0.85	0.18	0.26	0.85	0.18	0.15
GangionWire	-1.26	0.21	-1.27	0.21	-0.71	-1.27	0.21	-0.71
SexM	0.17	0.17	0.17	0.17	2.15	0.17	0.17	0.30
FL.t:Long.t	0.83	0.20	0.83	0.20	-0.42	0.83	0.20	-0.23
FL.t:Year2009	-5.47	1.19	-5.79	1.22	-5.73	-5.72	1.21	-4.49
FL.t:Year2010	-1.85	0.94	-2.16	0.97	-16.74	-2.09	0.96	-13.14
FL.t:Year2011	-3.61	0.96	-3.92	0.99	-8.65	-3.86	0.98	-6.80
Long.t: Season 2	-0.49	0.05	-0.49	0.05	0.10	-0.49	0.05	0.01
Long.t:SexM	-0.12	0.04	-0.12	0.04	0.90	-0.12	0.04	0.08
Year2009:GangionWire	0.04	0.29	0.05	0.29	-18.60	0.05	0.29	-18.96
Year2010:GangionWire	2.12	0.30	2.13	0.30	-0.42	2.13	0.30	-0.43
Year2011:GangionWire	1.42	0.24	1.43	0.24	-0.55	1.43	0.24	-0.64
Season 2:SexM	0.36	0.12	0.36	0.12	0.08	0.36	0.12	-0.22

In terms of model goodness-of-fit, both the simple effects and the model with interactions passed the Hosmer and Lemeshow test, with the simple effects model having a chi-square = 11.8 (p-value = 0.162) and the model with interactions having a slightly better fit (chi-square = 9.6, p-value = 0.295). The same type of improvement was observed for the Nagelkerke R^2 values, with the simple effects GLM having an R^2 of 0.149 and the model with interactions a higher R^2 of 0.165. Finally the discriminative capacity of the models also improved by adding the interactions, with the simple effects model having an AUC (estimated from the ROC curve) of 0.741, and the model with interactions a higher AUC value of 0.750, with a sensitivity of 74% and a specificity of 65% for a cut point of 0.144 (Figure II.9). Those AUC discriminative values are, according to Hosmer and Lemeshow (2000), considered acceptable.
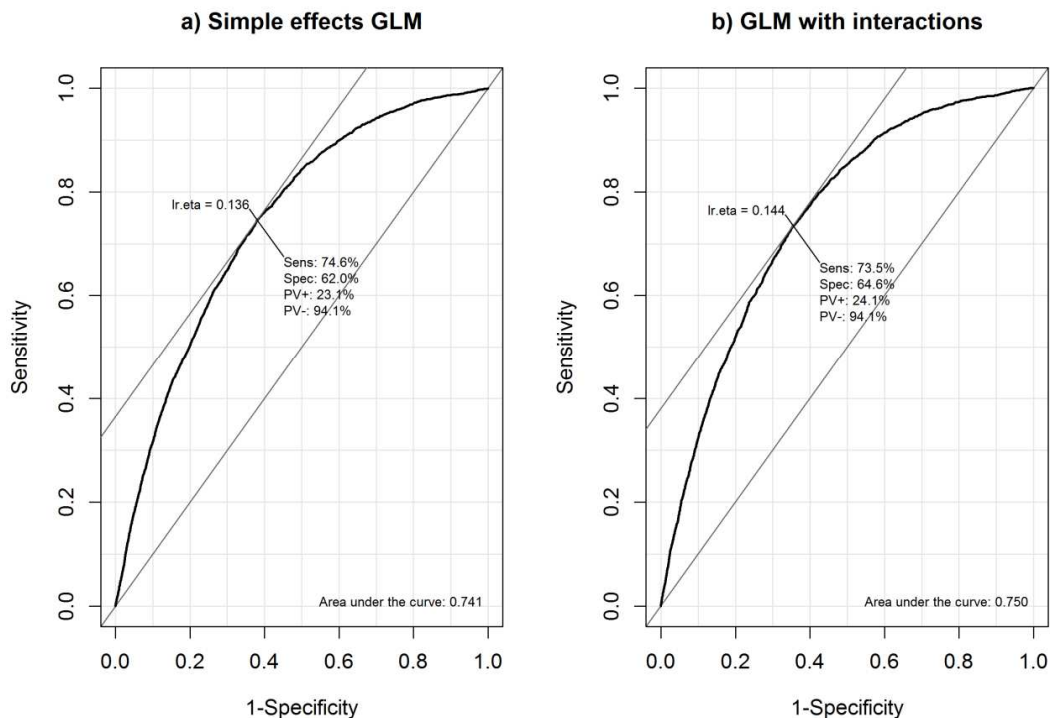
Figure II.9. Receiver Operating Characteristic (ROC) curves for the multivariate GLM using simple effects (a) and considering interactions (b), for the binomial response (alive or dead) status of blue sharks at-haulback. The Area Under the Curve (AUC) values are given, as well as the sensitivity (Sens), specificity (Spec) and predictive values (PV) at the optimal response cut-points (lr.eta).

The 10-fold cross validation procedure resulted in an estimated prediction error of 13.4% for the multivariate simple effects model, and a similar prediction error of 13.3% for the model with interactions. The bootstrapped cross-validation procedure resulted in an AUC = 0.748, which is very similar to the original AUC using the entire dataset (0.750) and also validates the models. Additionally, all bootstrapped models also passed the Hosmer and Lemeshow test (p-value > 0.05 on all cases) for model goodness-of-fit.

II.3.6. *Examples of model interpretation*

One parameter that in biological terms is particularly important to interpret and analyze is the influence of the specimen sizes on the probabilities of dying at-haulback, as well as the influence in the odds-ratios, and therefore the final model estimated was used for prediction and interpretation of the effects of changing specimen sizes on the mortality rates and odds-ratios. It was possible to see that the probabilities of a specimen dying at-haulback decreases with increasing specimen size, but those decreasing probabilities are steeper for the smaller specimens and tend to stabilize for the larger specimens (Figure II.10).

By interpreting the odds-ratios (in this case calculated for an increase of 10cm FL in specimen size), it is possible to see that as a shark increases in size the odds of dying decrease, but these odds are non-linear and vary with the size. For example, for a blue shark close to the size of birth (e.g. 50 cm FL) an increase of 10 cm FL in size will result in the odds of dying decreasing by 22%, with 95% CI varying between 14% and 30% (Figure II.10). On the other hand, for a larger adult blue shark with 250 cm FL, an increase of 10 cm FL in size with result in the odds of dying decreasing by only 11%, with 95% CI varying between 7% and 15% (Figure II.10).
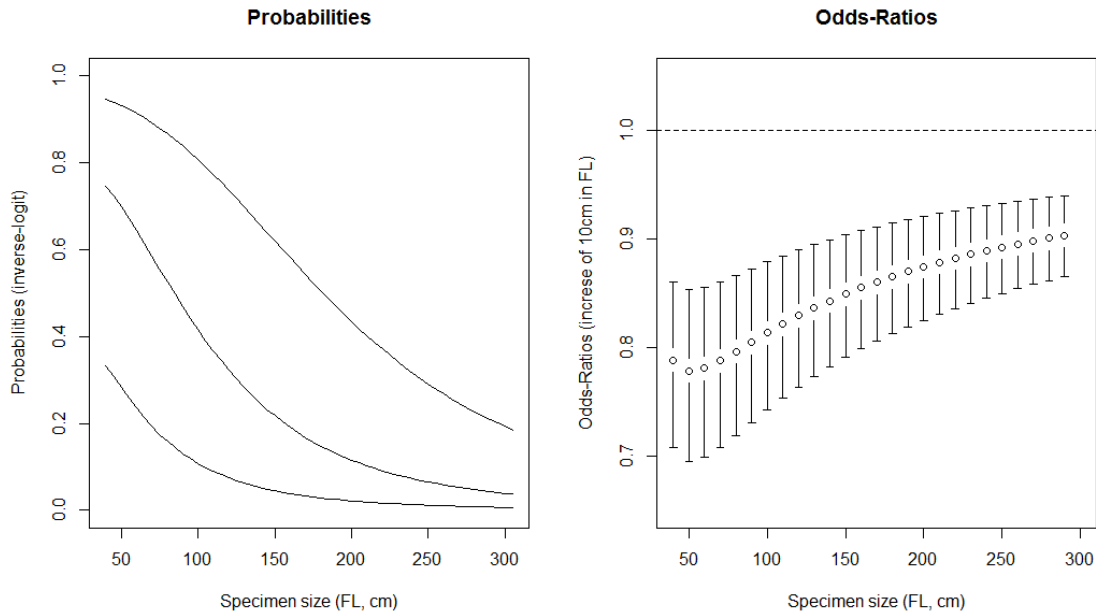
Figure II.10. Probabilities of a blue shark dying at haulback with varying specimen size (left), and the odds-ratios of a blue shark dying at-haulback (for an increased 10 cm FL of specimen size) along the size ranges of the captured specimens (right).

II.4. Discussion

This study focused on the parameters affecting blue shark at-haulback mortality in a large scale swordfish pelagic longline fishery in the Atlantic Ocean. In general, 13.3% of the blue shark capture was dead at-haulback, but it was possible to determine that several variables had significant effects on this mortality rates and a statistical model was produced.

Several studies have previously addressed blue shark at-haulback mortality in pelagic longline fisheries, including the works of Diaz and Serafy (2005) and Campana *et al.* (2009) in the Atlantic, and Moyes *et al.* (2006) in the Pacific. For the Canadian fleet in the northwest Atlantic, Campana *et al.* (2009) estimated the blue shark at-haulback mortality in the 12-13% range as measured by fishery observers, which is relatively similar to the 13.3% estimated in our study. However, using telemetry technology to account for the post-release mortality, Campana *et al.* (2009) also reported that the actual mortality values could be closer to 20% due to the added post-release mortality. In the Pacific Ocean, Moyes *et al.* (2006) also addressed post-release mortality using satellite telemetry, and in the case of blue shark noted that the

survivorship of sharks landed in an apparently healthy condition was likely to be high. This means that our estimates of 13.3% mortality probably represent accurately the at-haulback mortalities of blue shark in the Portuguese pelagic longline fishery, but at this stage the total mortalities (that also need to account for post release mortality) still remain unknown.

The most significant factors affecting mortality in our study were the year effect, followed by specimen size. The yearly variations may be related with inter-annual variability inherent to the species or the fishery spatial/seasonal patterns, or eventual changes in the fishery that may be contributing to changes in these rates. It should be mentioned, however, that the data analyzed in this study was collected by the fishery observer program that tries to cover the geographical/seasonal variability of the fleet in terms of catch rates, but it is a fishery-dependent source of data that cannot cover those geographical/seasonal patterns in a truly balanced design.

With regards to the specimen size, the probabilities and odds-ratios show that the larger specimens have lower probabilities of being dead at-haulback than the smaller specimens. However, these effects are non-linear, with the odds-ratios of surviving higher for the smaller specimens (as they grow in size) and then tending to stabilize as the sharks reach larger sizes. Some previous studies had already looked into effects of specimens sizes in the mortality rates (e.g. Diaz and Serafy, 2005; Campana *et al.*, 2009), and similar results were reached, with decreasing probabilities of at-haulback mortality as the specimens increase in size. These results have a direct effect on eventual management and conservation initiatives such as the establishment of minimum and/or maximum landing sizes, as the efficiency of such measures will have specific effects depending on the shark sizes. For example, the establishment of a minimum landing size would have a more limited conservation effect, as the smaller specimens are the ones that have higher probabilities of dying due to the fishing process, and would therefore tend to be discarded already dead.

Even though the models created and presented seem to be valid and perform well for predicting blue shark at-haulback mortality rates (as verified by the residual analysis, goodness-of-fit, and cross-validation procedures), some limitations need to be addressed and considered. One characteristic of our study was that the hook style effect was not considered, mainly because the Portuguese longline fleet uses exclusively J-

style hooks. Therefore, the values reported in our study refer specifically to fisheries using this type of hooks, while other pelagic longline fleets may use different hooks such as circle and/or tuna hooks. Some previous studies have reported that blue shark mortality rates were higher with J-style hooks when compared to circle hooks (Carruthers *et al.*, 2009), while on the other hand Coelho *et al.* (2012b) reported that for the elasmobranch species more commonly discarded (e.g. bigeye thresher and crocodile shark) the hook style (J-style *vs.* circle hooks) seemed unrelated to at-haulback mortality. Likewise, Kerstetter and Graves (2006) also showed that even though several target and bycatch species seemed to have higher rates of survival at-haulback with circle hooks, the effects were not statistically significant for most species. On the contrary, Afonso *et al.* (2011) compared J-style with circle hooks in the south-western Atlantic Ocean and concluded that circle hooks were efficient in reducing the mortality rates of most species caught, both in pelagic and coastal longline fisheries, observing at the same time that the catch rates of some species, including the blue shark, were higher with circle hooks. In the North Pacific Ocean, however, Yokota *et al.* (2006) showed that the hooks (circle *vs.* tuna hooks) had little effect on the catch rates and mortalities of blue shark. This variability in results seems to support the fact that specific studies and assessments should be carried out specifically for each fishery and fleet in question.

One possible shortcoming in our study was the fact that the fishing gear soaking time was not considered, with several previous studies (e.g. Campana *et al.*, 2009; Diaz and Serafy, 2005; Morgan and Burgess, 2007) having demonstrated that the soaking time was also a significant variable for predicting at-haulback mortality on elasmobranchs. Besides the fishing gear soaking time, Morgan and Carlson (2010) also demonstrated that the capture time (measured with hook timers) was also influential in the mortality rates of some demersal shark species captured in bottom longline fisheries. Finally, and even though in our study the gangion material had a relatively small effect on the mortality rates, other authors have shown that some components of longline gear may interact to influence catch rates and relative mortality estimates (e.g. Afonso *et al.*, 2012, Ward *et al.*, 2008). As suggested by these authors, it could be hypothesized that nylon leaders could catch relatively more dead blue sharks than wire leaders because healthy and robust specimens, which would be more likely to be alive at gear retrieval, may have more chances of biting through the nylon and escaping.

The logistic models used in our study seem adequate to evaluate the contribution of potential explanatory variables to blue shark at-haulback mortality, as the response variable is binomial (dead *vs.* alive sharks at fishing gear haulback). The models created used both biological factors such as specimen size and sex, as well as fishery operational factors such as geographical location and branch line material. In our study the vessel effect was tested but not considered significant, while a previous study by Campana *et al.* (2009) had verified that the vessel effect was significant. One important difference between the two studies is in the number of vessels monitored that was much larger in the Campana *et al.* (2009) study. Eventual differences between different vessels can hypothetically be due to: 1) vessels (in different trips and sets) targeting different species, and using therefore different gear specifications, such as monofilament *vs.* steel branch line materials; 2) vessels with different crews that may handle the sharks in different ways; 3) vessels using different fishing *metiers* that can result in different soaking times of the fishing gear, which will be influent in the mortality rates. Such possibilities are hypothesis that cannot be easily verified at this stage, but it is feasible to consider that a correlation in the mortality data within vessels, fishing trips or fishing sets may exist in those fishery-dependents datasets.

For addressing such eventual lack of independence in the sample, the ideal scenario would be to collect fishery-independent data, but for the large pelagic species such data would be extremely costly, and therefore fisheries-dependent data (either logbooks or fishery observer datasets) is usually the only available data for such analysis. However, models such as GLM or GAM assume that the data is independent, and therefore making inference from such data with such models may result is biased estimates. For such cases, the use of GEE models might be a valid alternative approach, as this modeling technique calculates a working correlation matrix that approximates the true correlation on the observations (Wang and Carey, 2003). Therefore, in our study we opted for a methodology of comparing GLM with GEE models, using the fishing sets as the grouping variable in the GEE models, and assuming therefore a possible lack of independence of data within each fishing set. With the GEE models a working correlation matrix is estimated, that is then used to correct the model parameters. However, the estimated correlation parameter was low, meaning that this lack of data independence between fishing sets does not seem to be significantly

affecting the GLM estimates, which could thus be considered also valid for predicting blue shark mortality rates.

This paper presents new and important information on the impacts of this pelagic longline fishery on blue shark populations in a wide Atlantic area. The results can now be used to predict the effects of the fishery on blue shark mortality, and specifically on how several factors are contributing to this mortality rates. One immediate application is, for example, to determine the efficiency of eventual future management and conservation initiatives such as the establishment of minimum and/or maximum landing sizes. The results can also be incorporated into future stock assessment models, including ecological risk assessment analysis carried out regularly by tRFMOs for bycatch species.

CHAPTER III.   **MODELING BLUE SHARK CATCH RATES IN A PELAGIC LONGLINE FISHERY IN THE SOUTHERN ATLANTIC OCEAN.[2]**

III.1. Introduction

The blue shark (*Prionace glauca*) is one of the most wide ranging of all pelagic sharks, found throughout tropical and temperate seas (Last and Stevens, 2009). It is capable of large scale migrations (Queiroz *et al.*, 2005; Silva *et al.*, 2010; Campana *et al.*, 2011), and has complex movement patterns and spatial structure probably related to the reproduction cycles and prey distribution (Pratt, 1979; Stevens, 1990). The blue shark is possibly the most abundant of all pelagic sharks, with most captures taking place as bycatch in pelagic longlines targeting tunas and swordfish (Aires-da-Silva *et al.*, 2008).

Modeling and understanding the catch rates dynamics of any species is an extremely important aspect for fisheries management and conservation, as it allows for a better understanding in terms of the species distribution and impacts by the fisheries. Some of the previous blue shark studies available in the literature are mainly descriptive in nature (e.g. Montealegre-Quijano and Vooren, 2010), while others have used advanced modeling techniques to analyze this type of data. Previous examples of such modeling approaches include the studies of Bigelow *et al.* (1999) and Walsh and Kleiber (2001) that used Generalized Additive Models (GAM) to model blue shark catch rates in the Hawaiian based USA pelagic longline fishery in the North Pacific; Megalofonou *et al.* (2009b) that used Generalized Linear Models (GLM) to determine the operational, spatial and seasonal effects affecting blue shark catch rates in the Mediterranean Sea by the Italian and Greek pelagic longline fleets; Carvalho *et al.* (2011) that created a GAM model for predicting blue shark catch rates in the southwest Atlantic in areas of operation of the Brazilian fleet; and Vögler *et al.* (2012) that also used GAM models for modeling the blue shark catch rates in the eastern tropical Pacific by the Mexican fleet. All of those examples have focused relatively small areas (when the oceanic nature of the species is considered), i.e., Hawaiian region in the Bigelow *et*

---

[2] ***Based on a manuscript in Preparation***: Coelho, R., Infante, P. & Santos, M.N. Modeling blue shark (*Prionace glauca*) catch rates in a pelagic longline fishery in the southern Atlantic Ocean using fixed and mixed effects generalized linear models.

al. (1999) and Walsh and Kleiber (2001) studies; eastern Mediterranean in the Megalofonou *et al.* (2009b) study; southwestern Atlantic off Brazil in the Carvalho *et al.* (2011) study; and eastern tropical Pacific off Mexico in the Vögler *et al.* (2012) study. At this stage, the authors are unaware of any previous study where such modeling analysis was performed for this species at a wide scale oceanic level.

Most of the previous studies used statistical models with the main objective of standardizing the catch rates of blue shark for creating annual indexes of abundance. This process is, in theory, used to remove factors other than the annual changes in abundance of the population that can have an impact on the catch rates over time. The primary objective of such studies is not necessarily to create models for understanding the factors affecting the catch rates, but to obtain indexes of relative abundance that reflect the actual changes of the species' abundances (Maunder and Punt, 2004), and where other factors that can affect the catches (e.g. seasonality, fishery-specific operations, regional effects) have been removed. Examples of such approaches are the works by Carvalho *et al.* (2010) for the Brazilian and Tavares *et al.* (2012) for the Venezuelan pelagic longline fisheries, in both cases with the authors using GLM approaches to standardize the annual catch rates. Likewise, most of the technical works carried out within the scope of the tuna Regional Management Fisheries Organizations (tRFMO) use such approaches for creating annual indexes of abundance for utilization within stock assessments (e.g. Coelho *et al.*, 2011; Hiraoka and Yokawa, 2012).

Given that many of the previous works focusing blue shark catch rates in pelagic fisheries have been mainly descriptive in nature, have been somewhat limited in spatial coverage and in comparisons of modeling techniques, and have been mainly created for standardizing annual catch rates, the authors considered that there was the need for a new study, covering a wider oceanic spatial scale and using/comparing several modeling techniques. Therefore, the present study was elaborated with the main objective of modeling blue shark catch rates in the Portuguese pelagic longline fishery targeting swordfish over a wide geographical area of the southern Atlantic Ocean. For achieving this main goal, and because of the characteristics of this fishery-dependant dataset (data collected by the commercial fisheries), a secondary objective was to explore and compare different modeling approaches, using different possible model types (GLM - Generalized Linear Models and GLMM - Generalized Linear Mixed Models) and distributions (Gamma, Poisson, Negative Binomial and Tweedie).

III.2. Material and Methods

III.2.1. Data collection

Data for this study was collected by fishery observers from the *Portuguese Sea and Atmospheric Research Institute* (IPMA, I.P.) placed onboard Portuguese longliners targeting swordfish in the Atlantic Ocean. Data was collected between October 2008 and December 2011. During that period, information from a total of 533 longline sets corresponding to 728,254 deployed hooks was collected, with this study focusing on the southern Atlantic region (Figure III.1). For the purposes of this study the southern Atlantic region was defined by the latitudes southern of 5ºN, following the stock delimitations of major shark species used by ICCAT (ICCAT, 2006-2009). For each fishing set that was carried out, information was recorded for the date, fishing set location (latitude and longitude), number of hooks used in the set, gangion material (monofilament or wire), and the species-specific catches in number (n) of specimens.

Additional variables relative to the fishing sets that were calculated *a posteriori* (using the fishing set location and date) included the lunar phase (category) and illumination (scaled luminosity from 0 to 1), sea surface temperature (SST, ºC) using data from the NOAA National Climatic Data Center (Reynolds *et al.*, 2007; NOAA, 2012), Chlorophyll-a (mg/m^2) using data from the NASA Ocean Color Group (NASA, 2012), salinity (mg/L), mixed layer depth (MLD, depth in meters at which the temperature drops 0.2ºC) and sea surface height (SSH) using HYCOM models (Bleck *et al.*, 2002, HYCOM, 2012), current velocity (m/s) and kinetic energy (m^2/s^2) from the NOAA Ocean Surface Current Analysis (OSCAR, 2012). All the data from those variables were interpolated from the different sources using the Marine Geospatial Ecology Tools (MGET) developed by Roberts *et al.* (2010).
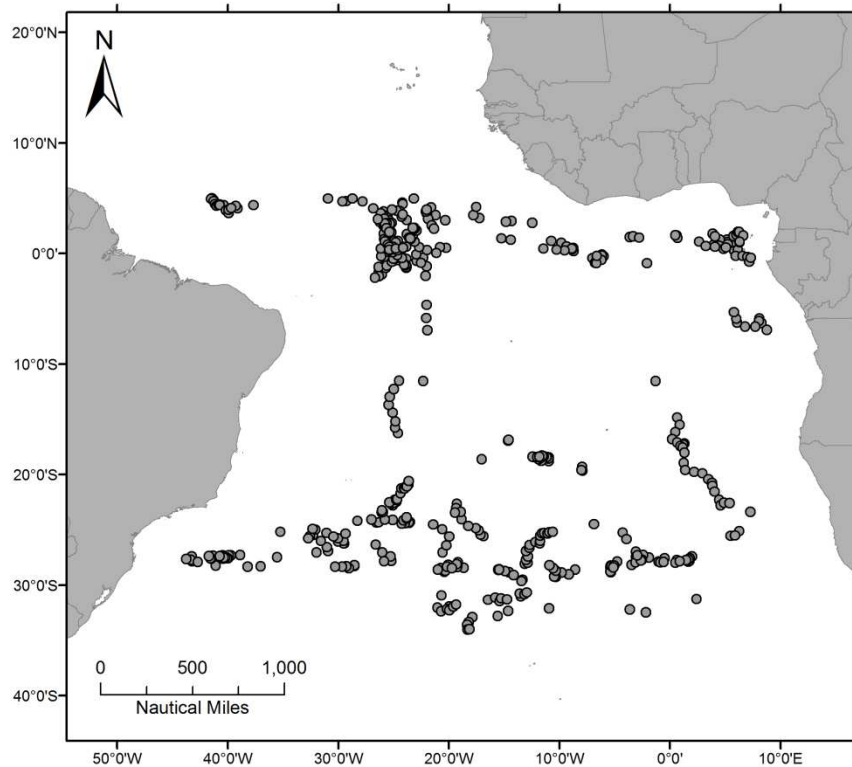
Figure III.1. Location of the 533 fishing sets carried out by the Portuguese longline fleet in the southern Atlantic Ocean (south of latitude 5ºN) that were analyzed for this study

III.2.2. Preliminary data analysis

The catch per unit of effort (CPUE - Y_i the response variable in this study) was calculated for each fishing set as the number of blue sharks captured (n) per 1000 deployed hooks (n/1000 hooks). This variable was analyzed in terms of shape with a histogram and a QQ plot, and tested for normality with a Kolmogorov–Smirnov test with Lilliefors correction.

The candidate continuous explanatory variables were analyzed with correlation matrices plots and by calculating non-parametric Spearman correlation coefficients. Those plots and correlation tests were mainly used for a preliminary analysis between the response variable and the candidate explanatory variables, as well as for eventual correlations between the explanatory variables. For the relationships between the response variable and the candidate categorical explanatory variables, boxplots and non-parametric tests were used to assess if differences occurred in the blue shark CPUE of the various categories of these variables. The non-parametric tests were used because

44

the assumptions of normality (tested with a Lilliefors test) and homogeneity of variances (tested with Levene tests) were not verified, even after transforming the response variable.

III.2.3. Statistical modeling

III.2.3.1. Modeling approaches

The first modeling approach carried out in this study to explain the blue shark catch rates in the southern Atlantic Ocean was performed with GLM models (McCullagh and Nelder, 1989; Agresti, 2002) that can be noted as:

$$\eta(Y_i) = \beta_0 + \beta_1 x_{1,i} + \beta_2 x_{2,i} + \cdots + \beta_k x_{k,i} + \varepsilon_i$$

Where η represents the link function, x_i the model variables, β the model coefficients that were estimated by maximum likelihood, and ε represents the errors.

The variable selection criteria followed the stepwise approach recommended by Hosmer and Lemeshow (2000). The univariate significance of each explanatory variable was determined by the Wald statistic and by the likelihood ratio tests, comparing each univariate model with the null model. The significant variables were then used to construct a simple effect multivariate GLM, with the non-significant variables (at the 5% level) eliminated consecutively from the model. At this stage, the variables that had been eliminated in the first step were further tested, in order to determine an eventual significance within the framework of a multivariate model. Once a final multivariate simple effects model was obtained, each pair of possible 1st degree interactions was tested, and those were considered for inclusion in the final model if significant at the 1% level, using Wald statistics and likelihood ratio tests.

In terms of the GLM assumptions regarding the explanatory variables, the assumption of linearity (in the continuous variables) with the linear predictor was assessed by creating and analyzing GAM plots. If evidences of non-linearity were present, then multivariate fractional polynomial transformations were carried out, and the transformed explanatory variables were used in the final models (Royston and Altman, 1994).

Another assumption in the GLM models that was considered was the fact that the samples should be independent (randomly collected), but in this case the data comes from a fishery-dependent source, i.e., from the commercial fishery. This means that some variables, for example the skill of the skipper and crew for handling the fishing gear that may have implications on the catch rates, were not controlled. To handle this lack of independence, an alternatively modeling approach with Generalized Linear Mixed Models (GLMM) was carried out (Jiang, 2007; Fitzmaurice *et al.*, 2009). GLMM are extensions of GLM and combine the properties of two statistical frameworks widely needed in biological studies, namely linear mixed models, which incorporate random effects, and generalized linear models, which handle non-normal data (Bolker *et at.*, 2008). Within GLMM, two types of variables can be considered, the fixed effects and the random effects. Random effects typically include blocks in experiments or observational studies that are replicated across sites or times, but can also encompass variation among individuals, species, regions or time periods (Bolker *et al.*, 2008). The choice of what should be a fixed and random effect can sometimes be a conceptual choice, and in our study we chose to use as random variables the vessel effect, given that conceptually the variability in the blue shark catch rates between vessels may depend on the intrinsic characteristics of each fishing vessel, skipper and crew, while the other possible explanatory variables (i.e. season, year, branch line material, SST, Chlorophyll, MLD, SSH, current velocity and lunar luminosity) were considered as fixed effects. Because the GLMM in this study were used mainly as a comparative technique with the more commonly used GLM approaches, the same variables that were selected for the final GLM models were used, adding only the vessel effect as a random variable. However, and for comparison and validation purposes, a GLMM with the variables selected using a stepwise top-down strategy, as recommended by Zuur et *al.* (2009) for these types of models, was also created.

The GLMM used in this study can be defined as

$$\eta(Y_{ij}) = \beta_0 + \beta_1 x_{1,ij} + \beta_2 x_{2,ij} + \cdots + \beta_k x_{k,ij} + a_j + \varepsilon_{ij}$$

Where η represents the link function, x_i the model fixed effects variables, β the model coefficients that were originally estimated with penalized quasi-likelihood (PQL) (Venables and Ripley, 2002) and then with Laplace approximations in the final models (Bolker *et al.*, 2008), *a* represents the random variable with a distribution defined by

$a \sim N(0, \sigma^2)$, and ε represents the errors. In this specific case the index j of the equation represents the different vessels, given that only the vessel effect was used as a random variable, and the index i represents the samples.

III.2.3.2. *Dealing with zeros in the response variable*

Both the blue shark catches in numbers per set (n) and the CPUE (n/1000 hooks) are types of response variables characterized for having some zero values, specifically in the fishing sets with zero catches of blue sharks. Because of these characteristics, several alternative approaches in terms of error distributions were used and tested. Most of the distributions used in this work can be defined in several different ways, and we used the notations presented by Zuur *et al.* (2009).

The first approach to deal with the zeros in the response variable was to add a small constant (δ) to all observed CPUE values, so that the response variable was transformed into CPUE + δ, and become a continuous positive variable that no longer contained zeros. The choice of the δ value to be added can be somewhat subjective, and after testing some possible alternatives (1 and 10% of the mean), we added the value of 1, which seems to be a common approach in many fisheries biology studies (e.g. Ortiz and Arocha, 2004; Punt *et al.*, 2000). With the response variable transformed in this way (Y_i now CPUE+1), a Gamma distribution defined by $Y_i \sim Gamma(\mu, \nu)$ was used to model the data, with the expected values and variance defined by:

$$E(Y) = \mu$$

$$Var(Y) = \frac{\mu^2}{\nu}$$

Where μ is the mean and ν^{-1} defines the dispersion, with small values of ν relative to μ^2 implying that the spread of the data is large.

The second alternative approach was to use distributions for categorical (count) data, such as the Poisson and the Negative Binomial (NB), and in those cases the response variable used was the blue shark catches in number (n). In both those cases the number of hooks in each specific set was used in the right-hand side of the model

equations as a variable offset, functioning as an exposure variable, which indicates the number of times the event (catches in number) could have occurred in terms of opportunities (number of hooks used).

Assuming a Poisson distribution the catches of blue sharks in number follow $Y_i \sim P(\mu)$, with expected value and variance defined by:

$$E(Y) = \mu$$

$$Var(Y) = \mu$$

Assuming a NB distribution the catches of blue sharks in number follow $Y_i \sim NB(\mu, k)$, with expected value and variance defined by:

$$E(Y) = \mu$$

$$Var(Y) = \mu + \frac{\mu^2}{k}$$

Where k defines the dispersion parameter; if k has large values (relative to μ^2) then the term μ^2/k approximates 0 and the variance of Y becomes μ, which approximates to a Poisson distribution. This NB can therefore be used instead of a Poisson distribution in cases where the data is overdispersed.

Additionally, and as another possible approach to deal with eventual overdispersion in the data in this type of models, a quasi-Poisson model was also used for comparative purposes with the Poisson and Negative Binomial. The quasi-Poisson estimation allows to deal with this type of problems (over or under-dispersion) as instead of specifying a probability distribution for the data it establishes a relationship between the mean and the variance in the form of a variance function, that can include a dispersion or scale parameter as a multiplicative factor (Faraway, 2006; Zuur *et al.*, 2009). In this model the mean and variance are given by:

$$E(Y) = \mu$$

$$Var(Y) = \varphi \times \mu$$

Where φ is the dispersion or scale parameter. Although in this models the Poisson distribution is not specified, it still uses the same type of model structure in terms of link function. If the φ value = 1 then, in theory, the estimated parameters and standard errors are equal to the Poisson GLM, while values of φ > 1 and φ < 1 refer to overdispersed and underdispersed data, respectively (Zuur *et al.*, 2009).

Finally, another approach to model this data was carried out with Tweedie distributions, also called compound Poisson–Gamma distributions (Dunn, 2004), that are defined by:

$$E(Y) = \mu$$

$$Var(Y) = \varphi \times \mu^p$$

In which φ is the dispersion parameter and *p* is the index parameter. When the index (*p*) parameter has values between 1 and 2, the distribution is continuous for positive real numbers, but has an added discrete mass at 0, which seems appropriate to model CPUE data (continuous data with an added mass of zeros). The index parameter for this specific dataset was calculated outside the models, by maximizing the likelihood profile function of possible values of *p* between 1 and 2.

In all the modeling approaches tested (both GLM and GLMM) the link function used in the models was the *log*. Within the Poisson and NB models, because of this *log* link, the model offsets were defined as *log*(number of hooks), constraining that parameter [*log* (number hooks)] to 1.

III.2.3.3. *Model comparison, validation and goodness-of-fit*

For each model that was run the residuals were plotted and analyzed, to determine visually if major problems were taking place, such as overdispersion problems, the presence of outliers or influencial observations. In general the deviance residuals were used (Zuur *et al.*, 2009), except in the Tweedie models where the quantile residuals were used as recommended by Dunn and Smyth (1996).

For each model the values of the AIC - Akaike Information Criterion (Akaike, 1974), and R^2 - Nagelkerke coefficient of determination (Nagelkerke, 1991), were calculated and used for model comparison in terms of goodness-of-fit. With the final models calculated, the estimated coefficients were compared between approaches (GLM *versus* GLMM) and between different distributions (Gamma, Poisson, Negative Binomial and Tweedie).

Because multiple explanatory variables were used in these models, which may potentially cause multicollinearity problems, Generalized Variance Inflation Factors (GVIF) were calculated for the models main effects (Fox and Monette, 1992). The definition of threshold values for these GVIF seems to be somewhat arbitrary, but as a general rule most authors recommend that values higher than 5 may be cause for concern, while values higher than 10 can indicate serious collinearity problems (Hair *et al.*, 1995; O'Brien, 2007).

Another validation measure used, particularly to determine eventual over-fitting problems in the models, was to measure the prediction error of the models by using a *k*-fold cross validation procedure, in this case using *k*=10 as this was demonstrated to produce the best results in large datasets with n>100 (Borra and DiCiaccio, 2010). With this procedure, the data was randomly divided into 10 equally sized groups, with new models fitted sequentially using data from only 9 groups (training sets) and used to predict and calculate the errors using the group that was left outside (testing set). The measure of error considered was the Mean Squared Error (MSE), defined as:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(\widehat{Y_i} - Y_i\right)^2$$

Where $\widehat{Y}$ is the vector of the fitted predictions and *Y* is the vector of the observed values. This MSE was calculated *k* times in each model being tested (using sequentially each of the training/testing sets), and a final cross-validation error (err^{CV}) calculated as the mean of the *k* MSE values obtained.

Finally, the candidate models were compared and interpreted in terms of blue shark catch rate predictions. Eight possible different scenarios were considered, reflecting theoretical changes in the fishery in terms of spatial, operational and seasonal

aspects. The spatial scenarios considered reflected four general regions in the study area, specifically the SW, NW, NE and SE, considering the 1st and 3rd quartiles of the latitude and longitude gradients; the operational scenarios reflected possible changes between wire and monofilament gangions; and the seasonal scenarios reflected the various seasons along the year.

Data analysis for this paper was carried out with the R Project for Statistical Computing 2.14.1 (R Development Core Team, 2012) and AD Model Builder 10.0 (Fournier *et al.*, 2012). In the R program some additional libraries were used for specific analyses, including the Levene tests and calculation of GVIF (library "car", Fox and Weisberg 2011), fitting Tweedie distributions and models including the maximum likelihood estimation of the index parameter (library "tweedie", Dunn 2010), fractional polynomials models and transformation (library "mfp", Ambler and Benner, 2010), GAM models and plots (library "gam", Hastie, 2011), fitting GLMM models using ADMB (library "glmmADMB, Skaug *et al.*, 2012), fitting GLMM models with the tweedie distribution (library "cplm", Zhang, 2012), fitting GLMM models with penalized quasi-likelihoods (library "MASS", Venables and Ripley, 2002), and cross validation procedures (library "boot", Canty and Ripley, 2011).

III.3. Results

III.3.1. Preliminary data analysis

Of the 533 fishing sets that were monitored and used in this study, positive blue shark catches occurred in almost all (specifically in 525), meaning that only 8 sets (1.52%) had zero catches of blue shark. In the positive sets the catches of blue shark in number varied from 1 to 184, with an overall average of 28.75 (SD=25.73) specimens per set. The catch rates (CPUE, n/1000 hooks) ranged from 0 to 176.9, with an average of 21.13 (SD=19.71) per set. The CPUE distribution was not normally distributed (K-S with Lilliefors correction: $D = 0.142$, p-value < 0.001), and was highly asymmetrical and skewed to the right (Figure III.2).

Blue shark CPUE distribution
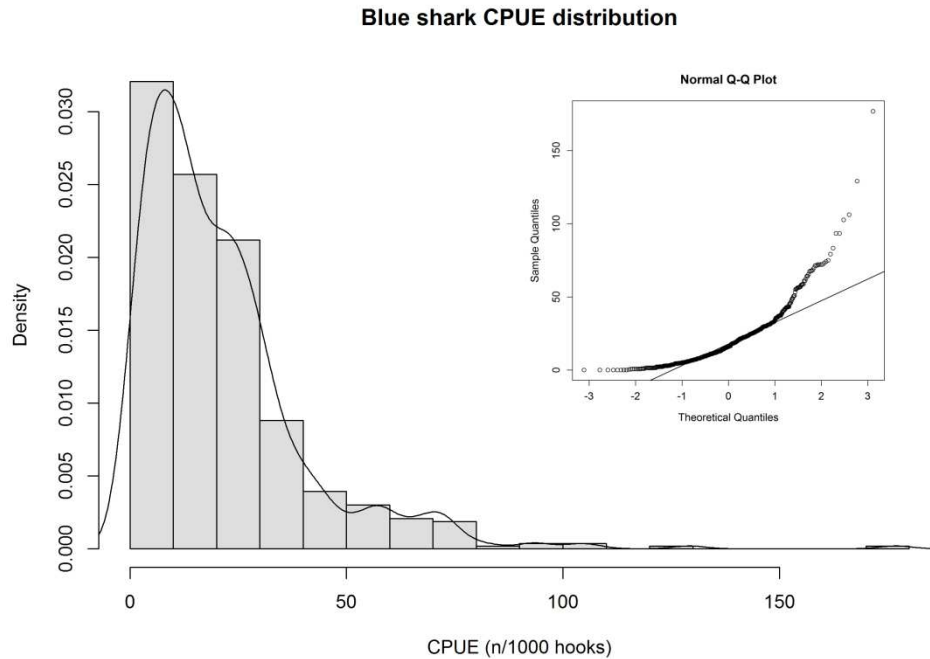
Figure III.2. Distribution of blue shark CPUEs (n/1000 hooks) for the Portuguese pelagic longline fishery in the southern Atlantic Ocean. A smooth kernel density line and a QQ plot with the distribution are also represented.

In terms of preliminary analysis of the explanatory variables, the southern blue shark CPUE had a significant and positive correlation with longitude, wind velocity, mixed layer depth and chlorophyll, and a significant negative correlation with latitude, sea surface high, salinity and SST (Figure III.3). By the contrary, blue shark CPUE did not show a significant correlation with intensity of moon illumination and current velocity (Figure III.3). Some of the possible explanatory variables were also correlated between themselves, such as for example sea surface height that was negatively correlated with both latitude and longitude, or SST that was highly correlated with latitude (Figure III.3).
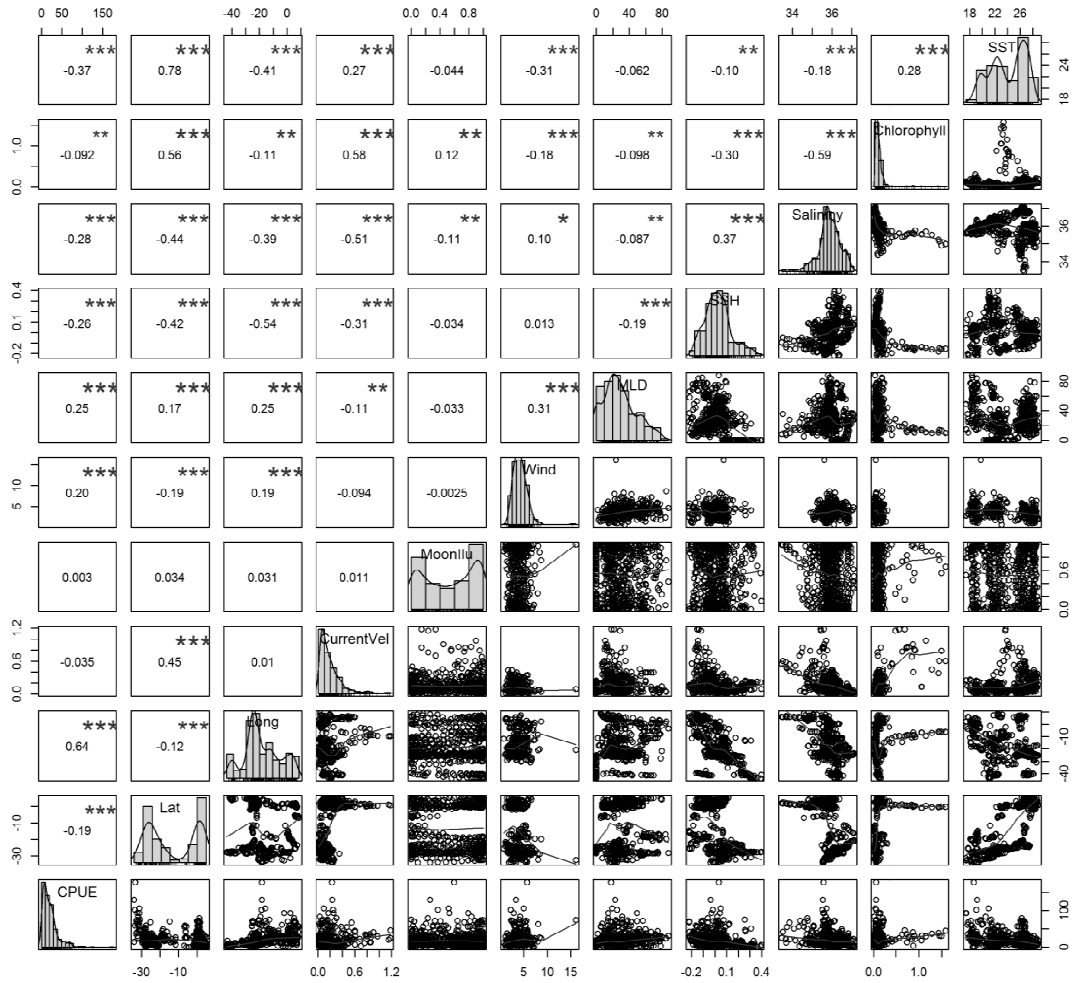
Figure III.3. Scatterplots matrices with the relationships between blue shark CPUE and the candidate continuous explanatory variables used in the models. The lower diagonal panels show the scatterplots with smooth lowess regression lines. The upper diagonal panels show the values of the non-parametric Spearman correlations and the respective significance levels: *** - p-value < 0.01; ** - p-value < 0.05; * - p-value < 0.1. The diagonals show the distribution of the variables with histograms and density lines.

While for most of the variables the information in the database was complete (i.e. there were no missing values), some of the variables that were extrapolated *a posteriori* had this additional problem of having some missing values. Particularly problematic were the cases of the current velocity with 103 missing values (corresponding to 19.3% of the fishing sets), and the wind velocity with 208 missing values (corresponding to 39.0% of the fishing sets). Because of those issues with missing values and because those two more problematic variables were highly correlated with some of the other possible explanatory variables, they were not considered for using in the final models.

When tested in univariate models, the current velocity was not significant (Wald statistic = 1.149, p-value = 0.251; likelihood ratio test p-value = 0.224), while the wind velocity was significant (Wald statistic = 2.646, p-value = 0.009; likelihood ratio test p-value = 0.008). It should be noted that for those univariate models and tests, new subsets of data were build with the removal of all missing values from those variables, so that the likelihood ratio tests for comparing nested models could be performed.

Some variations in the blue shark CPUE were detected in the categorical variables that were analyzed (Figure III.4). Specifically, and using non-parametric statistical tests, significant differences were detected in the blue shark CPUE between years (Kruskal-Wallis: Chi2=69.6, df=3, p-value < 0.001), seasons (Kruskal-Wallis: Chi2=105.5, df=3, p-value < 0.001) and gangion line material (Mann-Whitney: W=4664, p-value < 0.001).
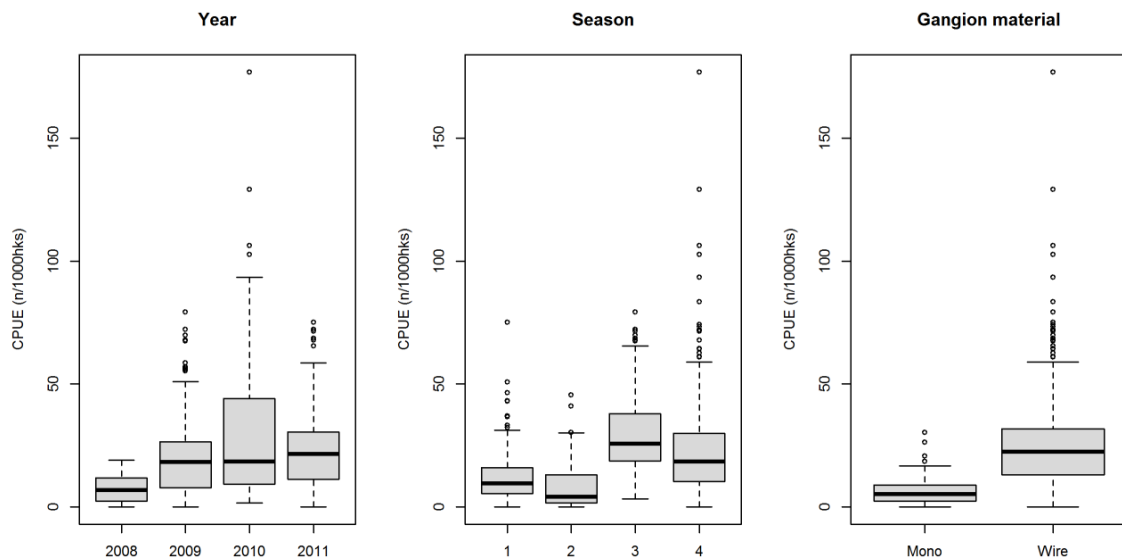
Figure III.4. Boxplots with the blue shark CPUE distribution for the categorical variables analyzed, specifically year of capture (2008, 2009, 2010, 2011), season (1=Jan-Mar, 2=Apr-Jun, 3=Jul-Sep, 4=Oct-Dec) and gangion material used in the longline (monofilament or wire).

III.3.2. Modeling blue shark catch rates with GLM

III.3.2.1. Gamma models adding a constant to the response variable

Given that the percentage of fishing sets with zero catches was relatively small (1.52%), a first approach to model the blue shark CPUEs was to add a small constant (δ, in this case set to 1) to the CPUE values, in a way that the response variable became

CPUE+1. With this transformation the Y_i values no longer contained zeros, but were still highly asymmetrical and skewed to the right (as noted in Figure III.2), and as such a Gamma distribution with a *log* link function seemed a good possible approach.

In this approach, the variables considered significant in the simple effects multivariate GLM, using the variable selection method described in the methods, were gangion material, season, year, latitude, longitude, chlorophyll and SST. In this multivariate simple effect model, the GVIFs were calculated and in all cases the values were < 10, meaning that severe collinearity problems between these explanatory variables were not likely to be occurring. Specifically, the GVIF factors calculated were: Gangion=2.00, Season=7.30, Year=4.73, Latitude=4.92, Longitude=3.99, Chlorophyll=1.61 and SST=6.92.

The functional form of the continuous explanatory variables used in this simple effects model (latitude, longitude, chlorophyll and SST) were tested for the assumption of linearity with the linear predictor using GAM plots, where it was possible to see that the expected CPUE seemed relatively linear with all variables, except with regards to the latitude and in the higher range of the SST values (Figure III.5). Specifically, the expected CPUE seemed to increase towards eastern longitudes and in waters with higher chlorophyll concentration, in both cases in a relatively linear fashion. With regards to the SST there was an increase along most of the observed water temperature values, with a peak at around 26-27 ºC, followed by a slight decrease for the highest temperatures in the range (Figure III.5). The effects of latitude showed a tendency for higher expected CPUE in the southern regions (south of 30ºS), followed by a decrease in CPUEs for the northern areas (towards equatorial waters) with this relation seeming to be non-linear (Figure III.5).
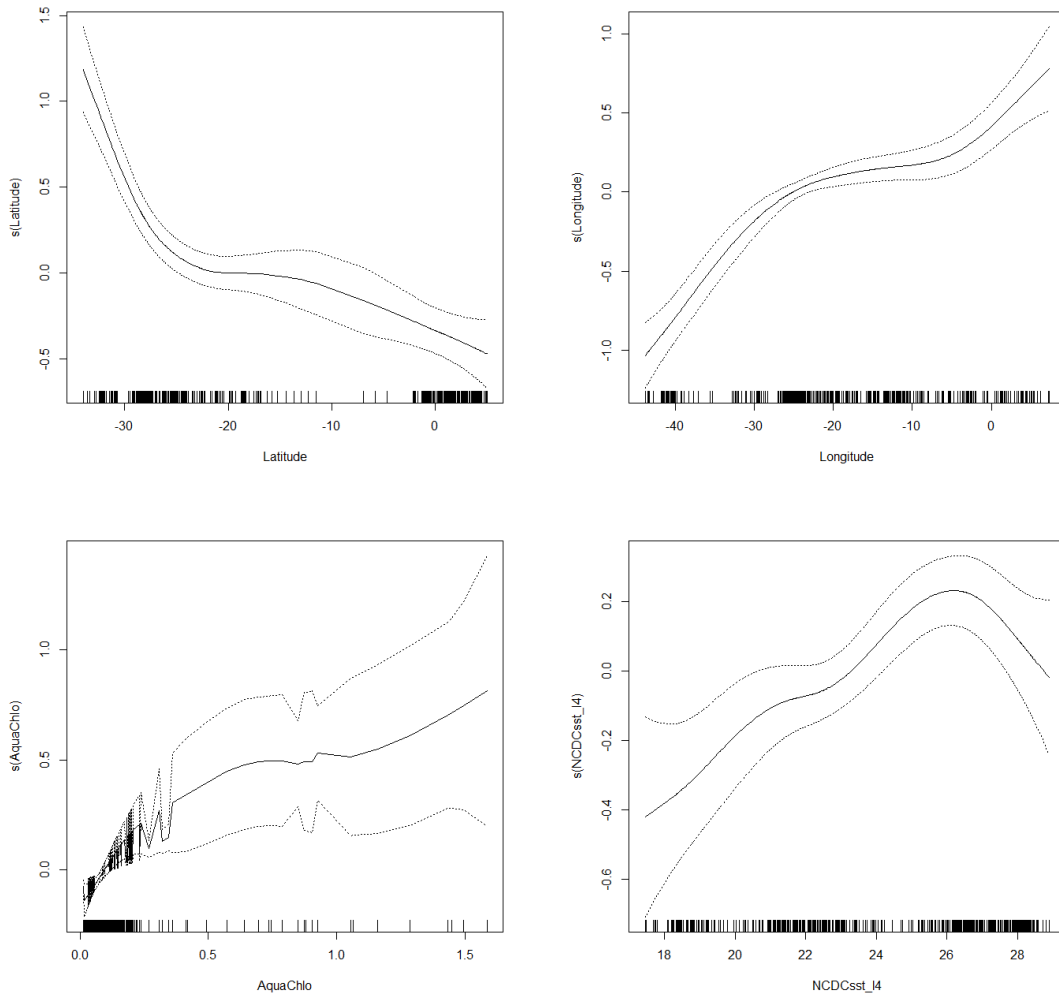
Figure III.5. Generalized Additive Model (GAM) plots with the shape of the continuous explanatory variables (Latitude - top left; Longitude - top right; Chlorophyll - bottom left and Sea Surface Temperature - bottom right) for modeling blue shark catch rates in the Southern Atlantic Ocean.

The eventual non-linearity of those continuous variables was also tested with multivariate fractional polynomials models, and only the latitude required functional form transformation, specifically with a *log* function. The other continuous variables were also transformed to be used in the final models, but only using scale transformations, defined as:

Latitude.t (transformed latitude) = log((Latitude+34.1)/10)

Longitude.t (transformed longitude) = (Longitude+43.8)/10

Chlorophyll.t (transformed chlorophyll-a) = Chlorophyll/0.1

SST.t (transformed sea surface temperature) = SST/10

After this step, all possible 1st degree interactions were tested in the multivariate simple effects model, and the ones significant at the 1% level were included. In this final model, the gangion material was responsible for explaining most of the deviance, followed by the factors year, latitude and longitude, and then the remaining variables. The interactions that were significant at the 1% level and used in this final model were gangion material with longitude, and gangion material with chlorophyll (Table III.1).

Table III.1. Deviance table for the Gamma GLM for predicting the expected blue shark CPUE in the southern Atlantic Ocean. Resid.df are the residual degrees of freedom and Resid.dev is the residual deviance. Significance is given by the p-values.

Variable	Df	Deviance	Resid.df	Resid.dev	p-value
Null			472	362.6	
Gangion	1	138.1	471	224.5	< 0.01
Season	3	12.7	468	211.8	< 0.01
Year	3	40.6	465	171.2	< 0.01
Latitude.t	1	12.2	464	159.0	< 0.01
Longitude.t	1	15.8	463	143.1	< 0.01
Chlorophyll.t	1	2.9	462	140.2	< 0.01
SST.t	1	10.8	461	129.4	< 0.01
Gangion:Longitude.t	1	5.8	460	123.6	< 0.01
Gangion:Chlorophyll.t	1	1.5	459	122.1	0.01

In terms of validation of this Gamma GLM model with a residual analysis, the variation of the deviance residuals along the fitted values seemed to be spread in a relatively random way around the value of zero, and without any major increases or decreases in variance. The QQ plot showed that most of the values were placed along the expected QQ line, except for a few outliers, and a similar conclusion was reached with the histogram of the distribution of the deviance residuals, that also showed a relatively normal distribution except for a few outliers (Figure III.6). In terms of Cook distances, two points were identified that could have leverage and be influential in the estimation of the parameters (Figure III.6).

The outliers identified in this analysis (data points 420 and 421) refer to two fishing sets with zero catches (CPUE=0) but that were carried out under situations that the model was expecting positive catches, specifically because those fishing sets were carried out in southeastern areas of the Atlantic Ocean in Season 4 and using wire

gangions. The remaining zeros in the dataset did not seem to constitute major problems for model fitting because they occurred in areas where those zeros are to be expected more often (northwest region) and in sets using monofilament gangions. In terms of the leverage points detected with the Cook distances (data points 95 and 229), both refer to fishing sets with positive catches that took place in the southwest Atlantic region.
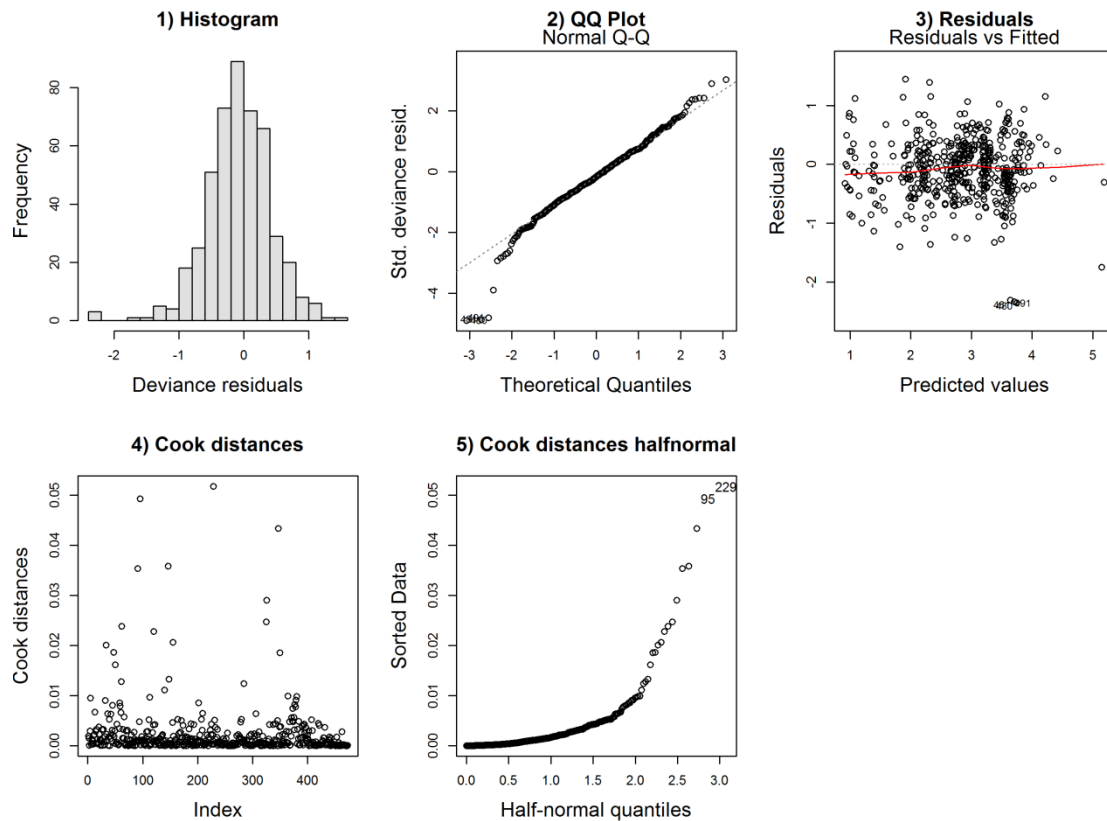
Figure III.6. Residual analysis for the Gamma GLM. The graphics plotted for the outlier analysis show 1) residuals histogram, 2) QQ-plot and 3) residuals along the predicted values; and the analysis to detect observations with leverage are 4) Cooks distances along the data and 5) Cook distances half-normal plot.

By removing each of the two outlier values the changes in the model parameters were minimal, while by removing the influential values the changes were a little larger, but still < 20% for all parameters (Table III.2). This means that, in general, the values detected in the residual analysis are not influential in the estimated model parameters. Aditionally, when those values were removed the improvements in terms of goodness-of-fit (R^2) were only slightly noticeable (Table III.2).

Table III.2. Estimated parameters (with the respective SE) of the original model using all data-points, and four alternative models each excluding one possible outlier or influential value. The differences in each of the estimated parameters in percentage are given, as well as each model goodness-of-fit measures in terms of R^2.

Parameter	All dataset		Excluding point 420			Excluding point 421			Excluding point 95			Excluding point 229		
	Estimate	SE	Estimate	SE	Dif(%)	Estimate	SE	Dif(%)	Estimate	SE	Dif(%)	Estimate	SE	Dif(%)
(Intercept)	-1.26	0.48	-1.26	0.48	0.0	-1.26	0.48	-0.7	-1.47	0.49	16.4	-1.40	0.47	10.9
GangionWire	0.78	0.20	0.78	0.19	-0.8	0.78	0.19	-0.6	0.78	0.19	-0.2	0.82	0.19	4.4
Season2	0.60	0.14	0.61	0.14	1.4	0.61	0.14	1.2	0.64	0.14	6.5	0.62	0.14	3.8
Season3	0.96	0.11	0.96	0.11	0.2	0.96	0.11	0.2	0.99	0.11	3.2	0.98	0.11	1.7
Season4	0.47	0.08	0.48	0.08	0.7	0.48	0.08	0.6	0.48	0.08	2.3	0.49	0.08	2.5
Year2009	-0.20	0.11	-0.20	0.11	2.1	-0.20	0.11	1.0	-0.21	0.11	2.6	-0.23	0.11	13.0
Year2010	0.56	0.11	0.56	0.11	-0.4	0.56	0.11	-0.1	0.55	0.11	-0.5	0.55	0.11	-1.2
Year2011	0.56	0.11	0.57	0.11	0.5	0.57	0.11	0.9	0.57	0.11	1.9	0.56	0.11	-1.2
Latitude.t	-0.46	0.05	-0.46	0.05	0.2	-0.46	0.05	0.1	-0.50	0.06	8.4	-0.47	0.05	1.9
Longitude.t	0.63	0.08	0.63	0.08	0.5	0.63	0.08	0.4	0.64	0.08	1.6	0.62	0.08	-2.0
Chlorophyll.t	-0.25	0.12	-0.25	0.12	0.6	-0.25	0.12	0.5	-0.24	0.12	-3.7	-0.21	0.12	-16.3
SST.t	0.93	0.18	0.93	0.18	-0.1	0.93	0.18	-0.6	1.01	0.19	8.4	0.97	0.18	4.0
GangionWire:Longitude.t	-0.35	0.09	-0.35	0.09	-0.3	-0.35	0.09	-0.1	-0.36	0.09	0.9	-0.33	0.08	-5.6
GangionWire:Chlorophyll.t	0.32	0.12	0.32	0.12	0.2	0.32	0.12	0.3	0.31	0.12	-2.2	0.26	0.12	-17.5
R^2	0.654		0.666			0.665			0.663			0.660		

III.3.2.2. Models for count data: Poisson, quasi-Poisson and Negative Binomial

A second possible approach to modeling was carried out with discrete distributions that can by themselves handle some zero values, specifically Poisson, quasi-Poisson and NB. In this case, and because the purpose was to compare modeling approaches, this type of models were applied to the complete model estimated previously. With these models, and similarly to what had been observed before in the Gamma model, the gangion material remained the most important variable for explaining part of the deviance, followed by the year effect, longitude, latitude and season (Table III.3).

The major difference and decision for using one of those three possible models has to do with the dispersion of the data. While the Poisson distribution assumes a dispersion parameter of 1, the NB model estimates θ (the dispersion parameter for the model) allowing for under or overdispersion in the data. In general, if this dispersion parameter in the NB is similar to 1 then a Poisson model may be more adequate, while much smaller or larger values represent under or overdispersed data that should not be modeled with a Poisson distribution. In this specific case the θ value was estimated at 4.510 (SE=0.367) meaning that the data is overdispersed and therefore a Poisson model might be producing biased estimates, especially with regards to the standard deviations of the parameters.

Other possible solutions for this overdispersion are to either use a quasi-Poisson model, in this case with a calculated dispersion parameter of 7.177, or to directly include the dispersion parameter in the model estimates to manually correct the SE of the parameters. With these approaches the point estimates of the models remain the same as in the Poisson model, but the SE values of the parameters are increased due to the overdispersion of the data. Consequently, and in this specific case, the interaction between gangion line material and chlorophyll lost significance when considering a significance level of 1% (Table III.3).

Table III.3. Deviance table for the Poisson, quasi-Poisson and Negative Binomial GLM for predicting the expected blue shark catches (in number) in the southern Atlantic Ocean, and using the number of hooks as a model offset. Resid.df are the residual degrees of freedom and Resid.dev is the residual deviance. Significance of the variables is given by the p-values.

Variable	Df	Deviance	Resid.df	Resid.dev	p-value
Poisson Model (with dispersion parameter)					
Null			472	9414.5	
Gangion	1	3229.6	471	6184.9	< 0.01
Season	3	227.9	468	5957.0	< 0.01
Year	3	1486.3	465	4470.7	< 0.01
Latitude.t	1	515.3	464	3955.4	< 0.01
Longitude.t	1	266.5	463	3688.9	< 0.01
Chlorophyll.t	1	69.7	462	3619.2	< 0.01
SST.t	1	188.8	461	3430.4	< 0.01
Gangion:Longitude.t	1	69.6	460	3360.7	< 0.01
Gangion:Chlorophyll.t	1	27.9	459	3332.8	0.05
NB Model					
Null			472	1515.8	
Gangion	1	578.3	471	937.5	< 0.01
Season	3	46.4	468	891.1	< 0.01
Year	3	178.9	465	712.3	< 0.01
Latitude.t	1	51.8	464	660.5	< 0.01
Longitude.t	1	58.2	463	602.3	< 0.01
Chlorophyll.t	1	12.6	462	589.7	< 0.01
SST.t	1	46.5	461	543.3	< 0.01
Gangion:Longitude.t	1	28.5	460	514.8	< 0.01
Gangion:Chlorophyll.t	1	6.1	459	508.7	0.01
quasi-Poisson Model					
Null			472	9414.5	
Gangion	1	3229.6	471	6184.9	< 0.01
Season	3	227.9	468	5957.0	< 0.01
Year	3	1486.3	465	4470.7	< 0.01
Latitude.t	1	515.3	464	3955.4	< 0.01
Longitude.t	1	266.5	463	3688.9	< 0.01
Chlorophyll.t	1	69.7	462	3619.2	< 0.01
SST.t	1	188.8	461	3430.4	< 0.01
Gangion:Longitude.t	1	69.6	460	3360.7	< 0.01
Gangion:Chlorophyll.t	1	27.9	459	3332.8	0.05

This overdispersion problem was also confirmed when comparing the residuals plots, with the variance of the residuals increasing with the expected values for the Poisson model, while in the NB model the residuals seemed more randomly distributed

61

along the value of zero, with no major increasing or decreasing trends in their variance (Figure III.7). Additionally, in the Poisson model there was one data point that is probably a severe outlier (point 141, observed in the residuals plots of Figure III.7) and another point with a very large influence (point 95, observed in the cooks distances plots of Figure III.7) and the removal of those could have large implications in the estimated parameters of the model. Therefore, if a model for counts was used to model this data, the Negative Binomial distribution approach that can handle overdispersed data would be preferable and probably more accurate than a Poisson distribution. Alternatively, a quasi-Poisson model or a Poisson model including a dispersion parameter to correct the SE of the parameters could also be used.
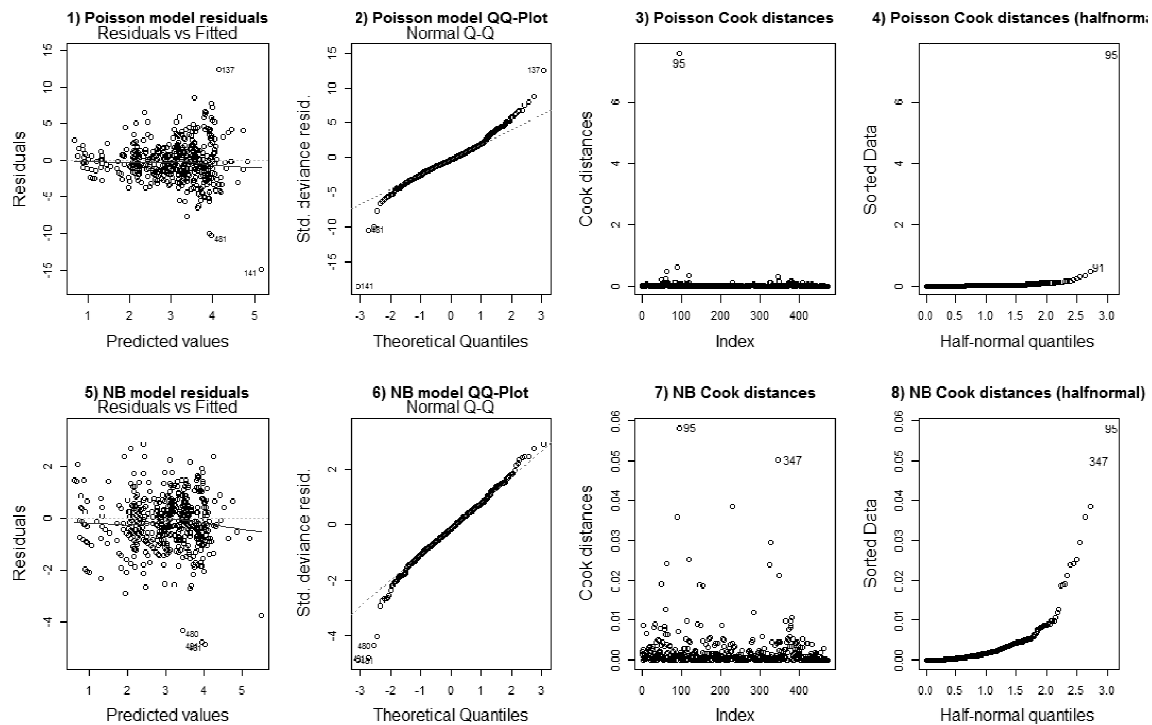
Figure III.7. Residual analysis for the Poisson (plots 1-4) and Negative Binomial (plots 5-8) GLM models for explaining the blue shark catch rates in the Southern Atlantic Ocean. The graphics for the outlier analysis show the deviance residuals along the fitted values and the QQ-plots, while the plots to detect observations with leverage represent the Cook distances along the data and the Cook distances half-normal plots.

III.3.2.3. Tweedie models

The last modeling approach was carried out with a Tweedie model. In this case, the index parameter (p) estimated by maximum likelihood with a likelihood function

profile, was estimated to be 1.492 (95% CI = [1.415, 1.571]) (Figure III.8). In this model, and again similarly to the previous cases, the gangion line material remained the most important variable for explaining part of the deviance, followed by the year effect, longitude, latitude and season (Table III.4).
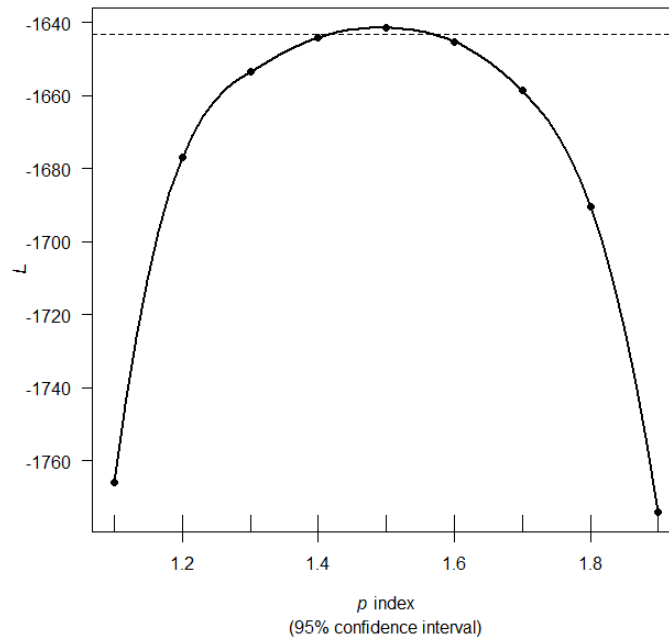
Figure III.8. Likelihood function profile (with 95% confidence interval) for the Tweedie model p-index (*p*), applied to the blue shark catch rate data in the Southern Atlantic Ocean.

Table III.4. Deviance table for the Tweedie GLM for predicting the expected blue shark catch rates (CPUE) in the southern Atlantic Ocean. Resid.df are the residual degrees of freedom and Resid.dev is the residual deviance. Significance of the variables is given by the p-values.

Variable	Df	Deviance	Resid.df	Resid.dev	p-value
Null			472	1725.34	
Gangion	1	620.92	471	1104.42	< 0.01
Season	3	46.36	468	1058.06	< 0.01
Year	3	236.34	465	821.72	< 0.01
Latitude.t	1	78.66	464	743.05	< 0.01
Longitude.t	1	53.81	463	689.24	< 0.01
Chlorophyll.t	1	14.15	462	675.09	< 0.01
SST.t	1	44.67	461	630.42	< 0.01
Gangion:Longitude.t	1	22.54	460	607.88	< 0.01
Gangion:Chlorophyll.t	1	6.03	459	601.85	0.02

In terms of model diagnostics, the variation of the quantile residuals along the fitted values seemed to be spread along the value of zero in a relatively random way, and without major increases or decreases in variance. The QQ plot showed that most of the values were placed along the expected QQ line, again with the exceptions of a few outliers, even though in this case they seemed to be more extreme than in the case of the Gamma GLM (Figure III.9). In terms of influential values detected with the Cook distances only a few points were detected, but in this case the leverage of the data point 95 (also identified in the previous models) seemed to be much more influential than in some of the other models, in particular when compared to the Gamma GLM (Figure III.9).

Figure III.9. Residual analysis for the Tweedie GLM model for explaining the blue shark catch rates in the Southern Atlantic Ocean. The graphics plotted for the outlier analysis show the 1) quantile residuals histogram, 2) QQ-plot and 3) quantile residuals along the fitted values; the analysis to detect influential observations are the 4) Cooks distances along the data and 5) Cook distances half-normal plot.
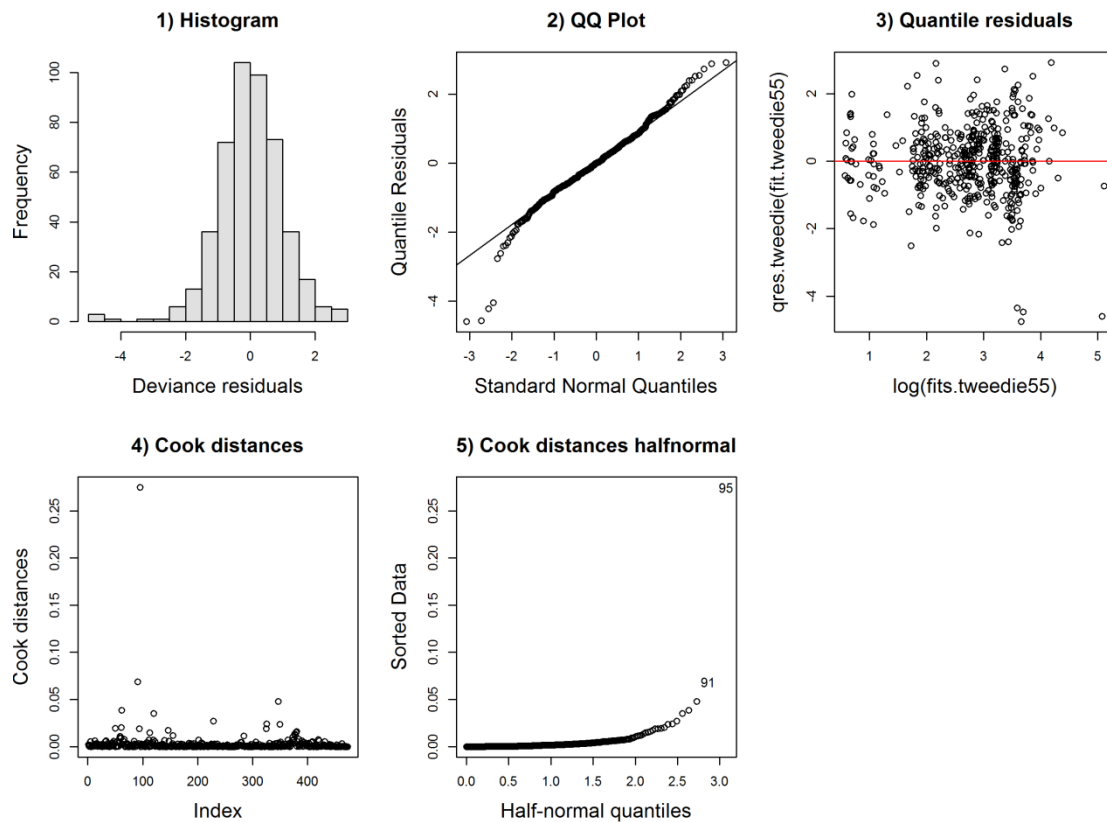
III.3.2.4. Comparing GLM models

A comparison of the estimated parameters for the various candidate GLM models is presented in Table III.5. It is interesting to note that regardless of the distribution used (Gamma, NB, quasi-Poisson or Tweedie) all parameters are providing relatively similar values, and always with the same signal, i.e. they are consistently either positive or negative, meaning that relatively similar conclusions are taken with all the candidate models in terms of the influence of the various explanatory variables in the expected blue shark catch rates. However, some parameters were either significant or non-significant depending on the distribution used. The most obvious case was the year 2009 (within the variable year) that was significant at the 5% level only when using the NB model, while it was only significant at the 10% level with the Gamma and Tweedie models, and non-significant with the quasi-Poisson model. Similarly, the interaction between gangion and chlorophyll was on the limit of significance at the 1% level using the Gamma and Negative Binomial models, while with the other distributions it was not significant at 1%.

In terms of the different models goodness-of-fit, the R^2 values obtained were 0.654, 0.641, 0.655 and 0.636 for the models using respectively the Gamma, Poisson (value equal to the quasi-Poisson), NB and Tweedie distributions. Those values were all relatively similar, with only very slightly better fits for the Gamma and NB, and a worst fit for the Poisson. In terms of AIC the obtained values were 3280.2, 3561.9 and 3312.8 for the Gamma, Negative Binomial and Tweedie models. Using this criterion, the Gamma model seemed to be performing better than the others, followed by the Tweedie and finally the NB. In this regards, the original Poisson model was performing much worse with a calculated AIC of 5575.9, which is consistent to the problems detected with the residuals in terms of overdispersion in the data.

Table III.5. Parameter estimation for the GLM Gamma, Negative Binomial, quasi-Poisson and Tweedie models for predicting the expected blue shark catches in the southern Atlantic Ocean. The Gamma and Tweedie models are predicting blue shark catch rates in CPUE (CPUE+1 for the Gamma), while the Negative Binomial and quasi-Poisson are predicting blue shark catches in numbers (n).

Model parameters	Gamma			Negative Binomial			quasi-Poisson			Tweedie		
	Est.	SE	p-value	Est.	SE	p-value	Est.	SE	p-value	Est.	SE	p-value
(Intercept)	-1.26	0.48	< 0.01	-8.83	0.53	< 0.01	-8.11	0.56	< 0.01	-1.57	0.51	< 0.01
GangionWire	0.78	0.20	< 0.01	1.07	0.24	< 0.01	0.99	0.38	0.01	1.01	0.27	< 0.01
Season2	0.60	0.14	< 0.01	0.62	0.16	< 0.01	0.58	0.17	< 0.01	0.60	0.16	< 0.01
Season3	0.96	0.11	< 0.01	1.01	0.11	< 0.01	0.90	0.10	< 0.01	0.96	0.11	< 0.01
Season4	0.47	0.08	< 0.01	0.51	0.09	< 0.01	0.43	0.09	< 0.01	0.47	0.09	< 0.01
Year2009	-0.20	0.11	0.07	-0.25	0.12	0.04	-0.19	0.15	0.20	-0.22	0.13	0.08
Year2010	0.56	0.11	< 0.01	0.56	0.13	< 0.01	0.69	0.14	< 0.01	0.62	0.13	< 0.01
Year2011	0.56	0.11	< 0.01	0.56	0.12	< 0.01	0.51	0.13	< 0.01	0.53	0.12	< 0.01
Latitude.t	-0.46	0.05	< 0.01	-0.49	0.06	< 0.01	-0.37	0.04	< 0.01	-0.44	0.05	< 0.01
Longitude.t	0.63	0.08	< 0.01	0.80	0.10	< 0.01	0.72	0.16	< 0.01	0.74	0.11	< 0.01
Chlorophyll.t	-0.25	0.12	0.04	-0.30	0.15	0.04	-0.39	0.22	0.08	-0.32	0.16	0.05
SST.t	0.93	0.18	< 0.01	1.04	0.20	< 0.01	0.81	0.18	< 0.01	0.93	0.19	< 0.01
GangionWire:Longitude.t	-0.35	0.09	< 0.01	-0.50	0.11	< 0.01	-0.44	0.17	0.01	-0.46	0.12	< 0.01
GangionWire:Chlorophyll.t	0.32	0.12	0.01	0.37	0.15	0.01	0.44	0.22	0.05	0.38	0.17	0.02
R^2	0.654			0.655			0.636			0.641		
AIC	3280.2			3561.9						3312.8		

For the 10-fold cross-validation procedure, the estimated mean MSE values were 236.5, 430.7, 394.6 and 245.2 for the Gamma, Poisson (equal value in the quasi-Poisson), NB and Tweedie models, respectively. Again with this approach the models for explaining the continuous catch rates (Gamma and Tweedie) seemed to be performing better (with lower associated errors) than the models for count data (Poisson and NB), again with the Gamma performing a little better than the Tweedie. However, the results obtained with this analysis should be interpreted with care, as the models being compared are not predicting exactly the same response, i.e., the continuous distribution models are predicting blue shark CPUE (CPUE+1 in the Gamma), while the count data models are predicting blue shark captures in numbers (n).

III.3.3. Modeling blue shark catch rates with GLMM

The same distributions that were used with the GLM analysis were used in GLMM approaches, using the effect of Vessel as a random effect. In terms of residual analysis of those new models again there seemed to be problems with the Poisson distribution, with the residuals showing an increasing variance along the fitted values, while no major problems were found in the other models (Figure III.10). The few possible outliers and influential values that had been previously detected in the GLM models seemed to be less severe in the GLMM models, meaning that the GLMM approach seems to be modeling better the variability in the dataset, including those more extreme values.
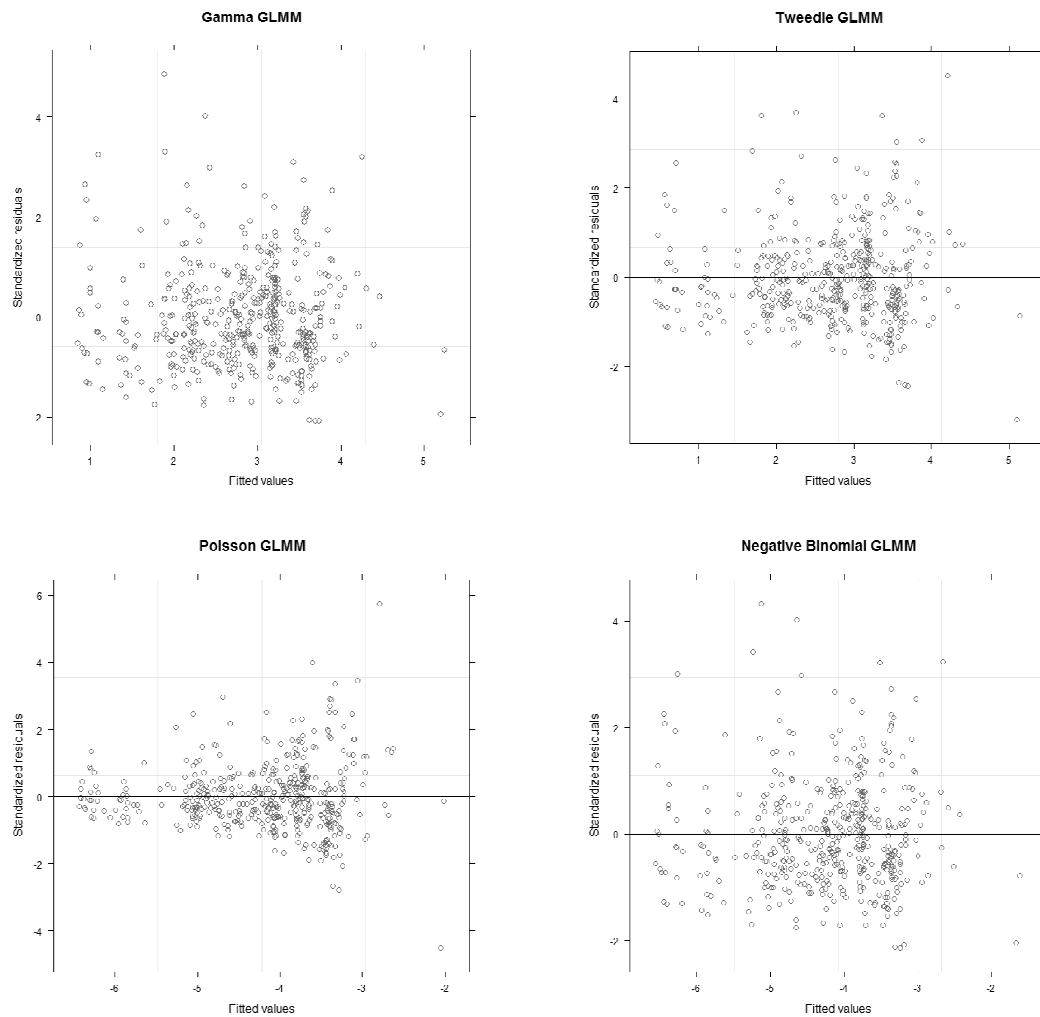
Figure III.10. Residual analysis (standardized residuals along the fitted values) for the Gamma, Poisson, Negative Binomial and Tweedie GLMM models.

In terms of AIC comparisons for the GLMM, the obtained values were 3279.4, 3560.2 and 3310.8 for the Gamma, NB and Tweedie models, respectively. Again, and similarly to the AIC analysis for the GLM, the best fit was obtained with the Gamma, followed by the Tweedie and NB models. Finally, and similarly to what was obtained for the original Poisson model, the AIC was much worse for the Poisson GLMM, in this case of 5529.7, again demonstrating that even thought the GLMM can account for more variability in the data than the GLM, the goodness-of-fit is still worse in a Poisson model due to overdispersion in the data. Therefore, and like in the GLM case, either a NB or a quasi-Poisson GLMM model should be used instead if a discrete distribution model was chosen.

When comparing the AIC values between GLM and GLMM approaches, the GLMM models always had slightly better AIC values, even though the differences were generally small. Specifically, the differences in the AIC values (Δ AIC) between GLM and GLMM approaches of the same family were 0.8, 46.2, 1.7 and 2.0 for the Gamma, Poisson, NB and Tweedie models, respectively.

With these GLMM models, and as defined in the model formulation, the random variable (vessel effect) followed a Normal distribution with mean 0, and in this case with estimated standard deviations of 0.0937, 0.1276 and 0.1205 for the Gamma, NB and Tweedie models, respectively. In terms of the fixed effects, the obtained coefficient values were again providing the same signal, i.e., being consistently either positive or negative and with similar estimated values (Table III.6). Still, and even though the differences in estimates were relatively minor, again some of the model parameters (e.g. 2009 within the factor year) were either significant or non-significant depending on the specific distribution used.

Table III.6. Parameter estimation for the GLMM Gamma, Negative Binomial, and Tweedie models for predicting the expected blue shark catches in the southern Atlantic Ocean. The Gamma and Tweedie models are predicting blue shark catch rates in CPUE (CPUE+1 for the Gamma), while the Negative Binomial is predicting blue shark catches in numbers (n). The Poisson GLMM results are not shown due to the overdispersion problems identified in the residual analysis.

Model parameters	Gamma GLMM			Negative Binomial GLMM			Tweedie GLMM		
	Estimate	SE	p-value	Estimate	SE	p-value	Estimate	SE	p-value
(Intercept)	-0.47	0.49	0.34	-8.34	0.58	< 0.01	-1.07	0.56	0.05
GangionWire	0.50	0.19	0.01	0.93	0.25	< 0.01	0.87	0.28	< 0.01
Season2	0.57	0.13	< 0.01	0.55	0.17	< 0.01	0.51	0.16	< 0.01
Season3	0.83	0.11	< 0.01	0.89	0.12	< 0.01	0.84	0.12	< 0.01
Season4	0.41	0.08	< 0.01	0.45	0.09	< 0.01	0.42	0.09	< 0.01
Year2009	-0.18	0.10	0.08	-0.28	0.12	0.02	-0.26	0.13	0.04
Year2010	0.54	0.11	< 0.01	0.53	0.12	< 0.01	0.58	0.13	< 0.01
Year2011	0.46	0.11	< 0.01	0.43	0.13	< 0.01	0.41	0.13	< 0.01
Latitude.t	-0.44	0.05	< 0.01	-0.49	0.06	< 0.01	-0.44	0.05	< 0.01
Longitude.t	0.49	0.08	< 0.01	0.77	0.11	< 0.01	0.71	0.12	< 0.01
Chlorophyll.t	-0.28	0.12	0.01	-0.37	0.15	0.01	-0.39	0.17	0.02
SST.t	0.81	0.18	< 0.01	0.94	0.20	< 0.01	0.84	0.19	< 0.01
GangionWire:Longitude.t	-0.22	0.08	0.01	-0.44	0.11	< 0.01	-0.39	0.12	< 0.01
GangionWire:Chlorophyll.t	0.33	0.12	< 0.01	0.43	0.15	< 0.01	0.44	0.17	0.01

III.3.4. Model interpretation and examples of predictions

In terms of model interpretation, and by looking to the various candidate models coefficients presented in Figure III.11, some interpretations can be taken with regards to the effects of the explanatory variables in the expected blue shark catch rates. In terms of seasonality it is expected for the fishery to have lower catch rates of blue sharks during the season 1 (baseline season, Jan-Mar), while higher catch rates are expected during the other seasons, specifically with highest catches during season 3 (Jul-Sep). With regards to the environmental variables, higher catch rates are expected with increasing SST and decreasing chlorophyll concentrations, but because there is a significant interaction between chlorophyll and gangion material this is only true for the baseline gangion material level (i.e. monofilament); if wire gangions are used then the significant interaction between these two variables needs to be taken into account. In terms of spatial variables, the expected catch rates increase towards southern latitudes and eastern longitudes, but again a significant interaction between gangion material and longitude needs to be taken into account. Finally, and in terms of annual variability, the expected catch rates deceased from 2008 to 2009, followed by an increase in the more recent years of 2010 and 2011.

In terms of general variability of the parameters, the gangion material coefficient showed the highest variability, while the latitude effect had a very small variability. Even though this was in general common for all candidate models, some models (for example the quasi-Poisson GLM) tended to have higher variability in all the parameters, including the loss of significance at the 5% level in some cases (e.g. parameter for year 2009).

In terms of model predictions, the blue shark catch rates were predicted for some possible theoretical scenarios under various fishing conditions, with the results presented in Table III.7. One immediate conclusion that can be taken is that the estimates, within each specific scenario, are relatively consistent between all candidate models, with predicted catch rate values relatively similar. Those scenarios were chosen to allow for spatial, operational and seasonal interpretations of the model outputs, and it is noted in terms of spatial predictions, that the expected blue shark catch rates are higher in the southern areas (when compared to the equatorial region), and also higher in the eastern areas (towards the African continent), when compared to the western

regions (Table III.7). In terms of operational characteristics of the fishing gear the expected catch rates with monofilament gangion are lower than when wire gangion are used. Finally, in terms of seasonal predictions the blue shark catch rates seem to be particularly higher during season 3 (between July and September) and lower in season 1 (between January and March), while the expected catch rates in seasons 2 and 4 (April to June and October to December) have similar values in the middle of the range (Table III.7).
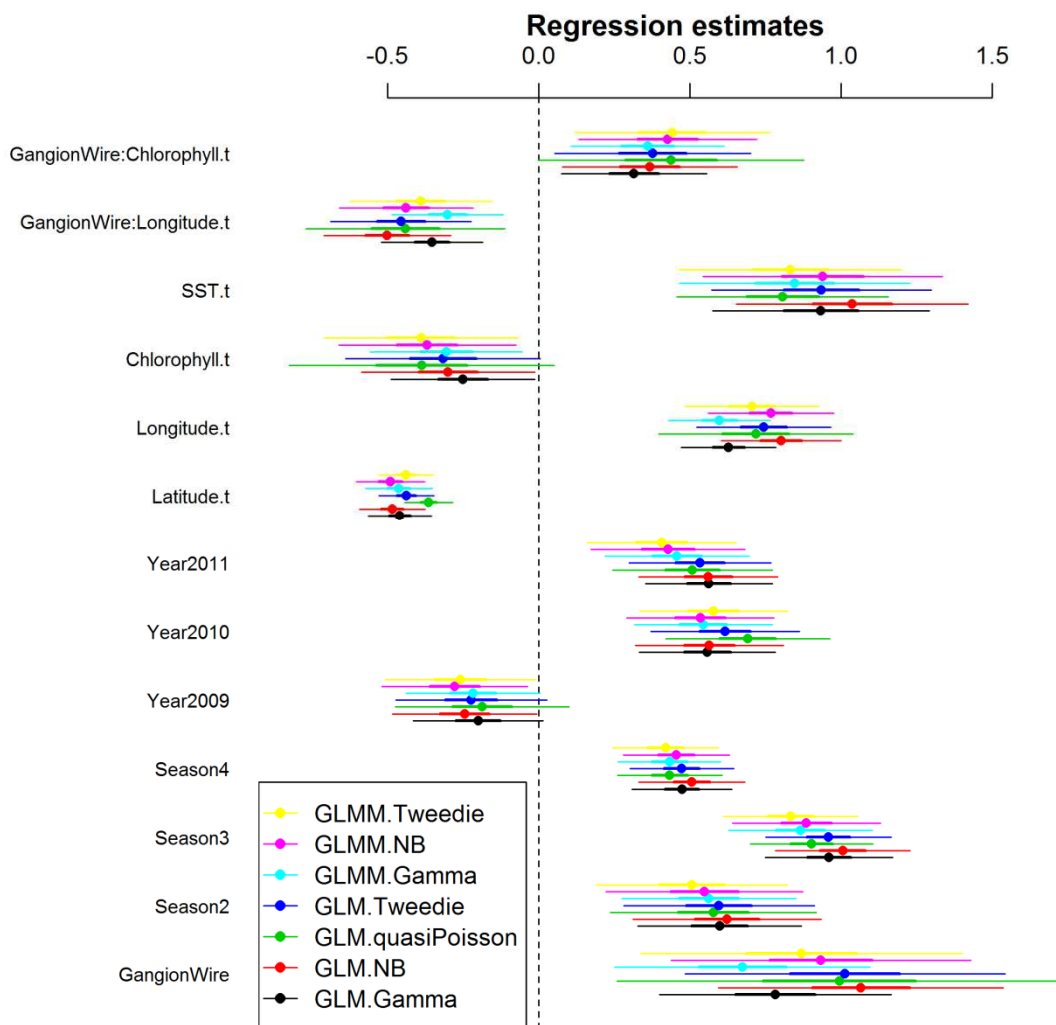

Figure III.11. Comparison of parameter estimates for the various candidate GLM and GLMM models for predicting blue shark catch rates in the Southern Atlantic Ocean. For each parameter it is indicated the point estimate, the 50% (thick lines) and the 95% (thin lines) confidence intervals.

Table III.7. Predictions of blue shark catch rates (n/1000 hooks) for possible theoretical fishing scenarios. The comparisons are established between one possible theoretical scenario (scenario 1) and others with spatial, operational and seasonal variations. For each model it is indicated the point estimate and respective standard deviations (between brackets). For the GLMMs only the variability associated with the fixed effects is being considered.

Scenario	Effects being compared	Specific teorethical conditions	GLM Gamma	GLM NB	GLM q.Poisson	GLM Tweedie	GLMM Gamma	GLMM NB	GLMM Tweedie
1	Spatial: SW area, all else remains constant	Southwest, Season 1, Year 2010, Wire gangion, median SST and Chlorophyll	19.5 (1.69)	19.3 (1.7)	21.9 (1.89)	20.6 (1.78)	22.1 (0.78)	22.2 (0.84)	23.3 (1.19)
2	Spatial: Same as scenario 1 but in SE area; all else remains constant	Southeast, Season 1, Year 2010, Wire gangion, median SST and Chlorophyll	29.3 (2.53)	29.6 (2.62)	32.4 (2.73)	30.9 (2.65)	34.1 (0.87)	35.4 (0.93)	36.3 (1.19)
3	Spatial: Same as scenario 1 but in NW area; all else remains constant	Northwest (equatorial), Season 1, Year 2010, Wire gangion, median SST and Chlorophyll	8.6 (1.20)	8.7 (1.17)	12.0 (1.49)	10.0 (1.28)	9.7 (0.77)	9.9 (0.83)	11.3 (1.19)
4	Spatial: Same as scenario 1 but in NE area; all else remains constant	Northeast (equatorial), Season 1, Year 2010, Wire gangion, median SST and Chlorophyll	13.1 (1.57)	13.2 (1.58)	17.7 (1.95)	15.0 (1.71)	15.3 (0.86)	15.7 (0.92)	17.6 (1.19)
5	Operational: Same as scenario 1 but with monofilament gangion; all else remains constant	Southwest, Season 1, Year 2010, Monofilament gangion, median SST and Chlorophyll	13.3 (1.21)	12.8 (1.19)	13.3 (1.47)	13.2 (1.26)	14.9 (0.72)	14.4 (0.77)	14.6 (1.19)
6	Seasonal: Same as scenario 1 but Apr-Jun; all else remains constant	Southwest, Season 2, Year 2010, Wire gangion, median SST and Chlorophyll	36.4 (5.12)	36.0 (5.60)	39.0 (6.50)	37.4 (5.77)	39.6 (0.80)	38.5 (0.84)	38.7 (1.20)
7	Seasonal: Same as scenario 1 but Jul-Sep; all else remains constant	Southwest, Season 3, Year 2010, Wire gangion, median SST and Chlorophyll	52.6 (5.93)	52.8 (6.05)	54.0 (5.23)	53.7 (5.55)	53.9 (0.77)	53.9 (0.82)	53.8 (1.19)
8	Seasonal: Same as scenario 1 but Oct-Dec; all else remains constant	Southwest, Season 4, Year 2010, Wire gangion, median SST and Chlorophyll	32.0 (2.31)	32.1 (2.35)	33.8 (2.12)	33.1 (2.23)	34.6 (0.80)	35.0 (0.85)	35.5 (1.19)

III.4. Discussion

This study presents and compares several possible modeling approaches for explaining the blue shark catch rates in pelagic longline fisheries carried out by the Portuguese fleet in the southern Atlantic Ocean. The model comparison strategy presented seems adequate, as different model types (GLM *versus* GLMM) and distributions (Gamma, Negative Binomial, Poisson and Tweedie) can be considered for using with this type of data. It was interesting to note and conclude that regardless of the model used, the estimates of the final parameters were relatively similar, meaning that all of the considered approaches (with the exception of the Poisson distribution due to overdispersion of the data) can probably be considered for use in this case. With regards to the Poisson model, the problems identified in the residuals and the value of the estimated dispersion parameter show that the dataset is highly overdispersed, and that in this case either using a quasi-Poisson model, or introducing a correction in the SE of the parameters, should be used instead.

In terms of deviance explained by the variables, the effects of the gangion material seemed to be the most important factor for explaining the blue shark catch rates, followed by the effects of year, spatial and seasonal effects. The major effect seen for the gangion material (with higher catch rates when wire gangions are used) seems to make sense from a biological point of view, as the wire gangions probably make bite-offs of the sharks less likely to occur while in the longlines, as was shown for the Brazilian fleet by Afonso *et al.* (2012). The conclusions regarding the spatial and seasonal effects taken from this study also seem to be in line to what is common knowledge from the commercial fishermen in the fleet, that empirically mention higher catches of sharks particularly in the second semester of the year and towards eastern longitudes, closer to mainland Africa. Those empirical observations by the fishermen (also in line with the conclusions from the models) may be related with spatial and seasonal migratory movements of this species in the Southern Atlantic Ocean, that are still not entirely know, but as with most pelagic sharks may involve complex spatial and seasonal migratory movements.

In the Atlantic Ocean, some previous studies have focused on modeling blue shark catch rates for other areas and using data from other fishing fleets. In the southwest Atlantic off Brazil, and using data from the Brazilian national and chartered

fleets, Carvalho *et al.* (2011) used GAM models to predict blue shark catch rates using spatial (latitude and longitude), seasonal (month) and environmental (SST and chlorophyll) effects. Even though the study areas are different, it was noteworthy that some of the effects seen by Carvalho *et al.* (2011) were similar to the conclusions of our study, such as for example the effect of latitude, also with higher catch rates in southern Brazil and lower towards equatorial waters; and the seasonal effects, also with higher catch rates between June and August, which is very similar to the higher catch rates of July to September seen in our study. One possible limitation in the Carvalho *et al.* (2011) study was that operational effects of the fishing gear, as well as the effects of the different vessels operating in the fleet, were not considered. Also for the Atlantic but in the Mediterranean Sea, Megalofonou *et al.* (2009b) also modeled the blue shark catch rates taken by the Italian and Greek fleets with a GLM, and also saw strong spatial (both in terms of longitude and latitude) and seasonal effects. One difference in the Megalofonou *et al.* (2009b) study was that, contrary to our study, the SST was not significant in the final models.

For other Oceans in the world such as the Pacific, Bigelow *et al.* (1999) used GAM models to evaluate the blue shark catch rates in the Hawaiian based US pelagic longline fleet, and again the spatial variables (latitude and longitude) seemed to be playing a major role in terms of effects. Other variables that were used in the Bigelow *et al.* (1999) GAM models included operational variables such as soaking time and number of light stick used, and environmental variables such as SST, wind velocity and lunar index. It was interesting to note that compared to our study, the lunar index was significant in the Bigelow *et al.* (1999) study in the Pacific Ocean, while in our case it was tested but excluded due to non-significance. By the contrary, the wind velocity that was used by Bigelow *et al.* (1999) in the Pacific was also significant in our preliminary univariate models, but excluded from the final multivariate models due to excessive missing values in the data.

One point that seems to be common in most of the previous studies (all also using fishery-dependant data from commercial fleets, as was the case of our study) is that the effects of having different vessels in the fleet do not seem to be accounted for in the final models. One example of a study were the vessel effect was also incorporated in the models, and even suggested as the variable explaining more of the deviance, was the study by Campana *et al.* (2009) for predicting blue shark mortality rates (binomial

models). However, the approach used in the Campana *et al.* (2009) study was different than the approach used in our study, as Campana *et al.* (2009) included the vessel effect as a categorical variable in the GLM (using 27 degrees of freedom to estimate all the parameters needed), while we used it as a random variable in a GLMM approach. In our opinion, our approach has the advantage of allowing for a measure of variability to be associated with a "general" vessel effects, rather than having to estimate one parameter for each of the n-1 vessels in the study as was done by Campana *et al.* (2009).

One of the difficulties in modeling this type of data is the presence of zeros in the response variable combined with positive continuous values (usually CPUE) in the non-zero cases. Several approaches to deal with this difficulty are available, and adding a constant to the response variable seems to be one possible and reasonable approach, particularly when the proportion of zeros is relatively low (Shono, 2008). That was the particular case of our study, where the proportion of zeros in the blue shark catch data from the southern Atlantic represented only 1.52% of the fishing sets. With this approach, one important issue that can be highly influential in the results is to decide the value of the small constant to be added, and in our case we chose to use 1, as that seems to be the most common in fisheries biology (e.g. Punt *et al.*, 2000, Ortiz and Arocha, 2004). However, it should be noted that other possibilities exist and have been used, such as for example adding 10% of the mean of the response variable (Campbell, 2004). One eventual problem with this later approach (10% of the mean) is that the actual value to be added depends on the dataset, so in specific cases of datasets with small mean values the constant being added will be very small, which can be problematic. In fact, while several approaches for adding different constants exist and have been used, one common consensus among researchers is that using very small values (e.g. 1^{-100}) can be problematic (Xiao, 1997), especially in models using a *log* link function, due to the properties of the *log* function as it approaches zero.

Another possibility also tested in our study was to consider discrete distributions that can account for some zero observations, such as the Poisson and the NB. In those cases, the major difference is that while in the Poisson distribution the expected variance is equal to the mean, the NB can account for overdispersion in the data, which seems to be a common characteristic of many biological and ecological datasets, sometimes due to the excess of zeros (Cunningham and Lindenmayer, 2005; Martin *et al.*, 2005). In both those cases the models were built in a conceptually different way, as

the response variable considered was the number of specimens caught (counts) per set, and the number of hooks per set were used in the set of explanatory variables, as an offset variable. This means that the parameter (β) for the variable "numbers of hooks" did not need to be calculated, and because a *log* link was used in the equations the parameter estimate for *log* (number hooks) was constrained to 1. Another eventual limitation of this approach with discrete distributions is that conceptually it can only be used if the models are being created to model the catches in number. While that was the case of the present study (i.e. catches accounted in number (n) and the respective CPUE calculated in n/1000 hooks), many fisheries biology studies use biomass (kg) instead of numbers, with the respective CPUE calculated as kg/1000 hooks. In such cases, and if a discrete distribution was to be applied, the response variable would be the catches in kg, that as mentioned by Maunder and Punt (2004) would be best modeled as a continuous (not a discrete) distribution.

In terms of the models used, it was interesting to note that both the GLM and GLMM approaches were producing relatively similar and consistent results. Even when comparative GLMM models were built using a top-down variable selection strategy as recommended by Zuur *et al.*, (2009), the variables excluded due to non-significance were exactly the same that had been previously excluded in the GLM models using a stepwise variable selection approach as recommended by Hosmer and Lemeshow (2000).

In terms of estimation, and while in the GLM modeling approach the use of maximum likelihood methodology is common to estimate the model parameters, for the GLMM approaches there are complications with the likelihood-based model fitting that comes from the difficulty of integrating over the random effects (Venables and Ripley, 2002). As mentioned by Zuur *et al.* (2009) the GLMM models are currently on the frontier of statistical research, and there is still very limited available documentation aimed to biologists and ecologists. For the simpler GLMM, such as the ones used in this work, in which the random effects were the subject-specific intercepts (in this case vessel specific) which were assumed to be Normally distributed, approximate methods based on quasi-likelihoods or numerical approximations to integrals can be used (Venables and Ripley, 2002). However, for more complex and elaborate specifications of the random effects the only possible integration seems to be based on Markov Chain Monte Carlo (MCMC) schemes within Bayesian statistics, and most of those are

currently ongoing research topics (Venables and Ripley, 2002). Bolker *et al.* (2008) revised the applications of GLMM to ecology data and ranked the currently available likelihood approximation methods as: 1) penalized quasi-likelihood, 2) Laplace approximation, 3) Gauss-Hermite quadrature and 4) Markov Chain Monte Carlo; with those methods being progressively more accurate in estimation, but also more complex and technically challenging to program and calculate. In our study the final GLMM models used Laplace approximation methods to estimate the model parameters, which seemed to be providing accurate estimates without being extremely time consuming in terms of model fitting. Even though it was not fully presented in this paper, we also tested to fit the GLMM models with penalized quasi-likelihoods (PQL), and while the estimated parameters were virtually equal to the Laplace approximations, the PQL had the disadvantage of not allowing the calculation of goodness-of-fit measures based on likelihoods, such as the AIC values for each model. While this is not essential for the model fitting process and parameter estimation, the AIC values were one of the goodness-of-fit measures that were chosen to be used for comparing different model approaches.

Another type of model framework that could have conceivably been used to deal with this type of fishery-dependant data would have been to use Generalized Estimation Equations (GEE) instead of GLMM. An example of an application of such methods, with a comparison to GLM models, was provided in the work of Coelho *et al.* (in press) for modeling blue shark mortality data (binomial models) in this same pelagic longline fishery. The main differences between those two modeling approaches (GEE *versus* GLMM) are that while the GEE use a correlation matrix to correct the estimated parameters values and standard deviations, the GLMM effectively calculate parameters for the variables with random effects. That is why Venables and Ripley (2002) stated that the mixed models are sometimes referred to as conditional or subject specific models, where the estimated coefficients apply to each individual but not necessarily to the population, while GEE are marginal models with the estimated regression coefficients representing the population means but not necessarily the individuals. In our study, and because the random variables considered where the vessels, the parameters estimated in the GLMM could now be used to predict vessel-specific blue shark catch rates, as well as to have an idea on the variability associated with the vessels in the fleet.

As a conclusion, we believe that this paper presents new and important information regarding blue shark catch rates in the Portuguese pelagic longline fishery in the Southern Atlantic Ocean, with the comparison of techniques and modeling approaches showing, in general, consistent results. The results presented in this paper can now be used to infer on the factors that are affecting more or less those blue shark catch rates, and eventually lead to proposals for more efficient management strategies on this shark species in the southern Atlantic Ocean.

CHAPTER IV.   FINAL REMARKS AND CONCLUSIONS

This study presented new information with statistical models created for predicting blue shark catch rates and at-haulback mortality rates in the Portuguese longline fishery targeting swordfish in the Atlantic Ocean. With the work presented it is now possible to better understand and predict the impacts of this fishery in this shark species, both in terms of the catches and mortality rates.

The chapter focusing on modeling the mortality rates has already been adapted for a peer-reviewed scientific journal and accepted for publication (Coelho *et al.*, 2013). The most important conclusion of that chapter (and paper) is that the at-haulback mortality of blue sharks is highly dependent on the specimen size, with specimens tending to have lower mortality rates as they grow into larger sizes. On the other hand, as the specimens grow in size the odds-ratios of dying also decrease (due to the non-linearity of the size effect), meaning that even though the probabilities of a small shark dying are higher than in a larger shark, if both are allowed to grow for example 10 cm in size the odds of dying are more reduced in the smaller specimen. Also interesting was to note that the differences between generalized linear models and generalized estimation equations was small, meaning the that even though the data comes from a fishery-dependant source, the use of generalized linear models that assume data independency, is probably also valid in this case.

The chapter that focused on modeling the catch rates is still under preparation to be submitted for a peer-reviewed journal, and some aspects that were not possible to incorporate in this thesis will be added before submission. Specifically, adding more data from different vessels that were not available at the time of writing this thesis. One shortcoming of that chapter is precisely the fact that only data from two vessels was available and used in the analysis, making the usefulness and interpretation of the random vessels effects within the generalized mixed models limited. Still, the results presented are already important and informative even if considered preliminary. Another result worthy to be highlighted was that the operational, spatial and seasonal variables are important to explain part of the blue shark catch rates in the Atlantic. Thus, with the model equations presented it is now possible to predict those catch rates with their associated uncertainties, under different fishing scenarios.

We believe that the results herein presented are important and may have implications in terms of fisheries management and conservation of the species. For example, if discards eventually become mandatory for this species, we can now use the catch rates models to predict the catches of blue sharks in different regions, seasons and fishing condition across the Southern Atlantic Ocean, and then use the mortality models to predict how many of those sharks are more likely to be caught and discarded already dead, allowing to have an estimate of the efficiency of such management measure. Another example of application of these models is the possibility to estimate the size-specific mortality rates of those specimens below a certain minimum size, namely if minimum landing sizes are implemented, thus allowing to have an estimate of the efficiently of such management measure. Finally, another example would be the establishment of any type of spatial and/or seasonal fishing closures, as the catch rates models allow to predict the impacts of such closures (in terms of reduced catches) on the blue shark stocks.

Even though the issues of the catch and mortality rates were addressed in this study, others that are also important in terms of management and conservation have remained unsolved. Therefore, a suggestion for future research priorities includes the development of statistical models aiming the prediction of the catch-at-size of the specimens, as it would also be important to determine if there are any spatial and/or seasonal effects in the catch sizes. If that is occurring, and once the respective models and equations are prepared, we could start predicting the expected catch-at-size under different fisheries management scenarios that would further contribute to our knowledge of the species and impacts of the fishery in the blue shark populations.

Finally, we would like to highlight the fact that these results of this work have been and will continue to be presented to the Shark Working Group of the Scientific Committee of Research and Statistics of ICCAT. Therefore, it is expected further discussions and development of these models to provide the best scientific advice for fisheries managers, further contributing to promote the sustainable exploitation of these migratory shark populations in the Atlantic Ocean.

CHAPTER V. REFERENCES

Afonso, A.S., Hazin, F.H.V., Carvalho, F., Pacheco, J.C., Hazin, H., Kerstetter, D.W., Murie, D., Burgess, G.H., 2011. Fishing gear modifications to reduce elasmobranch mortality in pelagic and bottom longline fisheries off Northeast Brazil. *Fisheries Research*, 108: 336-343.

Afonso, A.S., Santiago, R., Hazin, H., Hazin, F.H.V. 2012. Shark bycatch and mortality and hook bite-offs in pelagic longlines: Interactions between hook types and leader materials. *Fisheries Research*, 131-133: 9-14.

Agresti, A. 2002. *Categorical Data Analysis*. 2nd Edition. John Wiley and Sons, New Jersey, 691pp.

Aires-da-Silva, A.M., Gallucci, V.F. 2007. Demographic and risk analyses applied to management and conservation of the blue shark (*Prionace glauca*) in the North Atlantic Ocean. *Marine and Freshwater Research*, 58: 570-580.

Aires-da-Silva, A.M., Hoey, J.J., Gallucci, V.F. 2008. A historical index of abundance for the blue shark (*Prionace glauca*) in the western North Atlantic. *Fisheries Research*, 92: 41-52.

Akaike, H. 1974. A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19: 716-723.

Ambler, G., Benner, A. 2010. mfp: Multivariable Fractional Polynomials. R package version 1.4.9. Available from: http://CRAN.R-project.org/package=mfp.

Barker, M.J., Schluessel, V. 2005. Managing global shark fisheries: suggestions for prioritizing management strategies. *Aquatic Conservation: Marine and Freshwater Ecosystems* 15: 325–347.

Baum, J.K., Myers, R.A., Kehler, D.G., Worm, B., Harley, S.J., Doherty, P.A. 2003. Collapse and conservation of shark populations in the Northwest Atlantic. *Science*, 299: 389-392.

Beerkircher, L.R., Cortés, E., Shivji, M. 2003. Characteristics of shark bycatch observed on pelagic longlines off the southeastern United States, 1992–2000. *Marine Fisheries Review*, 64: 40-49.

Belcher, C.N., Jennings, C.A. 2011. Identification and evaluation of shark bycatch in Georgia's commercial shrimp trawl fishery with implications for management. *Fisheries Management and Ecology*, 18: 104–112.

Bigelow, K.A., Boggs, C.H., He, X. 1999. Environmental effects on swordfish and blue shark catch rates in the US North Pacific longline fishery. *Fisheries Oceanography*, 8: 178-198.

Blanco-Parra, M.P., Galván-Magaña, F., Márquez-Farías, F. 2008. Age and growth of the blue shark, *Prionace glauca* Linnaeus, 1758, in the Northwest coast off Mexico. *Revista de Biología Marina y Oceanografía*, 43: 513-520.

Bleck, R., Halliwell, G., Wallcraft, A., Carrol, S., Kelly, K., Rushing, K. 2002. *Hybrid Coordinate Ocean Model User's Manual*. Version 2.0.01. 199pp.

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 and Evolution*, 24: 127-135.

Bonfil, R. 1994. Overview of world elasmobranch fisheries. *FAO Fisheries Technical Paper*, 341. FAO: Rome. 119pp.

Borra, S., DiCiaccio, A. 2010. Measuring the prediction error. A comparison of cross-validation, bootstrap and covariance penalty methods. *Computational Statistics and Data Analysis*, 54: 2976-2989.

Buencuerpo, V., Rios, S., Moron, J. 1998. Pelagic sharks associated with the swordfish, *Xiphias gladius*, fishery in the eastern North Atlantic Ocean and the Strait of Gibraltar. *Fishery Bulletin*, 96: 667-685.

Burgess, G.H., Beerkircher, L.R., Caillet, G.M., Carlson, J.K., Cortés, E., Goldman, K.J., Grubbs, R.D., Musick, J.A., Musyl, M.K., Simpfendorfer, C.A. 2005. Is

the collapse of shark populations in the Northwest Atlantic Ocean and Gulf of Mexico real? *Fisheries*, 30: 19-26

Cailliet, G.M. 1992. Demography of the central California population of the leopard shark (*Triakis semifasciata*). *Australian Journal of Marine and Freshwater Research*, 43: 183-93

Cailliet, G.M., Martin, L.K., Harvey, J.T., Kusher, D., Welden, B.A. 1983. Preliminary studies on the age and growth of blue, *Prionace glauca*, common thresher, *Alopias vulpinus*, and shortfin mako, *Isurus oxyrinchus*, sharks from Californian waters. *In*: Tunas, billfishes and sharks. Proceedings of an international workshop on age determination of oceanic pelagic fishes (Prince, E.D., Pulos, L.M., eds.), pp. 179-188. NOAA Technical Report NMFS.

Camhi, M., Fowler, S., Musick, J., Bräutigam, A., Fordham, S. 1998. *Sharks and their relatives: ecology and conservation*. IUCN, Gland and Cambridge. 39pp.

Campana, S.E., Marks, L., Joyce, W., Kohler, N.E. 2006. Effects of recreational and commercial fishing on blue sharks (*Prionace glauca*) in Atlantic Canada, with inferences on the North Atlantic population. *Canadian Journal of Fisheries and Aquatic Science*, 63: 670–682.

Campana, S.E., Joyce, W., Manning, M.J. 2009. Bycatch and discard mortality in commercially caught blue sharks *Prionace glauca* assessed using archival satellite pop-up tags. *Marine Ecology Progress Series*, 387: 241–253.

Campana, S.E., Dorey, A., Fowler, M., Joyce, W., Wang, Z., Wright, D., Yashayaev, I. 2011 Migration pathways, behavioural thermoregulation and overwintering grounds of blue sharks in the northwest Atlantic. *PLoS ONE*, 6: e16854.

Campbell, D., Battaglene, T., Shafron, W. 1992. Economics of resource conservation in a commercial shark fishery. *Australian Journal of Marine and Freshwater Research*, 43: 251-262.

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: 209–227.

Canty, A., Ripley, B. 2011. boot: Bootstrap R (S-Plus) Functions. R package version 1.3-3.

Carstensen, B., Plummer, M., Laara, E., Hills, M. 2011. Epi: A Package for Statistical Analysis in Epidemiology. R package version 1.1.24. Available from: http://CRAN.R-project.org/package=Epi.

Carruthers, E.H., Schneider, D.C., Neilson, J.D. 2009. Estimating the odds of survival and identifying mitigation opportunities for common bycatch in pelagic longline fisheries. *Biological Conservation*, 142: 2620-2630.

Cartamil, D., Santana-Morales. O., Escobedo-Olvera, M., Kacev, D., Castillo-Geniz, L., Graham, J.B., Rubin, R.D., Sosa-Nishizaki, O. 2011. The artisanal elasmobranch fishery of the Pacific coast of Baja California, Mexico. *Fisheries Research*, 108: 393–403.

Carvalho, F.C., Murie, D.J., Hazin, F.H.V., Hazin, H.G., Mourato, B.L., Travassos, P., Burgess, G.H. 2010. Catch rates and size composition of blue sharks (*Prionace glauca*) caught by the Brazilian pelagic longline fleet in the southwestern Atlantic Ocean. *Aquatic Living Resources*, 23: 373-385.

Carvalho, F.C., Murie, D.J., Hazin, F.H.V., Hazin, H.G., Mourato, B.L., Burgess, G.H. 2011. Spatial predictions of blue shark (*Prionace glauca*) catch rate and catch probability of juveniles in the Southwest Atlantic. *ICES Journal of Marine Science*, 68: 890-900.

Castillo-Geniz, J.L., Marquez-Farias, J.F., Cruz, M.C.R., Cortés, E., Prado, A.C. 1998. The Mexican artisanal shark fishery in the Gulf of Mexico: towards a regulated fishery. *Marine and Freshwater Research*, 49: 611-620.

Castro, J.A., Mejuto, J.A. 1995. Reproductive parameters of blue shark, *Prionace glauca*, and other sharks in the Gulf of Guinea. *Marine and Freshwater Research*, 46: 967-73.

Castro, J.I., Woodley, C.M., Brudek, R.L. 1999. A preliminary evaluation of the status of shark species. *FAO Fisheries Technical Paper*, 380. FAO: Rome. 72pp.

Clarke, S., Milner-Gulland, E.J., Bjørndal, T. 2007. Social, economic, and regulatory drivers of the shark fin trade. *Marine Resource Economics*, 22: 305-327.

Coelho, R., Erzini, K. 2008. Effects of fishing methods on deep water shark species caught as by-catch off southern Portugal. *Hydrobiologia*, 606: 187-193.

Coelho, R., Bentes, L., Gonçalves, J.M.S., Lino, P., Ribeiro, J., Erzini. K. 2003. Reduction of elasmobranch by-catch in the hake semipelagic near-bottom longline fishery in the Algarve (Southern Portugal). *Fisheries Science*, 69: 293-299.

Coelho, R., Lino, P.G., Santos, M.N. 2011. Standardized CPUE of major shark species caught by the Portuguese longline fishery in the Indian Ocean. *Indian Ocean Tuna Commission Technical Document*, IOTC-2011-WPEB07-36. 14pp.

Coelho R., Fernandez-Carvalho J., Lino P.G., Santos M.N. 2012a. An overview of the hooking mortality of elasmobranchs caught in a swordfish pelagic longline fishery in the Atlantic Ocean. *Aquatic Living Resources*, 25: 311-319.

Coelho, R., Santos, M.N., Amorim, S. 2012b. Effects of hook and bait on targeted and bycatch fishes in an Equatorial Atlantic pelagic longline fishery. *Bulletin of Marine Science*, 88: 449-467.

Coelho, R., Infante, P., Santos, M.N. 2013. Application of generalized linear models and generalized estimation equations to model at-haulback mortality of blue sharks captured in a pelagic longline fishery in the Atlantic Ocean. *Fisheries Research*, 145: 66-75.

Compagno, L.J.V. 1984. *Sharks of the World. An annotated and illustrated catalogue of shark species to date. Part II (Carcharhiniformes).* FAO Fisheries Synopsis. FAO, Rome. 655pp.

Compagno, L.J.V. 2001. *Sharks of the World. An annotated and illustrated catalogue of shark species known to date. Volume 2. Bullhead, mackerel and carpet sharks (Heterodontiformes, Lamniformes and Orectolobiformes).* FAO Species Catalogue for Fishery Purposes. FAO, Rome. 269pp.

Compagno, L.J.V., Dando, M., Fowler, S. 2005. *Sharks of the World*. Collins, London. 368pp.

Cortés, E. 1995. Demographic analysis of the Atlantic sharpnose shark, *Rhizoprionodon terraenovae*, in the Gulf of Mexico. *Fishery Bulletin*, 93: 57-66.

Cortés, E. 1999. Standardized diet compositions and trophic levels of sharks. *ICES Journal of Marine Science*, 56: 707-717.

Cortés, E. 2000. Life history patterns and correlations in sharks. *Reviews in Fisheries Science*, 8: 299-344.

Cortés, E. 2002 Incorporating uncertainty into demographic modeling: application to shark populations and their conservation. *Conservation Biology*, 16: 1048-1062.

Cortés, E. 2007. Chondrichthyan demographic modeling: an essay on its use, abuse and future. *Marine and Freshwater Research*, 58: 4-6.

Cortés, E., Arocha, F., Beerkircher, L., Carvalho, F., Domingo, A., Heupel, M., Holtzhausen, H., Santos, M.N., Ribera, M., Simpfendorfer, C. 2010. Ecological risk assessment of pelagic sharks caught in Atlantic pelagic longline fisheries. *Aquatic Living Resources*, 23: 25-34.

Cortés, E., Domingo, A., Miller, P., Forselledo, R., Mas, F., Arocha, F., Campana, S., Coelho, R., Silva, C.D., Holtzhausen, H., Keene, K., Lucena, F., Ramirez, K., Santos, M.N., Semba-Murakami, Y., Yokawa, K. 2012. Expanded ecological risk assessment of pelagic sharks caught in Atlantic pelagic longline fisheries. *ICCAT SCRS Document*, SCRS/2012/167. 55pp.

Cunningham, R.B., Lindenmayer, D.B. 2005. Modeling count data of rare species: some statistical issues. *Ecology*, 86: 1135-1142.

Diaz G.A., Serafy J.E. 2005. Longline-caught blue shark (*Prionace glauca*): factors affecting the numbers available. *Fishery Bulletin*, 103: 720-724.

Dunn, P.K. 2004. Occurrence and quantity of precipitation can be modeled simultaneously. *International Journal of Climatology*, 24: 1231–1239.

Dunn, P.K. 2010. tweedie: Tweedie exponential family models. R package version 2.0.7. Available from: http://www.r-project.org/.

Dunn, K. P., Smyth, G. K. 1996. Randomized quantile residuals. *Journal of Computational and Graphical Statistics*, 5: 236-244.

Ellis, J.R., Cruz-Martínez, A., Rackham, B.D., Rogers, S.I. 2005. The distribution of chondrichthyan fishes around the British Isles and implications for conservation. *Journal of Northwestern Atlantic Fisheries Science*, 35: 195-213.

FAO, 2012. Fisheries Global Information System (FIGIS). Available from: www.fao.org/fishery/figis/en.

Faraway, J.J. 2006. *Extending the Linear Model with R: Generalized Linear, Mixed Effects and Nonparametric Regression Models*. Chapman & Hall/CRC, Boca Raton, FL. 331pp.

Fitzmaurice, G., Davidian, M., Verbeke, G., Molenberghs, G. 2009. *Longitudinal Data Analysis*. Chapman & Hall/CRC, Boca Raton, FL. 618pp.

Fournier, D.A., Skaug, H.J., Ancheta, J., Ianelli, J., Magnusson, A., Maunder, M.N., Nielsen, A., Sibert, J. 2012. AD Model Builder: using automatic differentiation for statistical inference of highly parameterized complex nonlinear models. *Optimization Methods and Software*, 27: 233-249.

Fox, J., Monette, G. 1992. Generalized collinearity diagnostics. *Journal of the American Statistical Association*, 87: 178-183.

Fox, J., Weisberg, S. 2011. *An {R} Companion to Applied Regression*. 2nd Edition. Thousand Oaks CA: Sage.

Francis, M.P. 1998. New Zealand shark fisheries: development, size and management. *Marine and Freshwater Research*, 49: 579-591.

Francis, M.P., Griggs, L.H., Baird, S.J. 2001. Pelagic shark bycatch in the New Zealand tuna longline fishery. *Marine and Freshwater Research*, 52: 165-178.

Froese, R., Pauly, D. (Editors). 2012. FishBase version 12/2012. Available from: www.fishbase.org.

Fushiki, T. 2011. Estimation of prediction error by using K-fold cross-validation. *Statistics and Computing*, 21: 137–146.

Grogan, E.D., Lund, R. 2004. The origin and relationships of early Chondrichthyes. *In*: Biology of sharks and their relatives (Carrier, J.C., Musick, J.A., Heithaus, M.R., eds.), pp. 3-32. CRC Press, Boca Raton, Florida.

Halekoh, U., Hojsgaard, S., Yan J. 2006. The R Package geepack for Generalized Estimating Equations. *Journal of Statistical Software*, 15: 1-11.

Hair, J.F., Anderson, R.E., Tatham, R.L., Black, W.C. 1995. *Multivariate Data Analysis*, 3rd Edition. New York: Macmillan.

Hastie, T. 2011. gam: Generalized Additive Models. R package version 1.04.1. Available from: http://CRAN.R-project.org/package=gam.

Hazin, F.H.V., Couto, A.A., Kihara, K., Otsuka, K., Ishino, M., Boeckman, C.E., Leal, E.C. 1994. Reproduction of the blue shark *Prionace glauca* in the south-western equatorial Atlantic Ocean. *Fisheries Science*, 60: 487-491.

Hilborn, R. 2006. Faith-based fisheries. *Fisheries*, 31: 554-555.

Hilborn, R., Walters, C.J. 1992. *Quantitative Fisheries Stock Assessment: Choice, Dynamics and Uncertainty*. Chapman & Hall, London. 570pp.

Hiraoka, Y., Yokawa, K., 2012. Standardized CPUE for blue sharks caught by Japanese longline fishery in the Atlantic Ocean. *Collective Volume of Scientific Papers ICCAT*, 68: 1893-1907.

Hosmer, D.W., Lemeshow, S. 2000. *Applied Logistic Regression*, 2nd Ed. Wiley, New York. 375pp.

Hurley, P.C.F. 1998. A review of the fishery for pelagic sharks in Atlantic Canada. *Fisheries Research*, 39: 107-113.

HYCOM. 2012. Hybrid Coordinate Ocean Model. Available from: http://hycom.org/hycom.

ICCAT. 2006-2009. *ICCAT Manual*. International Commission for the Conservation of Atlantic Tunas. Available from: http://www.iccat.int.

ICCAT, 2009. Report of the 2008 shark stock assessments meeting. *Collective Volume of Scientific Papers ICCAT*, 64: 1343-1491.

ICCAT, 2012a. Task 1 Data: nominal catch information. Available from: http://www.iccat.int.

ICCAT, 2012b. Report from the 2012 shortfin mako stock assessment and ecological risk assessment meeting. 11-18 June, Olhão, Portugal. 105pp. Available from: http://www.iccat.int.

Jiang, J. 2007. *Linear and Generalized Linear Mixed Models and Their Applications*. Springer, New York. 257pp.

Kerstetter, D.W., Graves, J.E., 2006. Effects of circle versus J-style hooks on target and non-target species in a pelagic longline fishery. *Fisheries Research*, 80, 239-250.

Kilpatrick, K., Podestá, G., Evans, R. 2001. Overview of the NOAA/NASA advanced very high resolution radiometer Pathfinder algorithm for sea surface temperature and associated matchup database. *Journal of Geophysical Research*, 106: 9179-9197.

Komsta, L., Novomestky, F. 2012. moments: Moments, cumulants, skewness, kurtosis and related tests. R package version 0.13. Available from: http://CRAN.R-project.org/package=moments

Last, P.R., Stevens, J.D. 2009. *Sharks and Rays of Australia*, 2nd Edition. CSIRO, Melbourne, Australia. 656pp.

Lessa, R., Santana, F.M., Hazin, F.H. 2004. Age and growth of the blue shark *Prionace glauca* (Linnaeus, 1758) off northeastern Brazil. *Fisheries Research*, 66: 19-30.

Lo, N., Jacobson, L.D., Squire, J.L. 1992. Indices of relative abundance from fish spotter data based on delta-log normal models. *Canadian Journal of Fisheries and Aquatic Sciences*, 49: 2515–2526.

Ludwig, D., Hilborn, R., Walters C. 1993. Uncertainty, resource exploitation, and conservation: lessons from history. *Science*, 260: 17-36.

Lynch, A.M.J., Sutton, S.G., Simpfendorfer, C.A. 2010. Implications of recreational fishing for elasmobranch conservation in the Great Barrier Reef Marine Park. *Aquatic Conservation: Marine and Freshwater Ecosystems*, 20: 312-318

Mandelman, J.W., Farrington, M.A. 2007. The estimated short-term discard mortality of a trawled elasmobranch, the spiny dogfish (*Squalus acanthias*). *Fisheries Research*, 83: 238-245.

Markaida, U., Sosa-Nishizaki, O. 2010. Food and feeding habits of the blue shark *Prionace glauca* caught off Ensenada, Baja California, Mexico, with a review on its feeding. *Journal of the Marine Biological Association of the United Kingdom*, 90: 977-994.

Martin, T.G., Wintle, B.A., Rhodes, J.R., Kuhnert, P.M., Field, S.A., Low-Choy, S.J., Tyre, A.J., Possingham, H.P. 2005. Zero tolerance ecology: improving ecological inference by modelling the source of zero observations. *Ecology Letters*, 8: 1235-1246.

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

McAllister, M.K., Starr, P.J., Restrepo, V.R., Kirkwood, G.P. 1999. Formulating quantitative methods to evaluate fishery-management systems: what fishery processes should be modelled and what trade-offs should be made? *ICES Journal of Marine Science*, 56: 900-916.

McAllister, M.K., Pikitch, E.K., Babcock, E.A. 2008. Why are Bayesian methods useful for the stock assessment of sharks? *In* Sharks of the Open Ocean: biology, fisheries and conservation (Camhi, M.D., Pikitch, E.K., Babcock, E.A, eds.), pp. 351-368. Blackwell Publishing, Oxford.

McCullagh, P., Nelder, J.A. 1989. *Generalized Linear Models*. 2nd Edition. Monographs on Statistics and Applied Probability 37. Chapman and Hall, London. 511pp.

McCulloch, C.E., Searle, S.R. 2001. *Generalized, Linear, and Mixed Models*. Wiley Series in Probability and Statistics. John Wiley and Sons, New York. 325pp.

McKinnell, S., Seki, M.P. 1998. Shark bycatch in the Japanese high seas squid driftnet fishery in the North Pacific Ocean. *Fisheries Research*, 39: 127-138.

McVean, A.R., Walker, R.C.J., Fanning, E. 2006. The traditional shark fisheries of southwest Madagascar: a study in the Toliara region. *Fisheries Research*, 82: 280-289.

Megalofonou, P., Yannopoulos, C., Damalas, D., DeMetrio, G., Deflorio, M., Serna, J.M., Macias, D. 2005. Incidental catch and estimated discards of pelagic sharks from the swordfish and tuna fisheries in the Mediterranean Sea. *Fishery Bulletin*, 103: 620-634.

Megalofonou, P., Damalas, D., DeMetrio, G. 2009a. Biological characteristics of blue shark, *Prionace glauca*, in the Mediterranean Sea. *Journal of the Marine Biological Association of the United Kingdom*, 89: 1233–1242.

Megalofonou, P., Damalas, D., Deflorio, M., DeMetrio, G. 2009b. Modeling environmental, spatial, temporal, and operational effects on blue shark by-catches in the Mediterranean long-line fishery. *Journal of Applied Ichthyology*, 25 (suppl. 1): 47–55.

Mollet, H.F., Cailliet, G.M. 2002. Comparative population demography of elasmobranchs using life history tables, Leslie matrices and stage-based matrix models. *Marine and Freshwater Research*, 53: 503-516.

Montealegre-Quijano, S., Vooren, C.M. 2010. Distribution and abundance of the life stages of the blue shark *Prionace glauca* in the Southwest Atlantic. *Fisheries Research*, 101: 168-179.

Morgan, A., Burgess, G.H. 2007. At-vessel fishing mortality for six species of sharks caught in the northwest Atlantic and Gulf of Mexico. *Gulf and Caribbean Research*, 19: 123–129.

Morgan, A., Carlson, J.K., 2010. Capture time, size and hooking mortality of bottom longline-caught sharks. *Fisheries Research*, 101: 32-37.

Moyes, C.D., Fragoso, N., Musyl, M.K., Brill, R.W., 2006. Predicting postrelease survival in large pelagic fish. *Transactions of the American Fisheries Society*, 135: 1389–1397.

Murua, H., Coelho, R., Santos, M.N., Arrizabalaga, H., Yokawa, K., Romanov, E., Zhu, J.F., Kim, Z.G., Bach, P., Chavance, P., Delgado de Molina, A., Ruiz, J. 2012. Preliminary ecological risk assessment (ERA) for shark species caught in fisheries managed by the Indian Ocean Tuna Commission (IOTC). *IOTC Technical Document*, IOTC-2012-SC15-INF10. 22pp.

Nagelkerke, N. 1991. A note on a general definition of the coefficient of determination. *Biometrika*, 78: 691-692.

Nakano, H. 1994. Age, reproduction and migration of blue shark in the North Pacific Ocean. *Bulletin of National Research Institute of Far Seas Fisheries*, 31: 141-256.

NASA. 2012. NASA Ocean Color Web. Available from: http://oceancolor.gsfc.nasa.gov/.

NOAA. 2012. NOAA National Climatic Data Center. Available from: http://www.ncdc.noaa.gov/.

O'Brien, R.M. 2007. A caution regarding rules of thumb for variance inflation factors. *Quality and Quantity*, 41: 673-690.

Ortiz, M., Arocha, F. 2004. Alternative error distribution models for standardization of catch rates of non-target species from a pelagic longline fishery: billfish species in the Venezuelan tuna longline fishery. *Fisheries Research*, 70: 275-297.

OSCAR, 2012. NOAA Ocean Surface Current Analysis. Available from: http://www.oscar.noaa.gov/.

Pauly, D., Palomares, M.L. 2005. Fishing down marine food web: it is far more pervasive than we thought. *Bulletin of Marine Science*, 76: 197–211.

Pauly, D., Christensen, V., Dalsgaard, J., Froese, R., Torres, F. 1998. Fishing down marine food webs. *Science*, 279: 860-863.

Pepperell, J.G. 1992. Trends in the distribution, species composition and size of sharks caught by gamefish anglers off South-eastern Australia, 1961-90. *Australian Journal of Marine and Freshwater Research*, 43: 213-225.

Petersen, S.L., Honig, M.B., Ryan, P.G., Underhill, L.G., Compagno, L.J.V. 2009. Pelagic shark bycatch in the tuna- and swordfish-directed longline fishery off southern Africa. *African Journal of Marine Science*, 31: 215-225.

Pratt, H.L. 1979. Reproduction in the blue shark, *Prionace glauca*. *Fishery Bulletin*, 77: 445-470.

Preti, A., Soykan, C.U., Dewar, H., Wells, R. J. D, Spear, N., Kohin, S. 2012. Comparative feeding ecology of shortfin mako, blue and thresher sharks in the California Current. *Environmental Biology of Fishes*, 95: 127-146.

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: 129-145.

Queiroz, N., Lima, F.P., Maia, A., Ribeiro, P.A., Correia, J.P., Santos, A.A.2005. Movement of blue shark, *Prionace glauca*, in the north-east Atlantic based on mark - recapture data. *Journal of the Marine Biological Association of the United Kingdom*, 85: 1107-1112.

Quinn, T.J., Deriso, R.B. 1999. *Quantitative Fish Dynamics*. Oxford University Press, New York. 542pp.

R Development Core Team. 2012. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Available from: http://www.R-project.org/.

Ready, J., Kaschner, K., South, A.B., Eastwood, P.D., Rees, T., Rius, J., Agbayani, E., Kullander, S., Froese, R. 2010. Predicting the distributions of marine organisms at the global scale. *Ecological Modelling*, 221: 467-478.

Reynolds, R.W., Smith, T.M., Liu, C., Chelton, D.B., Casey, K.S., Schlax, M.G. 2007. Daily high-resolution blended analyses for sea surface temperature. *Journal of Climate*, 20: 5473-5496.

Roberts, J.J., Best, B.D., Dunn, D.C., Treml, E.A., Halpin, P.N. 2010. Marine Geospatial Ecology Tools: An integrated framework for ecological geoprocessing with ArcGIS, Python, R, MATLAB, and C++. *Environmental Modeling & Software*, 25: 1197-1207.

Rodriguez-Cabello, C., Fernández, A., Olaso, I., Sánchez, F. 2005. Survival of small-spotted catshark (*Scyliorhinus canicula*) discarded by trawlers in the Cantabrian Sea. *Journal of the Marine Biological Association of the United Kingdom*, 85: 1145-1150.

Royston, P., Altman, D. G. 1994. Regression using fractional polynomials of continuous covariates: Parsimonious parametric modeling (with discussion). *Applied Statistics*, 43: 429-467.

Schindler, D.E., Essington, T.E., Kitchell, J.F., Boggs, C., Hilborn, R. 2002. Sharks and tunas: fisheries impacts on predators with contrasting life histories. *Ecological Applications*, 12: 735-748.

Shono, H. 2008. Application of the Tweedie distribution to zero-catch data in CPUE analysis. *Fisheries Research*, 93: 154–162.

Shotton, R. 1999. Case studies of the management of elasmobranch fisheries. *FAO Fisheries Technical Papers*, 378. FAO, Rome. 490pp.

Silva, C.D., Kerwath, S.E., Wilke, C.G., Meÿer, M., Lamberth, S.J. 2010. First documented southern transatlantic migration of a blue shark *Prionace glauca* tagged off South Africa. *African Journal of Marine Science*, 32: 639–642.

Simpfendorfer, C.A., Hueter, R.E., Bergman, U., Connett, S.M.H. 2002. Results of a fishery-independent survey for pelagic sharks in the Western North Atlantic, 1977-1994. *Fisheries Research*, 55: 175-192.

Skaug, H., Fournier, D., Nielsen, A., Magnusson, A., Bolker, B. 2012. glmmADMB: Generalized Linear Mixed Models Using AD Model Builder. R package version 0.7.2.5/r186. Available from: http://R-Forge.R-project.org/projects/glmmadmb/.

Skomal, G.B., Natanson, L.J. 2003. Age and growth of the blue shark (*Prionace glauca*) in the North Atlantic Ocean. *Fishery Bulletin*, 101: 627–639.

Smith, S.E., Au, D.W., Show, C. 1998. Intrinsic rebound potentials of 26 species of Pacific sharks. *Marine and Freshwater Research*, 49: 663-678.

Smith, W.D., Cailliet, G.M., Cortés, E., 2008. Demography and elasticity of the diamond stingray, Dasyatis dipterura: parameter uncertainty and resilience to fishing pressure. *Marine and Freshwater Research*, 59: 575-586.

Snelson Jr., F.F., Roman, B.L., Burgess, G.H. 2008. The reproductive biology of pelagic elasmobranchs. *In*: *Sharks of the Open Ocean: biology, fisheries and conservation* (Camhi, M.D., Pikitch, E.K., Babcock, E.A, eds.), pp. 24-53. Blackwell Publishing, Oxford.

Stevens, J.D. 1975. Vertebral rings as a means of age determination in the blue shark (*Prionace glauca* L). *Journal of the Marine Biological Association of the United Kingdom*, 55: 657-665.

Stevens, J.D. 1976. Preliminary results of shark tagging in the north-east Atlantic, 1972-1975. *Journal of the Marine Biological Association of the United Kingdom*, 56: 929-937.

Stevens, J.D. 1984. Biological observations on sharks caught by sport fishermen off New South Wales. *Australian Journal of Marine and Freshwater Research*, 35: 573-590.

Stevens, J.D. 1990. Further results from a tagging study of pelagic sharks in the north-east Atlantic. *Journal of the Marine Biological Association of the United Kingdom*, 70: 707-720.

Stevens, J.D. 1992. Blue and mako shark by-catch in the Japanese longline fishery off south-eastern Australia. *Australian Journal of Marine and Freshwater Research*, 43: 227-236.

Stevens, J. 2009. *Prionace glauca. In*: IUCN 2012. IUCN Red List of Threatened Species. Version 2012.1. Available from: www.iucnredlist.org.

Stevens, J.D., Bonfil, R., Dulvy, N.K., Walker, P. A. 2000. The effects of fishing on sharks, rays, and chimaeras (chondrichthyans), and the implications for marine ecosystems. *ICES Journal of Marine Science*, 57: 476-494.

Tavares, R., Ortiz, M., Arocha, F. 2012. Population structure, distribution and relative abundance of the blue shark (*Prionace glauca*) in the Caribbean Sea and adjacent waters of the North Atlantic. *Fisheries Research*, 129-130: 137-152.

Vannuccini, S. 1999. Shark utilization, marketing and trade. *FAO Fisheries Technical Paper*, 389. FAO, Rome. 470pp.

Vaske Jr., T., Lessa, R., Gadig, O.B. 2009. Feeding habits of the blue shark (*Prionace glauca*) off the coast of Brazil. *Biota Neotropica*, 9: 55-60.

Venables, W.N., Ripley, B.D. 2002. *Modern Applied Statistics with S*. 4th Edition. Springer, New York. 495pp.

Vögler, R., Beier, E., Ortega-García, S., Santana-Hernández, H., Valdez-Flores, J.J. 2012. Ecological patterns, distribution and population structure of *Prionace glauca* (Chondrichthyes: Carcharhinidae) in the tropical-subtropical transition zone of the north-eastern Pacific. *Marine Environmental Research*, 73: 37-52.

Walker, T.I. 1998. Can shark resources be harvested sustainably? A question revisited with a review of shark fisheries. *Marine and Freshwater Research*, 49: 553-572.

Walsh, W.A., Kleiber, P. 2001. Generalized additive model and regression tree analyses of blue shark (*Prionace glauca*) catch rates by the Hawaii-based commercial longline fishery. *Fisheries Research*, 53: 115-131.

Wang, Y.G., Carey, V. 2003. Working correlation structure misspecification, estimation and covariate design: Implications for generalized estimating equations performance. *Biometrika*, 90: 29–41.

Ward, P., Lawrence, E., Darbyshire, R., Hindmarsh, S., 2008. Large-scale experiment shows that nylon leaders reduce shark bycatch and benefit pelagic longline fishers. *Fisheries Research*, 90: 100-108.

Warnes, G.R. 2011a. gmodels: Various R programming tools for model fitting. R package version 2.15.1. Available from: http://CRAN.R-project.org/package=gmodels.

Warnes, G.R. 2011b. gplots: Various R programming tools for plotting data. R package version 2.10.1. Available from: http://CRAN.R-project.org/package=gplots.

White, W.T., Last, P.R. 2012. A review of the taxonomy of chondrichthyan fishes: a modern perspective. *Journal of Fish Biology*, 80: 901–917.

Xiao, Y. 1997. Subtleties in, and practical problems with, the use of production models in fish stock assessment. *Fisheries Research*, 33: 17-36.

Yokota, K., Kiyota, M., Minami, H., 2006. Shark catch in a pelagic longline fishery: comparison of circle and tuna hooks. *Fisheries Research*, 81: 337-341.

Zeger, S.L., Liang, K.Y. 1986. Longitudinal data analysis for discrete and continuous outcomes. *Biometrics*, 42: 121-130.

Zeger, S.L., Liang, K.Y., Albert, P.S. 1988. Models for longitudinal data: a generalized estimating equation approach. *Biometrics*, 44: 1049-1060.

Zhang, W., 2012. cplm: Tweedie compound Poisson linear models. R package version 0.5-1. Available from: http://CRAN.R-project.org/package=cplm.

Zhu, J., Dai, X., Xu, L., Chen, X., Chen, Y. 2011. Reproductive biology of female blue shark *Prionace glauca* in the southeastern Pacific Ocean. *Environmental Biology of Fishes*, 91:95-102.

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*. Series on Statistics for Biology and Health. Springer, New York. 529pp.

ANNEX 1: GLOSSARY

In this Annex we present a glossary of some abbreviations that were used in this thesis.

Abbreviation	Term
AIC	Akaike Information Criterion
AUC	Area Under the Curve
CPUE	Catches Per Unit of Effort
ERA	Ecological Risk Assessment
FL	Fork Length
F at MSY	Maximum Fishing Mortality at Maximum Sustainable Yield
GAM	Generalized Additive Models
GEE	Generalized Estimation Equations
GLM	Generalized Linear Models
GLMM	Generalized Linear Mixed Models
GVIF	Generalized Variance Inflation Factors
HYCOM	HYbrid Coordinate Ocean Model
ICCAT	International Commission for the Conservation of Atlantic Tunas
IOTC	Indian Ocean Tuna Commission
NASA	National Aeronautics and Space Administration
NOAA	National Oceanic and Atmospheric Administration
MCMC	Markov Chain Monte Carlo
MGET	Marine Geospatial Ecology Tools
MLD	Mixed Layer Depth
MSE	Mean Squared Error
MSY	Maximum Sustainable Yield
OR	Odds-ratios
PQL	Penalized quasi-Likelihood
RFMO	Regional Fisheries Management Organization
ROC	Receiver Operating Characteristic Curve
SSH	Sea Surface Height
SST	Sea Surface Temperature
TL	Total Length
tRFMO	Tuna-Regional Fisheries Management Organization

ANNEX 2: R-CODE

In this Annex we present examples of the the R-language code that was used for producing this thesis, including data analysis, statistical tests, plots and models. The complete scripts that were used in the thesis are not presented in full, mainly due to the very large size, but rather some examples on how the analysis were carried out are provided. We follow an R-language format where comments and notes to facilitate reading and understanding the code are preceded by the symbol #.

```
######################
##### CHAPTER I #####
######################
```

Plot FAO time series data
```
fis<-read.table("SQServlet_marine_fishes.csv", header=T, sep=",")
fis<-ts(fis[,2], start=c(1950,1), end=c(2010, 1), frequency=1)
elasmo<-read.table("SQServlet_elasmobranchs.csv", header=T, sep=",")
elasmo<-ts(elasmo[,2], start=c(1950,1), end=c(2010, 1), frequency=1)
par(mfrow=c(2,1))
plot(fis/1000, main="Marine fishes", ylab="Catches (thousand ton)")
plot(elasmo/1000, main="Elasmobranchs", ylab="Catches (thousand ton)")
```

Plot ICCAT-Task1 time series data
```
iccat.data<-read.table("Data_ICCAT-Task1.csv", header=T, sep=",")
all.sharks<-ts(iccat.data[,2], start=c(1950,1), end=c(2010, 1), frequency=1)
all.bsh<-ts(iccat.data[,3], start=c(1950,1), end=c(2010, 1), frequency=1)
pt.bsh<-ts(iccat.data[,4], start=c(1950,1), end=c(2010, 1), frequency=1)
par(mfrow=c(3,1))
plot(all.sharks/1000, main="All sharks, all fleets", ylab="Catches (thousand ton)")
plot(all.bsh/1000, main="Blue shark, all fleets", ylab="Catches (thousand ton)")
plot(pt.bsh/1000, main="Blue shark, EU.Portugal fleet", ylab="Catches (thousand ton)")
```

```
#######################
##### CHAPTER II #####
#######################
```

Load dataset
```
mort<-read.table("Mort_BSH.csv", header=T, dec=".", sep=",")
mort$Year<-factor(mort$Year)
mort$Quarter<-factor(mort$Quarter)
summary(mort)
str(mort)
```

Size distribution of the samples

```
fem<- subset(mort, Sex=="F")
mac<- subset(mort, Sex=="M")
summary(fem)
summary(mac)
#Plots
par(mfrow=c(2,1))
with(fem, hist(FL, breaks="scott", main="Females", xlim=c(40,310), ylim=c(0,1400),
xlab="Fork Length (cm)"))
with(mac, hist(FL, breaks="scott", main="Males", xlim=c(40,310), ylim=c(0,1400),
xlab="Fork Length (cm)"))
# Test normality and homogeneity of variances, with and without transformations
lillie.test(mort$FL)
lillie.test(log(mort$FL))
lillie.test(sqrt(mort$FL))
leveneTest(mort$FL~mort$Sex)
leveneTest(log(mort$FL)~mort$Sex)
leveneTest(sqrt(mort$FL)~mort$Sex)
# Hypothesis tests to compare sizes
wilcox.test(FL~Sex, data=mort)
```

Contingency table analysis

```
library(gmodels); citation("gmodels")
# Specimen Sex
table(mort$Sex, mort$Mort)
CrossTable(mort$Sex, mort$Mort, expected=T, chisq=T, format="SPSS", prop.r=F,
prop.c=T,prop.t=F,prop.chisq=F)
# Year
table(mort$Year, mort$Mort)
CrossTable(mort$Year, mort$Mort, expected=T, chisq=T, format="SPSS", prop.r=F,
prop.c=T,prop.t=F,prop.chisq=F)
# Quarter
table(mort$Quarter, mort$Mort)
CrossTable(mort$Quarter, mort$Mort, expected=T, chisq=T, format="SPSS", prop.r=F,
prop.c=T,prop.t=F,prop.chisq=F)
# Gangion material
table(mort$Gangion, mort$Mort)
CrossTable(mort$Gangion, mort$Mort, expected=T, chisq=T, format="SPSS",
prop.r=F, prop.c=T,prop.t=F,prop.chisq=F)
# Specimen size (categorized by the quartiles)
mort$FL_CAT<- cut(mort$FL, breaks=c(40, 188, 205, 221, 305),right=F,
include.lowest=T)
mort$FL_CAT<-factor(mort$FL_CAT)
table(mort$FL_CAT, mort$Mort)
CrossTable(mort$FL_CAT, mort$Mort, expected=T, chisq=T, format="SPSS",
prop.r=F, prop.c=T,prop.t=F,prop.chisq=F)
# Latitude (categorized by the quartiles)
mort$Lat_CAT<- cut(mort$Lat, breaks=c(-34, -24.290, 1.167, 14.370, 21.710),right=F,
include.lowest=T)
mort$Lat_CAT<-factor(mort$Lat_CAT)
```

```
table(mort$Lat_CAT,mort$Mort)
CrossTable(mort$Lat_CAT,mort$Mort, expected=T, chisq=T, format="SPSS",
prop.r=F, prop.c=T,prop.t=F,prop.chisq=F)
# Longitude (categorized by the quartiles)
mort$Long_CAT<- cut(mort$Long, breaks=c(-43.780, -26.650, -22.750, -11.100,
8.733),right=F, include.lowest=T)
mort$Long_CAT<-factor(mort$Long_CAT)
table(mort$Long_CAT,mort$Mort)
CrossTable(mort$Long_CAT,mort$Mort, expected=T, chisq=T, format="SPSS",
prop.r=F, prop.c=T,prop.t=F,prop.chisq=F)
# SST (categorized by the quartiles)
mort$SST_8day_CAT<- cut(mort$SST_8day, breaks=c(18.52, 23.10, 25.8, 26.77,
28.65), right=F, include.lowest=T)
mort$SST_8day_CAT<-factor(mort$SST_8day_CAT)
table(mort$SST_8day_CAT,mort$Mort)
CrossTable(mort$SST_8day_CAT,mort$Mort, expected=T, chisq=T, format="SPSS",
prop.r=F, prop.c=T,prop.t=F, prop.chisq=F)
# Plots
par(mfrow=c(2,4))
plot(mort$Mort~mort$Sex, main="Sex", ylab="", xlab="")
plot(mort$Mort~mort$Year, main="Year", ylab="", xlab="")
plot(mort$Mort~mort$Quarter, main="Quarter", ylab="", xlab="")
plot(mort$Mort~mort$Gangion, main="Branch Line", ylab="", xlab="")
plot(mort$Mort~mort$FL_CAT, main="Size (FL)", ylab="", xlab="")
plot(mort$Mort~mort$Lat_CAT, main="Latitude", ylab="", xlab="")
plot(mort$Mort~mort$Long_CAT, main="Longitude", ylab="", xlab="")
plot(mort$Mort~mort$SST_8day_CAT, main="SST (ºC)", ylab="", xlab="")

# Test univariate binomial GLM models
fit0<-glm(MortBIN~1, family=binomial(link="logit"), data=mort_2)
summary (fit0)
# Specimen size
fit2<-glm(MortBIN~FL, family=binomial(link="logit"), data=mort)
summary (fit2)
anova(fit0, fit2, test="Chisq")
# Specimen sex
fit3<-glm(MortBIN~Sex, family=binomial("logit"), data=mort)
summary (fit3)
anova(fit0, fit3, test="Chisq")
# Latitude
fit4<-glm(MortBIN~Lat, family=binomial("logit"), data=mort)
summary (fit4)
anova(fit0, fit4, test="Chisq")
# Longitude
fit5<-glm(MortBIN~Long, family=binomial("logit"), data=mort)
summary (fit5)
anova(fit0, fit5, test="Chisq")
# Year
fit6<-glm(MortBIN~Year, family=binomial("logit"), data=mort)
summary (fit6)
```

103

```
anova(fit0, fit6, test="Chisq")
# Quarter
fit8<-glm(MortBIN~Quarter, family=binomial("logit"), data=mort)
summary (fit8)
anova(fit0, fit8, test="Chisq")
# Vessel
fit9<-glm(MortBIN~Boat, family=binomial("logit"), data=mort)
summary (fit9)
anova(fit0, fit9, test="Chisq")
# SST
fit10<-glm(MortBIN~SST_8day, family=binomial("logit"), data=mort)
summary (fit10)
anova(fit0, fit10, test="Chisq")
# Gangion
fit12<-glm(MortBIN~Gangion, family=binomial("logit"), data=mort)
summary (fit12)
anova(fit0, fit12, test="Chisq")
```

Multivariate simple effects model
```
fit23_f3<-glm(MortBIN~FL+Lat+Long+Year+Season+Gangion+Sex,
family=binomial(link="logit"), data=mort)
summary (fit23_f3)
anova(fit23_f3, test="Chisq")
```

Assumption of sufficient binomial contrasts in the categorical variables
```
table(mort$Year, mort$MortBIN)
table(mort$Quarter, mort$MortBIN)
table(mort$Boat, mort$MortBIN)
table(mort$Gangion, mort$MortBIN)
table(mort$Sex, mort$MortBIN)
```

Assumption of linearity in the continuous variables
```
# Only 1 example is provided, but all continuous explanatory variables were tested
# Quartiles method
mort$FLCAT<- cut(mort$FL, breaks=c(40, 181, 201, 219, 305),right=F,
include.lowest=T)
table(mort$FLCAT)
fit20<-glm(MortBIN~FLCAT+Sex+Lat+Long+Year+Quarter+Boat+Gangion,
family=binomial("logit"), data=mort)
summary (fit20)
x<-c((40+181)/2, (181+201)/2, (201+219)/2, (219+305)/2)
y<-c(0, as.numeric(fit20$coef[2]),as.numeric(fit20$coef[3]), as.numeric(fit20$coef[4]))
plot(x, y,pch=19, main="Linearidade de idade com logit")
lines(lowess(x,y))
# Fractional polynomials
library(mfp)
mfp(MortBIN~fp(FL)+Sex+Lat+Long+Year+Quarter+Boat+Gangion,
family=binomial("logit"), data=mort)
# GAM plots
library(gam)
```

```
plot(gam(MortBIN~s(FL)+Sex+Lat+Long+Year+Quarter+Boat+Gangion,
family=binomial("logit"), data=mort), se=T, xlab="Size (FL, cm)")
```

Testing for significant interactions
```
# Only 1 example is provided, but all possible pairs were tested
fit101<-glm(MortBIN~FL1+Lat1+Long1+Year+Season+Gangion+Sex+FL1:Lat1,
family=binomial("logit"), data=mort_2)
summary (fit101)
anova (fit23_f3, fit101, test="Chisq")
```

Interaction plots
```
#Only 1 example is provided but all significant interactions were plotted
# Divide FL by the 10% percentiles
mort$FLCAT2<- cut(mort$FL, breaks=quantile(mort$FL, probs = seq(0, 1, 0.1)),
right=F, include.lowest=T)
# Divide Longitude by the 10% percentiles
mort$LongCAT2<- cut(mort$Long, breaks=quantile(mort$Long, probs = seq(0, 1,
0.1)),right=F, include.lowest=T)
#FL:Longitude interactions plots
par(mfrow=c(1,2))
interaction.plot(x.factor=mort$LongCAT2, trace.factor=mort$FLCAT2,
response=mort$MortBIN, fun=mean,  type = "l", legend=T, ylab="Mooking mortality",
col=rainbow(10), lwd=2, xlab="Longitude class", leg.bty = "o")
interaction.plot(x.factor=mort$FLCAT2, trace.factor=mort$LongCAT2,
response=mort$MortBIN, fun=mean, type = "l", legend=T, ylab="Mooking mortality",
col=rainbow(4), lwd=2, xlab="FL size class (cm)", leg.bty = "o")
```

Final model with transformations and interactions
```
fit205 <- glm(MortBIN ~ FL1 + Lat1 + Long1 + Year + Season + Gangion + Sex +
FL1:Long1 + FL1:Year + Long1:Season + Long1:Sex + Year:Gangion + Season:Sex,
family = binomial(link = "logit"), data = mort)
summary(fit205)
anova(fit205, test="Chisq")
```

R2 of the models
```
# R2 for the simple effects model
R2N_2 <- (1-exp((fit23_f3$dev - fit23_f3$null) / dim(mort_2)[1]))/(1-exp(-
fit23_f3$null / dim(mort_2)[1]))
# R2 for the model with interactions
R2N.multi.inter<-(1-exp((fit205$dev-fit205$null)/ dim(mort_2)[1]))/(1-exp(-
fit205$null/ dim(mort_2)[1]))
R2N.multi.inter #  [1] 0.164795
```

ROC Curves
```
library(Epi); citation("Epi")
# Simple effects model
ROC(form = MortBIN~FL1+Lat1+Long1+Year+Season+Gangion+Sex, data=mort,
plot="ROC", PV=T, MX=T, MI=F, AUC=T)
# Model with interactions
```

```
ROC(form = MortBIN ~ FL1 + Lat1 + Long1 + Year + Season +  Gangion + Sex +
FL1:Long1 + FL1:Year + Long1:Season + Long1:Sex + Year:Gangion + Season:Sex,
data=mort, plot="ROC", PV=T, MX=T, MI=F, AUC=T, cex.lab=1.1, cex.axis=1.1)
```

Hosmer and Lemeshow test

```
hosmerlem <- function(y, yhat, g=10) {
  cutyhat = cut(yhat,
  breaks = quantile(yhat, probs=seq(0, 1, 1/g)), include.lowest=T)
  obs = xtabs(cbind(1 - y, y) ~ cutyhat)
  expect = xtabs(cbind(1 - yhat, yhat) ~ cutyhat)
  chisq = sum((obs - expect)^2/expect)
  P = 1 - pchisq(chisq, g - 2)
   return(list(chisq=chisq,p.value=P))
}
# Simple effects model
hosmerlem(mort$MortBIN, fitted(fit23_f3))
# Model with interactions
hosmerlem(mort$MortBIN, fitted(fit205))
```

Residual analysis

```
#Residuals along the predicted values
plot (predict(fit205), residuals (fit205, type="deviance"))
plot (predict(fit205), residuals (fit205, type="pearson"))
# Cooks distances
plot(cooks.distance(fit205))
#Half-normal plots
halfnorm(cooks.distance(fit205), main="Cooks Distance")
halfnorm(residuals (fit205, type="pearson"), main="Pearson Residuals")
halfnorm(residuals (fit205, type="deviance"), main="Deviance Residuals")
# DF betas along the predicted values
rdf <- dfbetas(fit205)
head(rdf)
plot(predict(fit205), rdf[,2], main="Size", xlab="Predicted", ylab="Df Betas")
plot(predict(fit205), rdf[,3], main="Lat", xlab="Predicted", ylab="Df Betas")
plot(predict(fit205), rdf[,4], main="Long", xlab="Predicted", ylab="Df Betas")
plot(predict(fit205), rdf[,5], main="Year2009", xlab="Predicted", ylab="Df Betas")
plot(predict(fit205), rdf[,6], main="Year2010", xlab="Predicted", ylab="Df Betas")
plot(predict(fit205), rdf[,7], main="Year2011", xlab="Predicted", ylab="Df Betas")
plot(predict(fit205), rdf[,8], main="Quarter", xlab="Predicted", ylab="Df Betas")
plot(predict(fit205), rdf[,9], main="Branch line", xlab="Predicted", ylab="Df Betas")
plot(predict(fit205), rdf[,10], main="Sex", xlab="Predicted", ylab="Df Betas")
plot(predict(fit205), rdf[,11], main="FL:Long", xlab="Predicted", ylab="Df Betas")
plot(predict(fit205), rdf[,12], main="FL:Year2009", xlab="Predicted", ylab="Df Betas")
plot(predict(fit205), rdf[,13], main="FL:Year2010", xlab="Predicted", ylab="Df Betas")
plot(predict(fit205), rdf[,14], main="FL:Year2011", xlab="Predicted", ylab="Df Betas")
plot(predict(fit205), rdf[,15], main="Long:Quarter", xlab="Predicted", ylab="Df
Betas")
plot(predict(fit205), rdf[,16], main="Long:Sex", xlab="Predicted", ylab="Df Betas")
plot(predict(fit205), rdf[,17], main="Year2009:Branch", xlab="Predicted", ylab="Df
Betas")
```

```
plot(predict(fit205), rdf[,18], main="Year2010:Branch", xlab="Predicted", ylab="Df
Betas")
plot(predict(fit205), rdf[,19], main="Year2011:Branch", xlab="Predicted", ylab="Df
Betas")
plot(predict(fit205), rdf[,20], main="Quarter:Sex", xlab="Predicted", ylab="Df Betas")
```

Cross-validation
```
library(boot); citation("boot")
cv.10.err <- cv.glm(mort, fit205, cost, K = 10)
cv.10.err$delta
```

Calculate the odds-ratios
```
# Only one example is provided (specimen size), but calculated for all variables
exp(fit205$coef[2])
mc<-summary(fit205)$cov.scaled
exp(c(fit205$coef[2]-qnorm(0.975) * sqrt(mc[2,2]), fit205$coef[2] + qnorm(0.975) *
sqrt(mc[2,2])))
# Increase of 10cm in size
a<-10
exp(a* fit205$coef[2]); (1-exp(a* fit205$coef[2]))*100
exp(a* fit205$coef[2]-qnorm(0.975)*a*sqrt(mc[2,2]))
exp(a* fit205$coef[2]+qnorm(0.975)*a*sqrt(mc[2,2]))
```

Fit GEE models
```
library(geepack); citation("geepack")
# Simple effects GEE
gee4<-geeglm(MortBIN ~ FL1 + Lat1 + Long1 + Year + Season + Gangion + Sex,
id=Setcode, family=binomial("logit"), corstr="exchangeable", data=mort)
summary (gee4)
anova(gee4, test="Chisq")
# GEE with significant interactions
gee5 <- geeglm(MortBIN ~ FL1 + Lat1 + Long1 + Year + Season + Gangion + Sex +
FL1:Long1 + FL1:Year + Long1:Season + Long1:Sex + Year:Gangion + Season:Sex,
id=Setcode, family=binomial("logit"), corstr="exchangeable", data=mort)
summary (gee5)
anova(gee5, test="Chisq")
```

Plot inverse-logits
```
# Back-transform the FL variable
fl.t <- I((seq(40, 305, 5)/100)^-0.5)+log((seq(40, 305, 5)/100))
se.fl1 <- sqrt(mc[2,2] + mc[9,9] + mc[2,9])
# cycle to calculate the inverse logits for all size ranges
library(faraway)
logit.fl <- rep(0, times=length(fl.t))
ilogit.fl1 <- rep(0, times=length(fl.t))
ilogit.icl.fl1 <- rep(0, times=length(fl.t))
ilogit.icu.fl1 <- rep(0, times=length(fl.t))
for (i in 1: length(fl.t)){
    logit.fl[i] <- fit205$coef[1] + fit205$coef[2]*fl.t[i]
    ilogit.fl1[i] <- ilogit(logit.fl[i]) # point estimates
```

```
    ilogit.icl.fl1[i] <- ilogit(logit.fl[i] - 1.96 * se.fl1) # lower 95% CI
    ilogit.icu.fl1[i] <- ilogit(logit.fl[i] + 1.96 * se.fl1) # upper 95% CI
}
# Plots
plot(ilogit.fl1, type ="l", ylim=c(0,1))
lines(ilogit.icl.fl1)
lines(ilogit.icu.fl1)
```

Plot odds-ratios
```
# Back-transform the FL variable
fl.t.odds <-  I((seq(40, 305, by=10)/100)^-0.5)+log((seq(40, 305, by=10)/100))
# Cycle to calculate the odds-ratios for 10cm size increments
fl.odds <- rep(0, times=length(fl.t.odds))
fl.odds.up <- rep(0, times=length(fl.t.odds))
fl.odds.lo <- rep(0, times=length(fl.t.odds))
for (i in 1:length(fl.t.odds)){
    fl.odds[i] <- exp(fit205$coef[2]*(fl.t.odds[i+1]-fl.t.odds[i]))
    fl.odds.up[i] <- exp(fit205$coef[2]*(fl.t.odds[i+1]-fl.t.odds[i]) - qnorm(0.975) *
(fl.t.odds[i+1]-fl.t.odds[i])* sqrt(mc[2,2]))
    fl.odds.lo[i] <- exp(fit205$coef[2]*(fl.t.odds[i+1]-fl.t.odds[i]) + qnorm(0.975) *
(fl.t.odds[i+1]-fl.t.odds[i])* sqrt(mc[2,2]))
}
#Plots
library (gplots)
plotCI(y=fl.odds, x=Sizes2, uiw=(fl.odds-fl.odds.up), liw=(fl.odds.lo-fl.odds), err="y",
ylim=c(0.65, 1.05), ylab="Odds-Ratios (increse of 10cm in FL)", xlab="Specimen size
(FL, cm)", main="Odds-Ratios")
abline(h=1, lty=2)
```

```
#########################
##### CHAPTER III #####
#########################
```

Load dataset
```
cpue<-read.table("BD_South.csv", header=T, dec=".", sep=",", na.strings = "-9999")
summary(cpue)
dim(cpue)
str(cpue)
cpue$Year<-factor(cpue$Year)
cpue$Season<-factor(cpue$Season)
```

Plot distribution of the response variable
```
hist(cpue$cpueBSH,col="gray87", freq=F, breaks="Scott", main="Blue shark CPUE
distribution", xlab="CPUE (n/1000 hooks)")
lines(density(cpue$cpueBSH))
qqnorm(cpue$cpueBSH)
qqline(cpue$cpueBSH)
```

Correlation plots

```
# Diagonals with the histograms
panel.hist2 <- function(x) {
  par(new = T)
  hist(x, col = "light gray", probability = T, axes=F, main = "", breaks = "scott")
  lines(density(x, na.rm=T), col = "black", lwd = 1)
  rug(x)
}
# Panels with the Spearman correlations
panel.cor2 <- function(x, y, digits = 2, prefix = "", use = "pairwise.complete.obs",
cex.cor) {
  usr <- par("usr")
  on.exit(par(usr))
  par(usr = c(0, 1, 0, 1))
  r <- cor(x, y, use = use, method="spearman")
  txt <- format(c(r, 0.123456789), digits = digits)[1]
  txt <- paste(prefix, txt, sep = "")
  if (missing(cex.cor))
  cex <- 0.8/strwidth(txt)
  test <- cor.test(x, y, method="spearman")
  Signif <- symnum(test$p.value, corr = F, na = F,
  cutpoints = c(0, 0.01, 0.05, 0.1, 1), symbols = c("***", "**", "*", " "))
  text(0.5, 0.5, txt, cex = cex * log(r))
  text(0.8, 0.8, Signif, cex = cex, col = 2)
}
# Final plot
pairs(~cpueBSH +Lat+Long+CurrentVel+MoonIlu+Wind+MLD+SSH+ Salininy+
Chlorophyll+SST, data=bsh_corr, row1attop=F, lower.panel=panel.smooth,
upper.panel=panel.cor2, diag.panel=panel.hist2)
```

Boxplots for categorical explanatory variables
```
par(mfrow=c(1,3))
boxplot(cpueBSH ~  Year, data=cpue, ylab="CPUE (n/1000hks)", main="Year",
col="grey85")
boxplot(cpueBSH ~  Season, data= cpue, ylab="CPUE (n/1000hks)", main="Season",
col="grey85")
boxplot(cpueBSH ~  Gangion, data= cpue, ylab="CPUE (n/1000hks)", main="Gangion
material", col="grey85")
```

Hypothesis tests for categorical explanatory variables
```
# Test normality with and without transformations
library(nortest); citation("nortest")
lillie.test(cpue$cpueBSH)
lillie.test(log(cpue$cpueBSH+1))
lillie.test(1/(cpue$cpueBSH+1))
# Test homogeneity of variances with and without transformations
library(car); citation("car")
leveneTest(cpue$cpueBSH~cpue$Year)
leveneTest(log(cpue$cpueBSH+1)~cpue$Year)
leveneTest((1/(cpue$cpueBSH+1))~cpue$Year)
leveneTest(cpue$cpueBSH~cpue$Season)
```

```
leveneTest(log(cpue$cpueBSH+1)~cpue$Season)
leveneTest((1/(cpue$cpueBSH+1))~cpue$Season)
leveneTest(cpue$cpueBSH~cpue$Gangion)
leveneTest(log(cpue$cpueBSH+1)~cpue$Gangion)
leveneTest((1/(cpue$cpueBSH+1))~cpue$Gangion)
#Non-parametric hypothesis tests
kruskal.test(cpueBSH ~ Year, data=cpue)
kruskal.test(cpueBSH ~ Season, data=cpue)
wilcox.test(cpueBSH ~ Gangion, data=cpue)
```

Calculate % sets with 0 and positive catches
```
cpue$PositiveSet <- ifelse (cpue$nBSH>=1, 1, 0)
table(cpue$PositiveSet)
```

Test univariate models
```
cpue$cpue.add1 <- cpue$cpueBSH+1
# Null model
fit.gamma.uni0 <- glm(cpue.add1 ~  1, family=Gamma(link=log), data=cpue)
summary(fit.gamma.uni0)
# Gangion
fit.gamma.uni1 <- glm(cpue.add1 ~  Gangion, family=Gamma(link=log), data=cpue)
summary(fit.gamma.uni1)
anova(fit.gamma.uni0, fit.gamma.uni1, test="Chisq")
# Season
fit.gamma.uni2 <- glm(cpue.add1 ~  Season, family=Gamma(link=log), data=cpue)
summary(fit.gamma.uni2)
anova(fit.gamma.uni0, fit.gamma.uni2, test="Chisq")
# Year
fit.gamma.uni3 <- glm(cpue.add1 ~  Year, family=Gamma(link=log), data=cpue)
summary(fit.gamma.uni3)
anova(fit.gamma.uni0, fit.gamma.uni3, test="Chisq")
# Latitude
fit.gamma.uni4 <- glm(cpue.add1 ~  Latitude, family=Gamma(link=log), data=cpue)
summary(fit.gamma.uni4)
anova(fit.gamma.uni0, fit.gamma.uni4, test="Chisq")
# Longitude
fit.gamma.uni5 <- glm(cpue.add1 ~  Longitude, family=Gamma(link=log), data=cpue)
summary(fit.gamma.uni5)
anova(fit.gamma.uni0, fit.gamma.uni5, test="Chisq")
# Current velocity
fit.gamma.uni6 <- glm(cpue.add1 ~  OSCAR_mag, family=Gamma(link=log),
data=cpue)
summary(fit.gamma.uni6) #Not sig.
anova(fit.gamma.uni0, fit.gamma.uni6, test="Chisq")
# Moon Illumination
fit.gamma.uni7 <- glm(cpue.add1 ~  moon_ilu, family=Gamma(link=log), data=cpue)
summary(fit.gamma.uni7) #Not sig.
anova(fit.gamma.uni0, fit.gamma.uni7, test="Chisq")
# Wind velocity
```

```
fit.gamma.uni8 <- glm(cpue.add1 ~  AVISOWind, family=Gamma(link=log),
data=cpue)
summary(fit.gamma.uni8)
anova(fit.gamma.uni0, fit.gamma.uni8, test="Chisq")
# MLD
fit.gamma.uni9 <- glm(cpue.add1 ~  HYCOMmld, family=Gamma(link=log),
data=cpue)
summary(fit.gamma.uni9)
anova(fit.gamma.uni0, fit.gamma.uni9, test="Chisq")
# SSH
fit.gamma.uni10 <- glm(cpue.add1 ~  HYCOMssh, family=Gamma(link=log),
data=cpue)
summary(fit.gamma.uni10)
anova(fit.gamma.uni0, fit.gamma.uni10, test="Chisq")
# Salinity
fit.gamma.uni11 <- glm(cpue.add1 ~  HYCOMsali0, family=Gamma(link=log),
data=cpue)
summary(fit.gamma.uni11)
anova(fit.gamma.uni0, fit.gamma.uni11, test="Chisq")
# Chlorophyll
fit.gamma.uni12 <- glm(cpue.add1 ~  AquaChlo, family=Gamma(link=log), data=cpue)
summary(fit.gamma.uni12)
anova(fit.gamma.uni0, fit.gamma.uni12, test="Chisq")
# SST
fit.gamma.uni13 <- glm(cpue.add1 ~  NCDCsst_l4, family=Gamma(link=log),
data=cpue)
summary(fit.gamma.uni13)
anova(fit.gamma.uni0, fit.gamma.uni13, test="Chisq")
```

Multivariate simple effects model
```
fit.gamma3 <- glm(cpue.add1 ~ Gangion + Season + Year + Latitude.t + Longitude.t +
HYCOMsali0.t + AquaChlo.t + NCDCsst_l4.t, family=Gamma(link=log), data=cpue)
```

Assumption of linearity in the continuous variables
```
# Only 1 example is provided, but all continuous explanatory variables were tested
# Quartiles method
cpue$LatitudeCAT <- cut(cpue$Latitude, breaks=c(-34.0000, -27.3100 ,-18.4200,
0.8333, 4.9830 ),right=F, include.lowest=T)
table(cpue$LatitudeCAT)
fit.gamma3.1<-glm(cpue.add1 ~  LatitudeCAT+ Vessel + Gangion + Season + Year +
Longitude + HYCOMsali0+ AquaChlo + NCDCsst_l4, family=Gamma(link=log),
data=cpue)
summary (fit.gamma3.1)
x<-c((-34.0000-27.3100)/2, (-27.3100-18.4200)/2, (-18.4200+0.8333)/2,
(0.8333+4.9830)/2)
y<-c(0, as.numeric(fit.gamma3.1$coef[2]),as.numeric(fit.gamma3.1$coef[3]),
as.numeric(fit.gamma3.1$coef[4]))
plot(x, y, pch=19, main="Linearidade de latitude")
lines(lowess(x,y))
# Fractional polynomials
```

```
library(mfp); citation("mfp")
mfp(cpue.add1 ~ Vessel + Gangion + Season + Year + fp(Latitude) + Longitude +
HYCOMsali0+ AquaChlo + NCDCsst_l4, family=Gamma(link=log), data=cpue)
# GAM plot
library(gam); citation("gam")
plot(gam(cpue.add1~ Vessel + Gangion + Season + Year + s(Latitude) + Longitude +
HYCOMsali0+ AquaChlo + NCDCsst_l4, family=Gamma(link=log), data=cpue), se=T,
xlab="Latitude")
```

Test significance of interactions
```
# Only 1 example is provided, but all possible pairs were tested
fit.gamma12 <- glm(cpue.add1 ~ Gangion + Season + Year + Latitude.t + Longitude.t +
HYCOMsali0.t + AquaChlo.t + NCDCsst_l4.t + Gangion:Season,
family=Gamma(link=log), data=cpue)
summary(fit.gamma12)
anova(fit.gamma3, fit.gamma12)
```

Final Gamma GLM model
```
#Model
fit.gamma55 <- glm(cpue.add1 ~ Gangion + Season + Year + Latitude.t + Longitude.t
+ AquaChlo.t + NCDCsst_l4.t + Gangion:Longitude.t + Gangion:AquaChlo.t,
family=Gamma(link=log), data=cpue)
# Summary
summary(fit.gamma55)
# Significance and goodness-of-fit
anova(fit.gamma55)
R2.fit.gamma55 <- 1-(fit.gamma55$deviance / fit.gamma55$df.residual) /
(fit.gamma55$null.deviance / fit.gamma55$df.null)
AIC(fit.gamma55)
# Evaluate possible collinearity
library(car); citation("car")
vif(fit.gamma55)
# Residual analysis
res.fit.gamma55 <- residuals(fit.gamma55,type="deviance")
plot(residuals(fit.gamma55) ~ log(fitted(fit.gamma55)),ylab="Resíduos Deviance",
xlab=expression(log(hat(mu))))
abline(h=0)
plot(fit.gamma55, which=c(1,2), cex.main=1.5, cex.lab=1.5)
hist(res.fit.gamma55, breaks="Scott",main="Histogram", col="gray88",
cex.main=1.5,cex.lab=1.5, xlab="Deviance residuals")
```

Poisson GLM model
```
#Model
fit.poi55 <- glm(nBSH ~ Gangion + Season + Year + Latitude.t + Longitude.t +
AquaChlo.t + NCDCsst_l4.t + Gangion:Longitude.t + Gangion:AquaChlo.t +
offset(log(Hookn)), family=poisson(link=log), data=cpue)
# Summary
summary(fit.poi55)
# Poisson model with dispersion parameter
pd <-sum(residuals(fit.poi55,type="pearson")^2/fit.poi55$df.res)
```

```
summary (fit.poi55, dispersion=pd)
# Significance and goodness-of-fit
anova(fit.poi55)
R2.fit.poi55 <- 1-(fit.poi55$deviance / fit.poi55$df.residual) /
(fit.poi55$null.deviance/fit.poi55$df.null)
AIC(fit.poi55)
# Residual analysis
res.fit.poi55 <- residuals(fit.poi55,type="deviance")
plot(residuals(fit.poi55) ~ log(fitted(fit.poi55),ylab="Resíduos Deviance",
xlab=expression(log(hat(mu))))
abline(h=0)
plot(fit.poi55, which=c(1,2), cex.main=1.5, cex.lab=1.5)
hist(res.fit.poi55, breaks="Scott", main="Histogram", col="gray88",
cex.main=1.5,cex.lab=1.5, xlab="Deviance residuals")
```

Negative Binomial GLM model

```
library(MASS); citation("MASS")
fit.nb55 <- glm.nb(nBSH ~ Gangion + Season + Year + Latitude.t + Longitude.t +
AquaChlo.t + NCDCsst_l4.t + Gangion:Longitude.t + Gangion:AquaChlo.t +
offset(log(Hookn)), link=log, data=cpue)
# Summary
summary(fit.nb55)
# Significance and goodness-of-fit
anova(fit.nb55)
R2.fit.nb55 <- 1-(fit.nb55$deviance / fit.nb55$df.residual) /
(fit.nb55$null.deviance/fit.nb55$df.null)
AIC(fit.nb55)
# Residual analysis
res.fit.nb55 <- residuals(fit.nb55,type="deviance")
plot(residuals(fit.nb55) ~ log(fitted(fit.nb55),ylab="Resíduos Deviance",
xlab=expression(log(hat(mu))))
abline(h=0)
plot(fit.nb55, which=c(1,2), cex.main=1.5, cex.lab=1.5)
hist(res.fit.nb55, breaks="Scott", main="Histogram", col="gray88",
cex.main=1.5,cex.lab=1.5, xlab="Deviance residuals")
```

Quasi-Poisson GLM Model

```
# Model
fit.qpoi55 <- glm(nBSH ~ Gangion + Season + Year + Latitude.t + Longitude.t +
AquaChlo.t + NCDCsst_l4.t + Gangion:Longitude.t + Gangion:AquaChlo.t +
offset(log(Hookn)), family=quasipoisson(log), data=cpue)
# Summary
summary(fit.qpoi55)
# Significance and goodness-of-fit
anova(fit.qpoi55)
R2.fit.qpoi55 <- 1-(fit.qpoi55$deviance / fit.qpoi55$df.residual) /
(fit.qpoi55$null.deviance/fit.qpoi55$df.null)
```

Tweedie GLM Model

```
library(tweedie); citation("tweedie")
```

```
library(statmod); citation ("statmod")
# Fit the tweedie distribution
out <- tweedie.profile(cpueBSH ~ Gangion + Season + Year + Latitude.t + Longitude.t
+ AquaChlo.t + NCDCsst_l4.t + Gangion:Longitude.t + Gangion:AquaChlo.t,
data=cpue, p.vec=seq(1.1, 1.9, length=9), method="interpolation", do.ci=T,
do.smooth=T, do.plot=T, phi.method="saddlepoint")
p <- out$p.max
# Model
fit.tweedie55 <- glm(cpueBSH ~ Gangion + Season + Year + Latitude.t + Longitude.t
+ AquaChlo.t + NCDCsst_l4.t + Gangion:Longitude.t + Gangion:AquaChlo.t,
family=tweedie(var.power=p, link.power=0), data=cpue)
#Summary
summary(fit.tweedie55)
# Significance and goodness-of-fit
anova(fit.tweedie55, test="Chisq")
par(mfrow=c(2,2));plot(fit.tweedie55)
R2.fit.tweedie55 <- 1-(fit.tweedie55$deviance / fit.tweedie55$df.residual) /
(fit.tweedie55$null.deviance/fit.tweedie55$df.null)
AICtweedie(fit.tweedie55)
# Residual analysis
quantile.res.tweedie55 <- qres.tweedie(fit.tweedie55)
plot(log(fits.tweedie55), qres.tweedie(fit.tweedie55), main="Quantile residuals")
abline(0,0, col="red")
qqnorm(qres.tweedie(fit.tweedie55), main = "QQ Plot", xlab="Standard Normal
Quantiles", ylab="Quantile Residuals")
qqline(quantile.res.tweedie55)
hist(quantile.res.tweedie55, breaks="Scott",main="Histogram", col="gray87",
cex.main=1.5,cex.lab=1.5, xlab="Quantile residuals")
```

Cross-validation
```
library(boot); citation("boot")
k <- 10
cv.gamma55.10 <- cv.glm(cpue, fit.gamma55, K=k)
cv.gamma55.10$delta
cv.poi55.10 <- cv.glm(cpue, fit.poi55, K=k)
cv.poi55.10$delta
cv.poi55.10 <- cv.glm(cpue, fit.qpoi55, K=k)
cv.poi55.10$delta
cv.nb55.10 <- cv.glm(cpue, fit.nb55, K=k)
cv.nb55.10$delta
cv.tweedie55.10 <- cv.glm(cpue, fit.tweedie55, K=k)
cv.tweedie55.10$delta
```

GLMM models with penalized quasi-likelihoods
```
library(MASS); citation("MASS")
# Gamma
fit.mix.gamma100 <- glmmPQL(cpue.add1 ~ Gangion + Season + Year + Latitude.t +
Longitude.t + AquaChlo.t + NCDCsst_l4.t + Gangion:Longitude.t +
Gangion:AquaChlo.t, random = ~ 1| Vessel, family=Gamma(link=log), data=cpue)
summary(fit.mix.gamma100)
```

```
# Poisson
fit.mix.poi100 <- glmmPQL(nBSH ~  Gangion + Season + Year + Latitude.t +
Longitude.t + AquaChlo.t + NCDCsst_l4.t + Gangion:Longitude.t +
Gangion:AquaChlo.t + offset(log(Hookn)), random = ~ 1| Vessel,
family=poisson(link=log), data=cpue)
summary(fit.mix.poi100)
#Quasi-Poisson
fit.mix.qpoi100 <- glmmPQL(nBSH ~  Gangion + Season + Year + Latitude.t +
Longitude.t + AquaChlo.t + NCDCsst_l4.t + Gangion:Longitude.t +
Gangion:AquaChlo.t + offset(log(Hookn)),  random = ~ 1| Vessel,
family=quasipoisson(link=log), data=cpue)
summary(fit.mix.qpoi100)
# Negative Binomial
fit.mix.nb100 <- glmmPQL(nBSH ~  Gangion + Season + Year + Latitude.t +
Longitude.t + AquaChlo.t + NCDCsst_l4.t + Gangion:Longitude.t +
Gangion:AquaChlo.t + offset(log(Hookn)), random = ~ 1| Vessel,
family=negative.binomial(theta=4.510, link="log"), data=cpue)
summary(fit.mix.nb100)
# Tweedie
fit.mix.tweedie100 <- glmmPQL(cpueBSH ~  Gangion + Season + Year + Latitude.t +
Longitude.t + AquaChlo.t + NCDCsst_l4.t + Gangion:Longitude.t +
Gangion:AquaChlo.t, random = ~ 1| Vessel, family=tweedie(var.power=p,
link.power=0), data=cpue)
summary(fit.mix.tweedie100)
#Residuals plots
par(mfrow=c(2,2))
plot(fit.mix.gamma100, main="Gamma GLMM")
plot(fit.mix.poi100, main="Poisson GLMM")
plot(fit.mix.nb100, main="Negative Binomial GLMM")
plot(fit.mix.tweedie100, main="Tweedie GLMM")
```

GLMM models with Laplace approximations

```
library(glmmADMB); citation("glmmADMB")
# Gamma
fit.mix.gamma200 <- glmmadmb(cpue.add1 ~  Gangion + Season + Year + Latitude.t +
Longitude.t + AquaChlo.t + NCDCsst_l4.t + Gangion:Longitude.t +
Gangion:AquaChlo.t, random= ~1|Vessel, family="gamma", link="log", data=cpue,
verbose=F)
summary(fit.mix.gamma200)
AIC(fit.mix.gamma200)
#Poisson
fit.mix.poisson200 <- glmmadmb(nBSH ~  Gangion + Season + Year + Latitude.t +
Longitude.t + AquaChlo.t + NCDCsst_l4.t + Gangion:Longitude.t +
Gangion:AquaChlo.t + offset(log(Hookn)),  random= ~1|Vessel, family="Poisson",
link="log", data=cpue)
summary(fit.mix.poisson200)
AIC(fit.mix.poisson200)
# Negative Binomial
fit.mix.nb200 <- glmmadmb(nBSH ~  Gangion + Season + Year + Latitude.t +
Longitude.t + AquaChlo.t + NCDCsst_l4.t + Gangion:Longitude.t +
```

```
Gangion:AquaChlo.t + offset(log(Hookn)), random= ~1|Vessel, family="nbinom",
link="log", data=cpue)
summary(fit.mix.nb200)
AIC(fit.mix.nb200)
# Tweedie
library(cplm); citation("cplm")
fit.mix.tweedie200 <- cpglmm(cpueBSH ~ Gangion + Season + Year + Latitude.t +
Longitude.t + AquaChlo.t + NCDCsst_l4.t + Gangion:Longitude.t +
Gangion:AquaChlo.t+ (1|Vessel), link = "log", data=cpue)
summary(fit.mix.tweedie200)
AIC(fit.mix.tweedie200)
```

Plots to compare candidate models coefficients

```
library(coefplot2); citation("coefplot2")
vn <- c("GangionWire", "Season2" , "Season3" , "Season4" , "Year2009" , "Year2010" ,
"Year2011" , "Latitude.t", "Longitude.t", "Chlorophyll.t", "SST.t",
"GangionWire:Longitude.t" , "GangionWire:Chlorophyll.t")
coefplot2(list(GLM.Gamma=fit.gamma55, GLM.NB=fit.nb55,
GLM.quasiPoisson=fit.qpoi55, GLM.Tweedie=fit.tweedie55,
GLMM.Gamma=fit.mix.gamma200, GLMM.NB=fit.mix.nb200,
GLMM.Tweedie=fit.mix.tweedie200), varnames=vn, CI=2, legend=T,
legend.x="bottomleft", legend.args=c(ncol=1, cex=1))
```

Example of model predictions

```
#Back-transform the continuous variables
# Latitude (1Q:-27.5100): South Atlantic
log(((-27.5100+34.1)/10)) # [1] -0.4170317
# Latitude (3ºQ: 0.4417): Equatorial region
log(((0.4417+34.1)/10)) # [1] 1.239582
# Longitude (1ºQ: -25.620): Western Atlantic
I(((-25.72+43.8)/10)^1) #[1] 1.808
# Longitude (3ºQ: -9.167): Eastern Atlantic
I(((-11.59+43.8)/10)^1) #[1] [1] 3.221
# Clorofill  (Median: 0.06904): Median value
I((0.06904/0.1)^1) #[1] 0.6904
# SST  (Median: 23.73): Median value
I((23.73/10)^1)#[1] 2.373
# New data frames with different possible scenarios
x1 <- data.frame(Gangion="Wire",Season="1",Year="2010",Latitude.t=-
0.4170317,Longitude.t=1.808,AquaChlo.t=0.6904,NCDCsst_l4.t=2.373, Hookn=1000)
x2 <- data.frame(Gangion="Wire",Season="1",Year="2010",Latitude.t=-
0.4170317,Longitude.t=3.221,AquaChlo.t=0.6904,NCDCsst_l4.t=2.373, Hookn=1000)
x3 <- data.frame(Gangion= "Wire",Season="1", Year="2010", Latitude.t=1.239582,
Longitude.t=1.808,AquaChlo.t=0.6904,NCDCsst_l4.t=2.373, Hookn=1000)
x4 <- data.frame(Gangion="Wire", Season="1", Year="2010", Latitude.t=1.239582,
Longitude.t=3.221, AquaChlo.t=0.6904,NCDCsst_l4.t=2.373, Hookn=1000)
x5 <- data.frame(Gangion="Mono",Season="1",Year="2010",Latitude.t=-0.4170317,
Longitude.t=1.808, AquaChlo.t=0.6904,NCDCsst_l4.t=2.373, Hookn=1000)
x6 <- data.frame(Gangion="Wire",Season="2",Year="2010",Latitude.t=-0.4170317,
Longitude.t=1.808, AquaChlo.t=0.6904,NCDCsst_l4.t=2.373, Hookn=1000)
```

```
x7 <- data.frame(Gangion="Wire",Season="3",Year="2010",Latitude.t=-0.4170317,
Longitude.t=1.808, AquaChlo.t=0.6904,NCDCsst_l4.t=2.373, Hookn=1000)
x8 <- data.frame(Gangion="Wire",Season="4",Year="2010",Latitude.t=-0.4170317,
Longitude.t=1.808,AquaChlo.t=0.6904,NCDCsst_l4.t=2.373, Hookn=1000)
# Example of predictions for the GLM Gamma models, predicting CPUE+1
predict(fit.gamma55, newdata=x1, type="response", se.fit = T).
predict(fit.gamma55, newdata=x2, type="response", se.fit = T).
predict(fit.gamma55, newdata=x3, type="response", se.fit = T)
predict(fit.gamma55, newdata=x4, type="response", se.fit = T)
predict(fit.gamma55, newdata=x5, type="response", se.fit = T)
predict(fit.gamma55, newdata=x6, type="response", se.fit = T)
predict(fit.gamma55, newdata=x7, type="response", se.fit = T)
predict(fit.gamma55, newdata=x8, type="response", se.fit = T)
# Example of predictions for the GLM NB models, predicting catches (n) in 1000 hooks
predict(fit.nb55, newdata=x1, type="response", se.fit = T)
predict(fit.nb55, newdata=x2, type="response", se.fit = T)
predict(fit.nb55, newdata=x3, type="response", se.fit = T)
predict(fit.nb55, newdata=x4, type="response", se.fit = T)
predict(fit.nb55, newdata=x5, type="response", se.fit = T)
predict(fit.nb55, newdata=x6, type="response", se.fit = T)
predict(fit.nb55, newdata=x7, type="response", se.fit = T)
predict(fit.nb55, newdata=x8, type="response", se.fit = T)
```