

**UNIVERSIDADE FEDERAL DE SANTA CATARINA  
CENTRO DE CIÊNCIAS BIOLÓGICAS**

Fernando de Pol Mayer

**MODELOS PROBABILÍSTICOS PARA DADOS DE CAPTURA  
INFLACIONADOS DE ZEROS E SUPERDISPERSOS EM  
PESCARIAS COMERCIAIS**

Florianópolis  
2011



Fernando de Pol Mayer

**MODELOS PROBABILÍSTICOS PARA DADOS DE CAPTURA  
INFLACIONADOS DE ZEROS E SUPERDISPERSOS EM  
PESCARIAS COMERCIAIS**

Dissertação submetida ao Programa de Pós-  
graduação em Ecologia para a obtenção do  
Grau de Mestre em Ecologia.

Orientador: Dr. Humber Agrelli Andrade

Florianópolis  
2011

Catálogo na fonte pela Biblioteca Universitária  
da  
Universidade Federal de Santa Catarina

M468m Mayer, Fernando de Pol  
Modelos probabilísticos para dados de captura inflacionados  
de zeros e superdispersos em pescarias comerciais  
[dissertação] / Fernando de Pol Mayer ; orientador, Humber  
Agreli Andrade. - Florianópolis, SC, 2011.  
76 p.: grafs., tabs., mapas

Dissertação (mestrado) - Universidade Federal de Santa  
Catarina, Centro de Ciências Biológicas. Programa de Pós-  
Graduação em Ecologia.

Inclui referências

1. Ecologia. 2. Indústria pesqueira - Métodos estatísticos.  
3. Pesca - Aspectos econômicos. 4. Recursos marinhos -  
Conservação. I. Andrade, Humber Agreli. II. Universidade  
Federal de Santa Catarina. Programa de Pós-Graduação em  
Ecologia. III. Título.

CDU 577.4

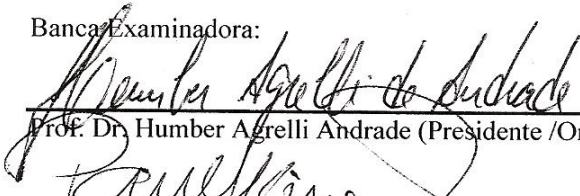
**“Modelos probabilísticos para dados de captura inflacionados de zeros e superdispersos em pescarias comerciais”**


por

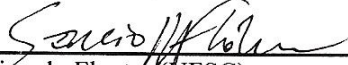
**Fernando de Pol Mayer**


Dissertação julgada e aprovada em sua forma final pelos membros titulares da Banca Examinadora (Port. 03/PPGECO/2011) do Programa de Pós-Graduação em Ecologia - UFSC, composta pelos Professores Doutores:

Banca Examinadora:

  
\_\_\_\_\_  
Prof. Dr. Humber Agrelli Andrade (Presidente /Orientador)

  
\_\_\_\_\_  
Prof. Dr. Paul Gerhard Kinas (Externo/IMEF-LEA/FURG)

  
\_\_\_\_\_  
Prof. Dr. Sérgio Ricardo Floeter (UFSC)

  
\_\_\_\_\_  
Prof. Dr. Mauricio Mello Petrucio  
Coordenador do Programa de Pós-Graduação em Ecologia

Florianópolis, 24 de fevereiro de 2011



## Agradecimentos

Aos meus pais por terem me apoiado em todas as etapas deste e de outros trabalhos.

À Debora, por toda ajuda, carinho e paciência nessa fase.

Ao prof. Humber, pelos ensinamentos que me passou até aqui.

Aos professores Paul G. Kinas, Sergio R. Floeter, e Paulo Pagliosa por terem aceitado o convite para a banca e prestado comentários valiosos.

Ao Programa de Pós-graduação em Ecologia, especialmente aos professores Mauricio Petrucio e Natalia Hanazaki por terem acreditado até o último momento que essa dissertação poderia ser terminada.

Ao Paulo, pela amizade e pelas pequenas farras no apartamento de Floripa.

À todos os meus amigos e amigas do mestrado, que sem eles, as poucas horas de distração não teriam sido tão felizes.

À Adriana, que me “abrigou” em sua sala — o único lugar fixo que consegui para estudar dentro da UFSC.

Ao Rodrigo Sant’Ana pela ajuda toda em Brasília, e ao Rodrigo Claudino pelas festas todas em Brasília.

Ao Linus Torvalds, Richard Stallman, Leslie Lamport, Ross Ihaka, Robert Gentlemen e muitos outros que fazem do mundo um lugar mais livre.





*Probability theory is nothing but common sense reduced to calculation.*

— Pierre Simon, Marquis de Laplace, 1819

*A new civilization is emerging in our lives, and blind men everywhere are trying to suppress it. This new civilization brings with it new family styles; changed ways of working, loving, and living; a new economy; new political conflicts; and beyond all this an altered consciousness as well. Pieces of this new civilization exist today. Millions are already attuning their lives to the rhythms of tomorrow. Others, terrified of the future, are engaged in a desperate, futile flight into the past and are trying to restore the dying world that gave them birth.*

— Alvin Toffler, 1980

*When dealing with numerical matrices you have to be prepared for the unexpected.*

— Bill Venables, 2010



## Resumo

As capturas incidentais (*i.e.* espécies capturadas sem intenção) geram, além do impacto ambiental, um problema econômico de larga escala em diversas pescarias comerciais. A pescaria com espinhel de superfície é uma das mais rentáveis do mundo, pois têm como alvo espécies de grande valor comercial como os atuns e o espadarte. No entanto, além destas espécies-alvo, também são capturadas espécies com populações frágeis e ameaçadas, como é o caso dos agulhões branco (*Tetrapturus albidus*) e negro (*Makaira nigricans*). A *International Commission for the Conservation of Atlantic Tunas* (ICCAT), instituição responsável pelo gerenciamento dos estoques de atuns e afins no oceano Atlântico, vêm solicitando para que seus países membros adotem medidas para a redução da captura destas espécies, devido à fragilidade dos estoques. Portanto, com a finalidade de dar apoio à medidas de manejo que possam auxiliar na redução da captura incidental, neste trabalho foram analisados dados da captura dos agulhões branco e negro no oceano Atlântico sul. Foram utilizados modelos com a intenção de se obter as probabilidades de captura (maior do que zero) destas espécies por área e época do ano. Com a identificação de áreas com probabilidades elevadas de captura, medidas de manejo baseadas em fechamento de áreas para a pesca em determinados períodos do ano podem ser utilizadas. Dados de espécies capturadas incidentalmente apresentam como característica um número de zeros excessivo, e frequentemente, também, superdispersão. Por esse motivo, modelos que consideram tanto o excesso de zeros quanto a superdispersão foram utilizados. Nesta abordagem, os modelos são chamados de “modelos de mistura” quando a variável resposta é modelada como uma mistura de uma distribuição de Bernoulli para a presença/ausência de zeros e uma outra distribuição para os valores positivos, e “modelos condicionais” quando a modelagem da variável resposta é feita em duas etapas, utilizando uma distribuição discreta truncada para os valores positivos e uma regressão logística tradicional para os valores nulos. Em ambas as abordagens, as distribuições de Poisson e binomial negativa (BN) foram utilizadas. A comparação entre os modelos foi feita através dos critérios de informação de Akaike (AIC) e bayesiano (BIC). O modelo com o menor valor destes critérios foi então selecionado para a continuidade das análises. Foram calculadas a Captura Por Unidade de Esforço (CPUE) padronizada pelo modelo, e as probabilidades de captura por ano, trimestre e área. Os resultados mostraram que para ambas as espécies, o padrão anual de CPUE está em queda, sugerindo também a redução dos estoques destas populações. As áreas com maiores probabilidades de captura são aquelas localizadas nas regiões tropicais do oceano Atlântico sul, especialmente nos meses do verão austral. Áreas mais centrais e à oeste, próximas à costa do Brasil, também mostraram valores elevados de probabilidade. Com estes resultados, pode-se concluir que os modelos utilizados foram apropriados, e devem ser considerados quando a intenção for modelar a captura de espécies capturadas incidentalmente, onde predominam o excesso de zeros e a superdispersão.

**Palavras-chave:** captura incidental, excesso de zeros, superdispersão, CPUE.



## Abstract

Incidental catches, or bycatch, bring up, besides the environmental impact, economical problems of large scale in several fisheries. The surface longline fishery is one of the most profitable in the world, since target species, as tuna and swordfish, have high commercial values. However, besides these, non-targeted species such as the white marlin (*Tetrapturus albidus*) and the blue marlin (*Makaira nigricans*) are also captured. The *International Commission for the Conservation of Atlantic Tunas* (ICCAT) is responsible for the management of tuna and tuna-like species in the Atlantic ocean, and for several years has appealing for reduction on the catch of unwanted species, due to the fragility of the stocks. In this work were analysed capture data of white and blue marlin in the south Atlantic ocean, aiming to give support for management measures attempting to reduce incidental catches. Models were applied for obtaining the catch probabilities of these species for area and season. The identification of high area/season probabilities may give support for management measures attempting for closing such fishing areas. Bycatch species data usually show an excessive number of zeros, and frequently, also, overdispersion. Models that consider as much as the excess zeros as overdispersion were used. These models are known as “mixture models”, when the response variable is modeled as a mixture of a Bernoulli distribution for the presence/absence of zeros and another distribution for positive values, and “conditional models”, when the modelling of the response variable is done in two stages, one using a discrete truncated distribution for positive values and another applying a traditional logistic regression for the null values. Poisson and negative binomial (NB) distributions were both used. Comparison between models were done using Akaike’s information criteria (AIC) and bayesian information criteria (BIC). The model with lower value of both these criteria was selected for the continuation of analysis. Standardized Catch Per Unit Effort (CPUE) was calculated based on model coefficients, and catch probabilities by year, quarter and area were also obtained. Results for both species showed that annual CPUE values are decreasing, also suggesting that these population stocks may be declining. Higher catch probabilities were found in tropical areas of the south Atlantic ocean, mainly in the austral summer. Central and western areas, located close to Brazilian coast, also exhibited high probability values. Considering such results, it may be concluded that the models utilized in this approach were appropriate, and are suggested to be used when the intention is modelling the capture of bycatch species, where both excess zeros and overdispersion are common.

**Keywords:** by-catch, excess zeros, overdispersion, CPUE.



## Lista de Tabelas

- 1 Número de registros presente na base de dados da ICCAT por ano e trimestre, referentes à pescarias de espinhel de superfície realizados no oceano Atlântico sul pela frota japonesa. Os valores são referentes à base de dados atualizada pela última vez no *site* da ICCAT em novembro de 2010. . . . . 35
- 2 Número de parâmetros, logaritmo da versossimilhança ( $\log L$ ), critério de informação de Akaike (AIC), critério de informação bayesiano (BIC), e número de zeros preditos (o número de zeros observados foi de 9037) pelos diferentes modelos ajustados para a captura do agulhão branco. . . . . 51
- 3 Estimativas dos parâmetros do modelo condicional ajustado com a distribuição binomial negativa para a captura do agulhão branco. As colunas “Contagem” se referem aos coeficientes do modelo BN para as contagens. As colunas “Zero” representam os coeficientes para o modelo logístico considerando a presença/ausência de zeros. . . . . 52
- 4 Número de parâmetros, logaritmo da versossimilhança ( $\log L$ ), critério de informação de Akaike (AIC), critério de informação bayesiano (BIC), e número de zeros preditos (o número de zeros observados foi de 6343) pelos diferentes modelos ajustados para a captura do agulhão negro. . . . . 61
- 5 Estimativas dos parâmetros do modelo condicional ajustado com a distribuição binomial negativa para a captura do agulhão negro. As colunas “Contagem” se referem aos coeficientes do modelo BN para as contagens. As colunas “Zero” representam os coeficientes para o modelo logístico considerando a presença/ausência de zeros. . . . . 63





## Lista de Figuras

1	Ilustração do agulhão branco ( <i>Tetrapturus albidus</i> ). Fonte: FAO. . . . .	26
2	Ilustração do agulhão negro ( <i>Makaira nigricans</i> ). Fonte: FAO. . . . .	27
3	Divisão e nome das áreas no oceano Atlântico Sul, utilizadas para a modelagem das capturas dos agulhões branco e negro. . . . .	42
4	Esforço de pesca (número de anzóis $\times$ 1000) realizado pela frota japonesa de espinhel de superfície no oceano Atlântico, somado por trimestre para o período 1970–2009. . . . .	47
5	Captura por Unidade de Esforço (número de peixes capturados por lance com 1000 anzóis) trimestral do agulhão branco, obtida pela frota japonesa de espinhel de superfície no oceano Atlântico. Os quadrantes marcados com o símbolo + indicam o valor zero. . . . .	48
6	Distribuição da Captura por Unidade de Esforço (número de peixes capturados por lance com 1000 anzóis) trimestral do agulhão branco, obtida pela frota japonesa de espinhel de superfície em cada área no oceano Atlântico, entre 1970–2009. Os valores foram transformados para a escala $\log(x + 0,5)$ para facilitar a visualização. . . . .	49
7	Média da Captura por Unidade de Esforço (número de peixes capturados por lance com 1000 anzóis) trimestral do agulhão branco, obtida pela frota japonesa de espinhel de superfície no oceano Atlântico, entre 1970–2009. Os valores foram transformados para a escala $\log(x + 0,5)$ para facilitar a visualização. . . . .	50
8	Proporção do número de agulhões branco capturados no Atlântico Sul no período 1970–2009. Como os valores de captura variaram no intervalo $[0, 843]$ , aqueles maiores que 19 foram agregados na classe 20+. . . . .	51
9	Histograma de frequência (esquerda) e gráfico de dispersão (direita) dos resíduos do modelo condicional BN para o agulhão branco. . . . .	55
10	Captura por Unidade de Esforço (número de peixes capturados por lance com 1000 anzóis) do agulhão branco observada e estimada pelo modelo condicional BN. Os valores foram padronizados, tendo o primeiro ano como referência. . . . .	56

11	Probabilidade de captura trimestral por grande área do agulhão branco no oceano Atlântico Sul. São apresentados os resultados para o primeiro ano (1970), o ano central (1989) e o último ano (2009) da série temporal analisada (1970–2009).	57
12	Captura por Unidade de Esforço (número de peixes capturados por lance com 1000 anzóis) trimestral do agulhão negro, obtida pela frota japonesa de espinhel de superfície no oceano Atlântico. Os quadrantes marcados com o símbolo + indicam o valor zero. . . . .	58
13	Distribuição da Captura por Unidade de Esforço (número de peixes capturados por lance com 1000 anzóis) trimestral do agulhão negro, obtida pela frota japonesa de espinhel de superfície em cada área no oceano Atlântico, entre 1970–2009. Os valores foram transformados para a escala $\log(x + 0,5)$ para facilitar a visualização. . . . .	59
14	Média da Captura por Unidade de Esforço (número de peixes capturados por lance com 1000 anzóis) trimestral do agulhão branco, obtida pela frota japonesa de espinhel de superfície no oceano Atlântico, entre 1970–2009. Os valores foram transformados para a escala $\log(x + 0,5)$ para facilitar a visualização. . . . .	60
15	Proporção do número de agulhões negro capturados no Atlântico Sul no período 1970–2009. Como os valores de captura variaram no intervalo $[0, 681]$ , aqueles maiores que 19 foram agregados na classe 20+. . . . .	60
16	Histograma de frequência (esquerda) e gráfico de dispersão (direita) dos resíduos do modelo condicional BN para o agulhão negro. . . . .	65
17	Captura por Unidade de Esforço (número de peixes capturados por lance com 1000 anzóis) do agulhão negro observada e estimada pelo modelo condicional BN. Os valores foram padronizados, tendo o primeiro ano como referência. . . . .	65
18	Probabilidade de captura trimestral por grande área do agulhão negro no oceano Atlântico Sul. São apresentados os resultados para o primeiro ano (1970), o ano central (1989) e o último ano (2009) da série temporal analisada (1970–2009).	67

# Sumário

<b>1</b>	<b>Introdução</b>	<b>21</b>
1.1	Distribuição e biologia dos agulhões . . . . .	24
1.1.1	Agulhão branco ( <i>T. albidus</i> Poey, 1860) . . . . .	25
1.1.2	Agulhão negro ( <i>M. nigricans</i> Lacepède, 1802) . . . . .	27
1.2	Modelos para dados inflacionados de zeros . . . . .	28
<b>2</b>	<b>Objetivos</b>	<b>31</b>
<b>3</b>	<b>Materiais e Métodos</b>	<b>33</b>
3.1	Dados . . . . .	33
3.2	Análise . . . . .	36
3.2.1	Análise descritiva e exploratória . . . . .	36
3.2.2	Formulação dos modelos . . . . .	36
3.2.3	Definição de variáveis e seleção de modelos . . . . .	41
3.2.4	Diagnósticos . . . . .	43
3.2.5	Estimativas dos parâmetros . . . . .	44
<b>4</b>	<b>Resultados</b>	<b>47</b>
4.1	Esforço de pesca . . . . .	47
4.2	Agulhão branco . . . . .	48
4.2.1	Análise descritiva . . . . .	48
4.2.2	Modelagem . . . . .	49
4.3	Agulhão negro . . . . .	58
4.3.1	Análise descritiva . . . . .	58
4.3.2	Modelagem . . . . .	59
<b>5</b>	<b>Discussão</b>	<b>68</b>
5.1	Modelos . . . . .	68
5.2	Aplicações à ecologia dos agulhões . . . . .	69
5.3	Sugestões para trabalhos futuros . . . . .	71
	<b>Referências</b>	<b>73</b>



# 1 Introdução

Pescarias comerciais são atividades extrativistas de larga escala, praticadas em todos os oceanos. Há grandes variações quanto às espécies capturadas e ao ambiente em que se encontram, desde pequenos animais bentônicos (moluscos e camarões) até grandes peixes pelágicos (atuns e tubarões). O impacto gerado pela extração de organismos pela pesca vai desde a alteração do ambiente físico (*e.g.* redes de arrasto passando sobre fundos coralinos) até a depleção dos estoques pesqueiros.

Uma das modalidades de pesca de maior importância econômica é aquela praticada com espinhéis de superfície, que tem como alvo a captura de atuns e espécies afins em regiões oceânicas. No Atlântico, estas pescarias são realizadas principalmente por frotas de países como Estados Unidos, Espanha e Japão. Este tipo de espinhel pelágico consiste de uma extensa linha principal suspensa por bóias. Nesta linha principal são conectadas diversas linhas secundárias, as quais tem na extremidade um único anzol. Um espinhel tradicional possui em torno de 60 km de extensão, por onde são distribuídos cerca de 1100 anzóis (Arfelli, 1996).

Em pescarias oceânicas, como as de espinhel de superfície, muitos dos estoques explorados são compartilhados por diversos países. Isso se deve à ampla distribuição e a capacidade de migração dos grandes peixes pelágicos (Sund *et al.*, 1981). Sendo assim, torna-se inviável que cada país tenha suas próprias medidas de manejo. Portanto, para que medidas reguladoras fossem elaboradas e compartilhadas entre as diferentes nações que possuem alguma frota atuando sobre atuns e espécies afins no oceano Atlântico, criou-se a *International Commission for the Conservation of Atlantic Tunas* (ICCAT). Esta comissão tem como finalidade: (a) requerer e organizar as informações pesqueiras de todos os países membros; (b) promover reuniões para a avaliação dos estoques; e (c) recomendar medidas de manejo para as espécies de atuns e afins.

Inicialmente, o objetivo da ICCAT era recomendar medidas de manejo visando apenas a pescaria de atuns capturados em quantidade e/ou de grande importância econômica (*e.g.* as espécies do gênero *Thunnus*). No entanto,

recentemente, espécies consideradas como fauna acompanhante vêm recebendo maior atenção por parte da comissão, tanto pela tendência de aumento da captura e/ou pela fragilidade das populações. Duas das espécies que se enquadram neste último caso são os agulhões branco (*Tetrapturus albidus*) e negro (*Makaira nigricans*). A preocupação da ICCAT se deve não somente ao aumento das capturas, mas também ao fato dos agulhões serem espécies ecologicamente emblemáticas e de grande apelo conservacionista. Além disso, a pescaria esportiva e recreacional destas espécies é uma atividade de alta rentabilidade, especialmente em países desenvolvidos como, por exemplo, os Estados Unidos. Soma-se ainda a esta problemática, o fato de que o valor comercial dos agulhões é alto. Assim, ainda que não sejam alvo das pescarias comerciais e o volume capturado seja baixo, é compensatório para os pescadores capturar e vender agulhões.

A redução das mortalidades dos agulhões branco e negro tem sido tomada como meta internacional devido à tendência atual de declínio da abundância dos estoques (ICCAT, 2003, 2007a). A gravidade do problema fez com que, em 2000, fosse criado um *Plano de Recuperação das Populações de Agulhões Branco e Negro* (Recomendação 00-13 da ICCAT), recentemente aprimorado e substituído pela Recomendação 06-09 (2006). Este Plano de Recuperação estabelece duas medidas principais: (a) a captura total anual que pode ser desembarcada não deve ultrapassar 578 t (agulhão branco) e 2303 t (agulhão negro); e (b) todos os exemplares capturados e que ainda estiverem vivos no momento do recolhimento do espinhel devem ser liberados. Mesmo com esta regulamentação, a captura anual dos últimos anos vêm se mantendo acima das capturas máximas recomendadas (ICCAT, 2007a, 2008). Além disso, a liberação de indivíduos capturados vivos é ainda uma questão em aberto pois não existem informações suficientes sobre as estimativas da proporção de capturados vivos/mortos, e nem da frequência com que os vivos são liberados (ICCAT, 2007a, 2008).

No caso da frota brasileira de espinhel de superfície, a situação é ainda mais grave pois ela é uma das responsáveis pelas maiores capturas de agulhões no Atlântico (ICCAT, 2007b). Devido à toda esta mobilização internacional para a recuperação das populações dos agulhões, em 2005 o Governo

brasileiro publicou uma Instrução Normativa (IN n° 12, de 14 de julho de 2005) com a intenção de diminuir o desembarque destas espécies. Como principais medidas, esta Normativa estabelece que: (a) é proibido o descarte dos agulhões encontrados mortos no momento do recolhimento do espinhel (os exemplares devem ser destinados à doação e não podem ser comercializados); e (b) é obrigatória a devolução ao mar dos espécimes encontrados vivos no recolhimento do espinhel.

Apesar do cenário de declínio na abundância das populações de agulhões, se for considerado o elevado esforço realizado pelas frotas pesqueiras, a captura destas espécies em pescarias comerciais pode ser considerada pouco frequente. Ou seja, a captura de agulhões é um evento que ocorre raramente em um grande número de lances de pesca. De fato, a maioria dos pescadores considera a captura destas espécies como incidental (Andrade, 2007).

Desta forma, embora haja um esforço do Governo para proibir a comercialização e estimular a devolução dos indivíduos vivos, ainda não existe nenhum mecanismo que impeça a captura destas espécies, de forma que medidas que auxiliem na identificação de alternativas de manejo voltadas à diminuição das mortalidades por pesca dos agulhões branco e negro tornam-se valiosas, tanto como estratégias ecológicas quanto políticas. Este trabalho visa justamente dar apoio às medidas nacionais e internacionais que vêm sendo realizadas para promover a conservação das populações de agulhões no Atlântico, através da identificação de cenários onde as probabilidades de captura (maior do que zero) sejam altas. Assim, mediante a identificação de uma região e trimestre de maior probabilidade de captura de indivíduos de uma ou outra espécie, haveria a alternativa de fechar aquela área de pesca durante determinado período.

Por se tratar de espécies pouco frequentes, foram utilizados na análise modelos estatísticos que levam o excesso (ou a “inflação”) de zeros em consideração. Genericamente, foi utilizada a designação “modelos para dados inflacionados de zeros” para se referir a uma grande gama de abordagens. Sob o prisma conservacionista e ecológico, este trabalho traz informações que podem auxiliar no desenvolvimento de estratégias que promovam a redução da mortalidade dos agulhões branco e negro, promovendo a conservação

das populações destas espécies. Sob o ponto de vista político, é fundamental que o Brasil consiga diminuir a mortalidade dos agulhões para atender as recomendações da ICCAT. A elevada captura realizada pelo país pode gerar uma situação de desconforto em reuniões internacionais, como por exemplo, uma menor abertura nas discussões onde existe interesse nacional, como a definição das cotas de captura.

Ressalta-se que a metodologia empregada nesta análise pode ser empregada estrategicamente na investigação sobre a conservação de outras pescarias. A aplicação de modelos estatísticos para resolver problemas práticos em áreas que não são relacionadas às Ciências Exatas vem crescendo muito em outros países. No entanto, no Brasil o uso da modelagem estatística em Ecologia está restrita a poucos pesquisadores e instituições. Portanto, a geração de informações relacionadas à essa área da ciência deve também servir para a difusão (e a discussão) da modelagem entre pesquisadores brasileiros que trabalham com Ecologia.

Neste sentido, ressalta-se que os resultados encontrados podem auxiliar tanto no fortalecimento da posição política do Brasil no cenário internacional, demonstrando que o país vem fazendo esforços para solucionar o problema da elevada mortalidade de agulhões, quanto favorecer o crescimento da aplicação de modelos estatísticos na área científica, considerada importante em âmbito internacional.

Para maior compreensão sobre o problema, será apresentada uma breve revisão sobre a distribuição, a biologia e a pesca dos agulhões, seguida de uma revisão sobre o tema dos modelos para dados inflacionados de zeros.

## 1.1 Distribuição e biologia dos agulhões

Os agulhões são peixes que pertencem a uma classe denominada “peixes de bico”, e que incluem os membros das famílias *Xiphiidae* (espadarte) e *Istiophoridae* (diversas espécies de agulhões). São caracterizados pelo alongamento da mandíbula superior, muito além da inferior, formando um longo *rostrum* em forma de espada (espadarte) ou arredondado e em forma de lança (agulhões). A família *Istiophoridae*, na qual os agulhões são classi-



ficados, possui três gêneros: *Istiophorus*, *Tetrapturus*, e *Makaira*.

Estas espécies são oceânicas e epipelágicas, e devido ao formato alongado do corpo e à habilidade de rápida natação, habitam águas tropicais e temperadas, e ocasionalmente águas frias de todos os oceanos (Nakamura, 1985). Geralmente estão confinadas às camadas de água acima da termoclina, mas algumas podem ocorrer em maiores profundidades. Saito & Yokawa (2006) encontraram evidências de que a profundidade de ocorrência está correlacionada com a profundidade da camada de mistura, sendo que com o aumento desta camada, maior poderá ser a profundidade de “mergulho” dos agulhões. Em particular, os agulhões branco e negro são espécies que se distribuem por todo o oceano Atlântico, sendo que *M. nigricans* é tipicamente mais oceânica. O padrão de migração envolve movimentos sazonais para águas temperadas ou frias para alimentação e para águas tropicais ou subtropicais para desova (Nakamura, 1985). Para fins de manejo, a ICCAT considera que as populações de agulhões branco e negro no Atlântico se dividem em dois estoques, um ao norte e outro ao sul de 5°N. No entanto, discussões recentes sobre a possibilidade de se considerar um único estoque estão ocorrendo em reuniões internacionais, embora ainda não hajam pesquisas conclusivas (ICCAT, 2007a).

Todas as espécies são dióicas (sexos separados) (ICCAT, 2003), mas não existe dimorfismo sexual em características morfológicas ou nos padrões de coloração, apenas constata-se que as fêmeas podem atingir tamanhos maiores que os machos (Nakamura, 1985).

A seguir são apresentadas algumas informações específicas das espécies de interesse neste projeto. Devido à escassez na literatura a respeito da biologia dos agulhões, esta revisão foi baseada principalmente no texto de Nakamura (1985).

### **1.1.1 Agulhão branco (*T. albidus* Poey, 1860)**

**Distribuição geográfica** O agulhão branco (Figura 1) se distribui por todo oceano Atlântico, atingindo águas mais rasas do que o agulhão negro. Sua extensão latitudinal de distribuição vai de 45°N até 45°S no oeste, e até 35°S no leste do Atlântico.

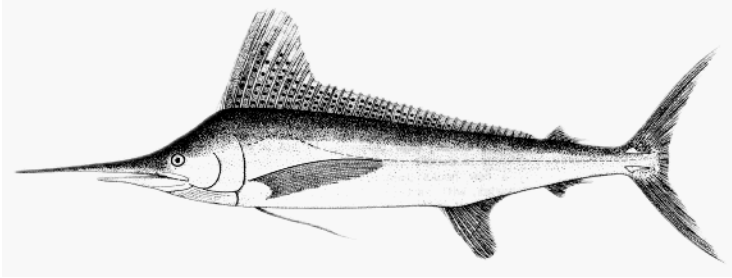


Figura 1: Ilustração do agulhão branco (*Tetrapturus albidus*). Fonte: FAO.

**Habitat e biologia** É uma espécie pelágica e oceânica, geralmente situada acima da termoclina. Sua distribuição varia sazonalmente, atingindo as altas latitudes em ambos hemisférios somente durante os períodos mais quentes. Geralmente *T. albidus* é encontrado em águas claras e relativamente mais profundas (abaixo dos 100 m), com TSM em torno de 22°C. No Golfo do México, por exemplo, pescadores relatam que o sucesso na captura desta espécie aumenta proporcionalmente com o aumento da claridade da água (Nakamura, 1985). Em algumas áreas, pode ser encontrada na interface de duas camadas de água ou em concentrações de algas. Sua distribuição também é influenciada pela topografia do fundo oceânico. Ambientes escarpados, cânions submarinos e bancos submersos, quando localizados em áreas onde as propriedades da água favorecem, são importantes locais de concentração e alimentação, o que também favorece uma alta produtividade pesqueira. Não é considerada uma espécie que forma cardumes, e são encontrados mais frequentemente indivíduos solitários ou aos pares. No entanto, ocasionalmente podem formar cardumes de acordo com tamanho ou sexo em algumas épocas do ano. Os agulhões branco normalmente utilizam o bico para atordoar ou matar suas presas. As lulas são o item alimentar mais frequente na dieta desta espécie, embora dourados, cavalas e bonitos, além de outros peixes menores e crustáceos, também fazem parte de sua alimentação.

**Tamanho** Esta espécie atinge um tamanho máximo de cerca de 280 cm (comprimento total) e em torno de 82 kg em peso. O tamanho dos agulhões

branco capturados pelos barcos de espinhel de superfície varia de 85–90 cm até 200–205 cm (Amorim & Arfelli, 2003).

### 1.1.2 Agulhão negro (*M. nigricans* Lacepède, 1802)

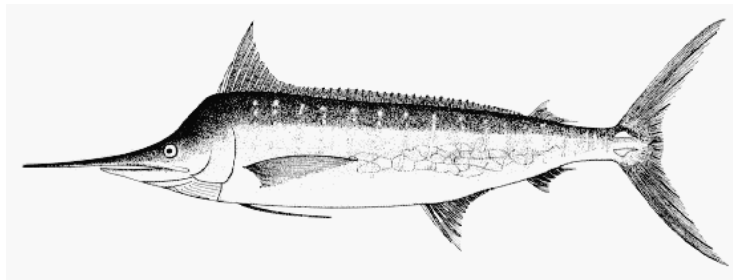


Figura 2: Ilustração do agulhão negro (*Makaira nigricans*). Fonte: FAO.

**Distribuição geográfica** O agulhão negro (Figura 2) distribuiu-se principalmente em águas temperadas e tropicais do oceano Atlântico, sendo considerada a mais oceânica de todas as espécies de peixes de bico. Sua amplitude latitudinal de distribuição se estende desde 45°N até 40°S no oeste, e 35°S no leste do Atlântico.

**Habitat e biologia** Esta é uma espécie epipelágica e oceânica, geralmente encontrada em águas com Temperatura Superficial do Mar (TSM) entre 22°C e 31°C. Sua ocorrência também pode estar condicionada à ocorrência de outras espécies de agulhões, provavelmente devido às interações ecológicas. Por exemplo, Nakamura (1985) observou que pescadores tendem a capturar mais agulhões negro quando as capturas de agulhões branco são menores, e vice-versa. No oceano aberto raramente formam cardumes, e geralmente são encontrados indivíduos solitários e dispersos. Sua alimentação ocorre próxima à superfície da água, mas algumas vezes podem realizar migrações verticais profundas em busca de alimento. Os itens alimentares mais comumente encontrados em estômagos de agulhões negro são os dourados e atuns e espécies afins como a cavala.

**Tamanho** O agulhão negro possui peso médio entre 136 kg e 181 kg. O tamanho máximo pode chegar a 375 cm (comprimento do corpo) e 580 kg em peso. O tamanho dos espécimes capturados em pescarias comerciais com espinhel de superfície encontrado por Amorim & Arfelli (2001) variou entre 160 cm (22 kg) e 345 cm (428 kg).

## 1.2 Modelos para dados inflacionados de zeros

Normalmente, em avaliações de estoques pesqueiros, a questão de interesse é estimar a abundância de uma espécie (geralmente em peso), e também a tendência desta abundância ao longo dos anos. Existe uma longa tradição na ciência pesqueira no uso da Captura Por Unidade de Esforço (CPUE) como um índice relativo de abundância populacional (Gulland, 1983; Hilborn & Walters, 1992). No entanto, a CPUE sofre a interferência de diversos outros fatores além da abundância (Hilborn & Walters, 1992). Modelos Lineares Generalizados (MLGs) (Nelder & Wedderburn, 1972) têm sido frequentemente utilizados para identificar e minimizar o efeito de fatores que não são de interesse, na busca por um índice de abundância razoável (Maunder & Punt, 2004). Quando esta modelagem envolve dados de espécies pouco frequentes (*i.e.* existe uma grande parcela de capturas iguais a zero), abordagens mais complexas devem ser consideradas.

Neste trabalho foi utilizada uma abordagem voltada ao ajuste de distribuições de probabilidade para a variável “captura” de maneira integrada, ou seja, capturas positivas e iguais a zero. A maioria dos modelos desenvolvidos para essa finalidade é direcionada à análise de dados discretos, em forma de contagens.

A distribuição de Poisson normalmente é utilizada para modelar dados em forma de contagens. Seja  $Y = \{Y_1, \dots, Y_j\}$  com  $j = 1, \dots, n$  eventos de pesca, a variável aleatória “número de agulhões capturados”. Então, segundo a distribuição de Poisson, a probabilidade de  $Y$  assumir um valor qualquer  $y_i$  em um dado evento  $j$  é dada por

$$P\{Y = y_i\} = \frac{\lambda^{y_i} e^{-\lambda}}{y_i!}, \quad y_i = 0, 1, 2, \dots$$

onde o parâmetro  $\lambda$  é o número médio de indivíduos capturados. Uma propriedade da distribuição de Poisson é que a esperança é igual a variância, ou seja,  $E(Y) = Var(Y) = \lambda$ . As realizações de  $Y$  são normalmente observadas em um intervalo de tempo ou espaço. No caso dos dados de pesca, poderia se considerar o número de agulhões capturados por lance, já que a duração de um lance de pesca é aproximadamente a mesma (lançamento no final da tarde e recolhimento no início da manhã) (Arfelli, 1996). Os efeitos de diversas variáveis explicativas sobre  $Y$  podem ser modelados através de ajustes no parâmetro  $\lambda$ . Este procedimento é chamado de regressão de Poisson, um método bastante utilizado para a modelagem de dados em forma de contagem (Ridout *et al.*, 1998).

No entanto, na prática é muito comum que a variância seja maior, e não igual, à média em relação à distribuição de Poisson. Este fato é denominado de *superdispersão* (McCullagh & Nelder, 1989; Hinde & Demétrio, 1998). Não considerar esta superdispersão pode levar a sérias subestimativas de erros-padrões e a uma inferência deficiente dos parâmetros da regressão (Hinde & Demétrio, 1998). Considerar esta superdispersão no modelo tradicional de Poisson ou ajustar um modelo através da distribuição binomial negativa (que permite uma “extra-dispersão”), pode ser uma maneira simples de lidar com dados onde existe um número moderado de zeros (Cunningham & Lindenmayer, 2005). No entanto, quando existe ainda um excesso de zeros, ambos procedimentos sofrem da desvantagem de que a proporção de zeros deve necessariamente estar relacionada com a distribuição dos valores positivos, o que geralmente resulta em ajustes insatisfatórios no caso de dados ecológicos (Welsh *et al.*, 1996).

Há abordagens alternativas, à princípio mais satisfatórias, para lidar com o excesso de zeros e com a superdispersão em modelos que envolvem contagens. Duas delas envolvem os chamados *modelos de mistura* (Lambert, 1992), e os *modelos condicionais* (Welsh *et al.*, 1996).

O primeiro trabalho a considerar modelos de mistura para contagens com muitos zeros foi o de Lambert (1992), que aplicou a técnica à dados de defeitos de fabricação em produtos manufaturados, e chamou a metodologia de Regressão de Poisson Inflacionada de Zeros, ou, do inglês, *Zero-Inflated*

*Poisson Regression* (ZIP). Nesta abordagem a variável resposta é modelada como uma mistura de uma distribuição de Bernoulli com uma distribuição de Poisson ou binomial negativa.

Os modelos condicionais foram inicialmente apresentados por Welsh *et al.* (1996), e também são chamados de “modelos de barreira” (Ridout *et al.*, 1998). Nesta abordagem, a modelagem da variável resposta é realizada em duas etapas. A primeira etapa é quando nenhum animal ocorre, e a segunda é quando os animais ocorrem em diferentes níveis de abundância. Se o problema for modelar apenas o primeiro estado, então uma regressão logística tradicional pode ser aplicada normalmente (Welsh *et al.*, 1996). Dado que houve captura, o número de animais observados (capturados) pode ser modelado através de uma distribuição discreta “truncada”, como a Poisson ou a binomial negativa (Grogger & Carson, 1991). O modelo completo, considerando os dois estados, é chamado de modelo condicional.

Os modelos de mistura e condicionais foram utilizados neste trabalho para o ajuste das distribuições de probabilidade para a variável “captura” dos agulhões branco e negro. Maiores detalhes sobre como estes modelos foram implementados estão descritos no item 3.2.2 deste documento.

Modelos como os mencionados acima têm sido utilizados no estudo de dados com excesso de zeros em diversas áreas como Econometria (Grogger & Carson, 1991; Mullahy, 1997) e Agricultura (Borgatto *et al.*, 2006; Sileshi, 2008). No entanto, a aplicação de tais modelos na área de Ecologia é pouco frequente (Welsh *et al.*, 2000; Martin *et al.*, 2005; Potts & Elith, 2006), e ainda mais rara quando se trata especificamente de dados pesqueiros (Fletcher *et al.*, 2005). Dessa forma, os resultados deste trabalho, pode também contribuir para a difusão da utilização potencial desta abordagem de grande aplicação na área de Ecologia.

## 2 Objetivos

O objetivo geral é estimar parâmetros das distribuições de probabilidade para a captura de espécies pouco frequentes em pescarias comerciais. Em particular, será estudado o caso dos agulhões branco e negro capturados pelas frotas de espinhel de superfície que atuam no oceano Atlântico sul. A análise deste problema tem como meta contribuir tanto para a conservação quanto para o manejo das populações de agulhões.

De maneira mais específica, os objetivos são:

- Estimar parâmetros das distribuições de probabilidade para a captura dos agulhões branco e negro no Atlântico sul, por área e trimestre;
- Identificar a relação entre as distribuições de probabilidade e fatores explanatórios como área e trimestre;
- Estimar o número médio de agulhões capturados e desembarcados ao longo das últimas décadas, permitindo uma avaliação aproximada das tendências da situação dos estoques;
- Gerar mapas que representem temporalmente (*e.g.* trimestres) e espacialmente as probabilidades de captura diferente de zero para cada espécie de agulhão;





## 3 Materiais e Métodos

### 3.1 Dados

Os países signatários têm o direito de participar de todo o processo de debate e deliberação que ocorre no âmbito da ICCAT. Em contrapartida é requisitado que todas as nações reportem anualmente informações sobre as diversas pescarias de tunídeos que ocorrem sob sua jurisdição. Neste trabalho foram utilizadas estas informações armazenadas no banco de dados da ICCAT, as quais podem ser acessadas via consulta pública no *site* [www.iccat.int](http://www.iccat.int). O banco de dados específico que é de interesse primordial neste projeto é referenciado na ICCAT como sendo o oriundo dos formulários da “Tarefa 2” (ICCAT *Task 2 Database*). No caso das pescarias com espinhel pelágico, constam os seguintes itens em cada entrada do banco de dados:

- A quantidade (em peso e/ou número) de peixes capturados e desembarcados de cada espécie ou categoria de pescado;
- O esforço de pesca utilizado para obter a captura, usualmente expresso em número de anzóis;
- O ano, o trimestre ou o mês em que ocorreu a captura;
- Uma identificação da frota (país) a que pertence(m) o(s) barco(s) que realizou(aram) a captura. No caso das frotas arrendadas, é indicada qual é a bandeira do país de origem;
- A posição geográfica, usualmente com uma resolução de 5° de latitude por 5° de longitude.

Uma categoria de pescado, como citado no primeiro item listado acima, pode conter espécies de vários gêneros e famílias. Por exemplo, há a categoria “outros tubarões” onde são incorporadas as capturas de uma série de elasmobrânquios de interesse comercial secundário para os pescadores.

É importante ressaltar que as entradas do banco de dados não correspondem necessariamente ao registro de um único lance de pesca. Em muitas situações são reportados os resultados agregados de dois ou mais lances, e

mesmo de duas ou mais embarcações, desde que elas pertençam à mesma frota e que os lances tenham sido realizados na mesma área e período.

A princípio, foram selecionadas as entradas correspondentes a todas as frotas que atuaram sobre o estoque sul dos agulhões branco e negro (latitudes ao sul de 5° N). A base de dados utilizada foi atualizada pela última vez em novembro de 2010, de acordo com o próprio site da ICCAT. Através de uma breve análise do número de informações por ano, trimestre e frota, constatou-se que apenas a frota do Japão apresenta dados suficientes para uma análise temporal da captura dos agulhões. As demais frotas apresentaram informações dispersas ao longo do tempo, e muitas vezes com longos intervalos anuais sem dados. Por esse motivo, optou-se por utilizar os dados referentes apenas à frota do Japão, a partir do ano de 1970 até 2009. Na Tabela 1 consta o número de informações por ano e trimestre utilizada nas análises.

Tabela 1: Número de registros presente na base de dados da ICCAT por ano e trimestre, referentes à pescarias de espinhel de superfície realizados no oceano Atlântico sul pela frota japonesa. Os valores são referentes à base de dados atualizada pela última vez no *site* da ICCAT em novembro de 2010.

Trimestre					Trimestre				
Ano	1	2	3	4	Ano	1	2	3	4
1970	141	99	70	88	1990	74	127	114	97
1971	115	104	64	69	1991	95	146	136	108
1972	110	84	63	68	1992	94	95	93	76
1973	81	68	56	58	1993	79	120	115	84
1974	41	31	32	36	1994	82	147	137	95
1975	42	34	66	60	1995	88	128	112	104
1976	38	16	22	34	1996	86	114	113	107
1977	29	44	32	31	1997	78	105	107	98
1978	39	36	56	20	1998	66	102	91	95
1979	43	43	55	56	1999	74	103	93	77
1980	72	88	64	57	2000	70	96	105	114
1981	98	94	60	56	2001	85	92	66	87
1982	98	91	57	79	2002	70	57	66	68
1983	70	25	34	85	2003	79	103	97	88
1984	62	45	65	77	2004	89	91	71	82
1985	67	98	111	89	2005	64	84	87	81
1986	95	72	81	77	2006	63	97	102	85
1987	67	70	62	53	2007	59	70	83	74
1988	68	110	103	107	2008	83	106	111	97
1989	82	106	111	102	2009	86	116	104	94

## 3.2 Análise

### 3.2.1 Análise descritiva e exploratória

Nesta etapa inicial, foram calculados sumários estatísticos (*e.g.* médias, medianas, quartis) e construídos elementos básicos de uma análise exploratória dentro do contexto clássico definido por Tukey (1977). Esta análise inclui, por exemplo, a confecção de diagramas de caixa e distribuições de frequência para o número de indivíduos capturados e o esforço de pesca (número de anzóis). Estratificações envolvendo ano, trimestre, área de pesca e tipo de frota foram consideradas.

Padrões de associação entre a variável resposta (número de agulhões capturados) e covariáveis (variáveis quantitativas) e fatores (variáveis qualitativas) foram avaliados através de gráficos de dispersão e diagramas de caixa, respectivamente.

A Captura Por Unidade de Esforço (CPUE) também foi calculada para uma análise descritiva dos padrões temporais e espaciais das capturas. A CPUE é definida como

$$CPUE = \frac{\sum C}{\sum f}$$

onde  $C$  é a captura e  $f$  é o esforço. Sob condições adequadas, a CPUE pode ser considerada como proporcional à abundância de um estoque (Hilborn & Walters, 1992).

### 3.2.2 Formulação dos modelos

Os modelos de regressão para dados de contagem podem ser classificados como um tipo particular da classe de Modelos Lineares Generalizados, uma abordagem abrangente para diversos tipos de regressão, proposta inicialmente por Nelder & Wedderburn (1972). De maneira geral, os MLGs descrevem a relação entre a variável resposta  $y_i$  ( $i = 1, \dots, n$ ) através de preditores  $x_i$ . A distribuição condicional de  $y_i|x_i$  deve pertencer à família exponencial, com função dada por

$$f(y; \theta, \phi) = \exp\{(y\theta - b(\theta))/a(\phi) + c(y, \phi)\} \quad (1)$$

para funções  $a(\cdot)$ ,  $b(\cdot)$  e  $c(\cdot)$  conhecidas. O parâmetro de dispersão,  $\phi$ , é geralmente conhecido, e quando isso ocorre,  $\theta$  é o parâmetro canônico da função acima (McCullagh & Nelder, 1989).

Se esta condição for satisfeita, então a média de  $y_i$  condicionada aos preditores  $x_i$  é  $E(y_i|x_i) = \mu_i$ , e existe uma transformação de  $\mu_i$  de forma que

$$g(\mu_i) = x_i^T \beta \quad (2)$$

onde  $g(\cdot)$  é uma função de ligação conhecida, e  $\beta$  é o vetor de parâmetros a ser estimado. A variância de  $y_i$  é então dada por  $Var(y_i) = \phi V(\mu_i)$ , onde  $\phi$  é o parâmetro de dispersão (geralmente constante) e  $V(\mu_i)$  é a função variância (McCullagh & Nelder, 1989). Os modelos de contagem utilizados nesta análise são definidos a seguir.

**Modelo de Poisson** A distribuição de Poisson é dada pela função de massa de probabilidade (f.m.p.)

$$f(y; \mu) = \frac{e^{-\mu} \mu^y}{y!} \quad (3)$$

onde o parâmetro  $\mu$  representa o número médio de ocorrências em um intervalo (Dobson, 2002). Uma propriedade desta distribuição é que  $E(y) = Var(y) = \mu$ , e, por isso, o parâmetro de dispersão é fixo em  $\phi = 1$ . Agora suponha que  $y_i$  seja o número de ocorrências de um evento para um dado número de “exposições”  $n_i$ . Nesse caso, a esperança de  $y_i$  pode ser escrita como  $E(y_i) = \mu_i = n_i \theta_i$ . Se  $\theta_i$  depende de variáveis explicativas, então pode ser modelado através de  $\theta_i = e^{x_i^T \beta}$ . Portanto, se  $y_i \sim \text{Poisson}(\mu_i)$ , o MLG é dado por  $E(y_i) = \mu_i = n_i e^{x_i^T \beta}$ . Como a função de ligação canônica é a logarítmica, o modelo resultante fica

$$\log \mu_i = \log n_i + x_i^T \beta$$

Note que este modelo difere da especificação tradicional do componente linear  $x_i^T \beta$ , devido à inclusão do termo  $\log n_i$  (Dobson, 2002). Este termo (chamado de *offset*) é uma “compensação” nos casos onde a variável resposta é

observada em intervalos de comprimento conhecido (portanto fixo), porém não necessariamente constante.

**Modelo quase-Poisson** Uma maneira de lidar com a superdispersão no modelo de Poisson é assumir que o parâmetro  $\phi$  não seja fixo em 1, mas permitir que ele seja irrestrito e estimado a partir dos próprios dados (Zeileis *et al.*, 2008). O modelo de quase-Poisson obtêm as mesmas estimativas dos coeficientes que o modelo de Poisson, porém elas são ajustadas para “acomodar” a superdispersão. Como a distribuição da variável resposta para esse modelo depende apenas das funções média e variância (Venables & Ripley, 2002), ele não possui verossimilhança totalmente especificada.

**Modelo binomial negativo** A distribuição binomial negativa surge a partir de um modelo de dois estágios para a variável discreta  $y$ , no qual assume-se que  $y|z \sim \text{Poisson}(z)$  e  $Z \sim \text{gama}(\mu, \theta)$ , onde  $\theta$  é independente de  $\mu$ . A combinação destas distribuições é a BN com f.m.p. dada por

$$f(y; \mu, \theta) = \binom{y + \theta - 1}{y} \left( \frac{\mu}{\mu + \theta} \right)^y \left( \frac{\theta}{\mu + \theta} \right)^\theta \quad (4)$$

A esperança desta distribuição é  $E(y) = \mu$  e a variância é dada por  $\text{Var}(y) = \mu + \mu^2/\theta$  (Venables & Ripley, 2002). Note que, ao contrário da distribuição de Poisson, aqui a variância é maior do que a média e  $\theta \geq 0$  é frequentemente chamado de parâmetro de superdispersão. Devido à inclusão desse parâmetro, esta distribuição tem uma maior capacidade de lidar com dados superdispersos.

**Modelos de mistura** Para o caso da utilização da distribuição de Poisson, e sendo  $y_i, i = 1, \dots, n$ , o vetor da variável resposta em forma de contagens (e.g. número de agulhões capturados) para um evento de pesca  $j$ , a definição do modelo é dada por

$$\begin{aligned} y_i = 0 & \quad \text{com probabilidade } 1 - p(x) \text{ e} \\ y_i = \text{Poisson}(\lambda(z)) & \quad \text{com probabilidade } p(x) \end{aligned}$$

de forma que

$$\begin{aligned} P(Y = 0|x, z) &= 1 - p(x) + p(x)e^{-\lambda(z)} \\ P(Y = r|x, z) &= \frac{p(x)e^{-\lambda(z)}\lambda(z)^r}{r!}, \quad r = 1, 2, \dots \end{aligned} \quad (5)$$

Aqui,  $p(x)$  representa a probabilidade de que o número de animais em um dado cenário (*e.g.* área e trimestre) seja proveniente de uma distribuição de Poisson, e dado que este número segue uma distribuição de Poisson,  $\lambda(z)$  é o número médio de animais esperado para aquele dado cenário. As modificações apropriadas para o caso da consideração de uma distribuição binomial negativa são triviais, e maiores detalhes podem ser encontrados em Lambert (1992). A distribuição binomial negativa deve ser considerada quando o excesso de zeros causa uma superdispersão muito grande em relação à distribuição de Poisson (Welsh *et al.*, 2000). Ridout *et al.* (2001) apresentam uma metodologia para o cálculo de estatísticas de teste para a comparação entre modelos inflacionados de zero ajustados utilizando as distribuições de Poisson e binomial negativa. Quando o modelo depende de covariáveis (*e.g.* trimestre, área), é necessário o uso de funções de ligação entre  $p(x)$  e  $\lambda(z)$  e estas covariáveis. Se considerarmos as funções logística e log-linear usuais teremos

$$\text{logit}(p(x)) = \log\left(\frac{p(x)}{1-p(x)}\right) = x^T \beta \quad \text{e} \quad \log(\lambda(z)) = z^T \alpha \quad (6)$$

onde  $x$  e  $z$  são as matrizes de covariáveis, e  $\beta$  e  $\alpha$  são vetores de parâmetros a serem estimados. As matrizes de covariáveis podem ou não serem coincidentes. Caso elas sejam, o modelo pode ser desenvolvido sob a suposição de que os dois preditores lineares ( $x^T \beta$  e  $z^T \alpha$ ) são relacionados de alguma forma (Ridout *et al.*, 1998). Isto é, escrevendo  $p$  como uma função de  $\lambda$ , o número de parâmetros pode ser reduzido e o modelo pode ser simplificado. Portanto, a variável resposta pode ser modelada através de

$$\text{logit}(p(x)) = \log\left(\frac{p(x)}{1-p(x)}\right) = \tau z^T \alpha \quad \text{e} \quad \log(\lambda(z)) = z^T \alpha \quad (7)$$

onde  $\tau$  é um parâmetro escalar, o que implica em  $p(x) = (1 + \lambda(z)^\tau)^{-1}$  (Lambert, 1992). Na terminologia de Modelos Lineares Generalizados,  $\text{logit}(p(x))$  e  $\log(\lambda(z))$  são as funções de ligação canônicas, ou as transformações que linearizam as probabilidades de sucesso da distribuição de Bernoulli, e as médias da distribuição de Poisson, respectivamente. É importante ressaltar que para o modelo de mistura, os parâmetros estão relacionados conjuntamente à probabilidade de encontrar um animal e ao número médio de animais. Portanto, a interpretação destes parâmetros não é direta. Por exemplo, para saber a estimativa da captura média esperada em um determinado cenário, deve-se multiplicar a estimativa da média  $\lambda$  pela probabilidade  $p(x)$  (Martin *et al.*, 2005).

**Modelos condicionais** No caso da utilização de uma distribuição de Poisson, e sendo  $y_i, i = 1, \dots, n$ , como definido na abordagem anterior, a definição do modelo é dada por

$$\begin{aligned} y_i = 0 & && \text{com probabilidade } 1 - p(x) \text{ e} \\ y_i = \text{Poisson}(\lambda(z)) \text{ truncada} & && \text{com probabilidade } p(x) \end{aligned}$$

de forma que

$$\begin{aligned} P(Y = 0|x) &= 1 - p(x) \\ P(Y = r|x, z) &= \frac{p(x)e^{-\lambda(z)}\lambda(z)^r}{r!(1 - e^{-\lambda(z)})}, \quad r = 1, 2, \dots \end{aligned} \quad (8)$$

Neste caso,  $p(x)$  é a probabilidade de se observar pelo menos um animal (*i.e.* capturar pelo menos um exemplar) em um dado cenário, e dado que existe pelo menos um,  $\lambda(z)$  é o parâmetro da distribuição de Poisson truncada que descreve o número de animais que observamos. A adaptação deste modelo para o caso da utilização da distribuição binomial negativa pode ser encontrada em Welsh *et al.* (1996). As funções de ligação e a formulação do modelo considerando covariáveis para  $p(x)$  e  $\lambda(z)$  são as mesmas demonstradas em (6). Os modelos condicionais possuem uma vantagem em relação



aos modelos de mistura, no sentido de que os parâmetros  $\beta$  são interpretados diretamente em termos de  $P(Y = 0|x)$ , e os parâmetros  $\alpha$  representam separadamente a abundância média, dado que pelo menos um animal ocorreu (Welsh *et al.*, 1996).

### 3.2.3 Definição de variáveis e seleção de modelos

Para cada uma das espécies foram ajustados os sete modelos descritos acima, incluindo as devidas modificações no caso da distribuição BN para os modelos de mistura e condicional (Poisson, quase-Poisson, BN, mistura Poisson e BN, condicional Poisson e BN). A variável resposta é a captura, e a matriz de covariáveis é: ano (fator; 1970–2009), trimestre (fator; 1–4) e área (fator; NW, NC, NE, CW, CC, CE, SW, SC, SE). Para os modelos de mistura e condicionais, foi utilizado  $x = z$ , ou seja, a mesma matriz do modelo (matriz com as variáveis explicativas) foi utilizada tanto para a modelagem das contagens, quanto para a modelagem dos zeros. A definição das áreas foi feita com base no tamanho e na quantidade de informações disponíveis em cada região. A intenção foi dividir o oceano Atlântico Sul considerando três áreas latitudinais: norte, central e sul, e três áreas longitudinais: leste, central e oeste. A divisão foi realizada com a intenção de se obter o melhor balanço entre número de informações e tamanho da área por ano e trimestre. Na Figura 3 está ilustrada a divisão e as siglas das áreas utilizadas como fator no modelo.

Com estas covariáveis fixas, as diferenças entre as estimativas dos parâmetros, e por consequência das predições, deve ser resultado apenas da especificação de cada modelo. Convém ressaltar que em todos eles, o *offset* com o logaritmo do número de anzóis foi utilizado para compensar os valores agregados de captura.

Devido ao fato de que os termos de MLGs são geralmente não-ortogonais, a ordem de entrada das variáveis explicativas em um modelo afeta o resultado da contribuição de cada variável para o modelo final (McCullagh & Nelder, 1989). Para se definir a ordem de entrada nos modelos testados, cada variável explicativa foi ajustada sozinha, com a mesma variável resposta, e o critério de informação de Akaike (AIC) (Akaike, 1974) foi calculado para cada mo-

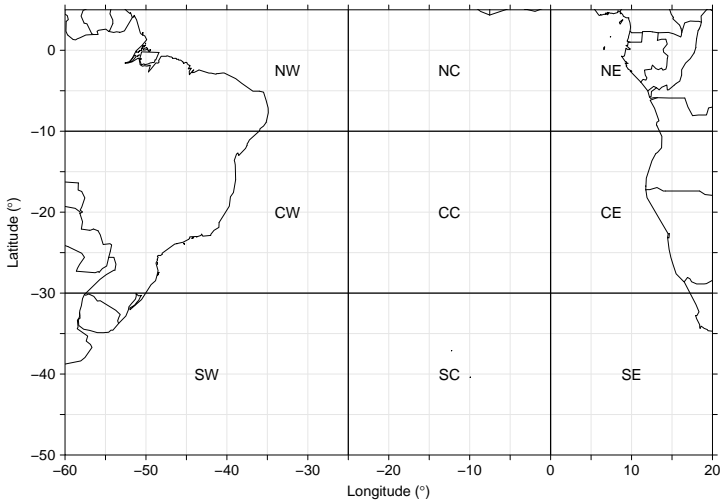


Figura 3: Divisão e nome das áreas no oceano Atlântico Sul, utilizadas para a modelagem das capturas dos agulhões branco e negro.

delo. Com estes resultados, os modelos foram classificados do menor para o maior AIC, classificando, portanto, as variáveis em termos de representatividade (*e.g.* redução individual na deviança) em relação à variável resposta. Esta ordem possivelmente representa a alternativa mais apropriada para a definição da entrada das variáveis (sequencialmente) nos modelos.

A comparação entre os modelos ajustados para cada espécie foi realizada através dos critérios de informação de Akaike (AIC) e de Bayes (BIC) (Schwarz, 1978). À título de ilustração, o número de zeros predito por cada modelo também foi estimado para uma comparação com o número de zeros observado. O modelo que apresentou o menor valor de AIC/BIC foi selecionado para continuidade das análises.

Após a comparação e a seleção do modelo mais adequado para cada espécie, interações de primeira ordem foram incluídas uma a uma ao modelo principal (método “passo à frente” – *step forward*), desde que promovessem

uma mudança significativa em termos de AIC e na redução na deviência. Esta significância foi determinada através de um teste de razão de verossimilhança (TRV) entre o modelo inicial (com todos os efeitos principais), e este modelo mais cada interação candidata inserida. O TRV utiliza duas vezes a diferença do logaritmo da verossimilhança de dois modelos ( $M_1$  e  $M_2$ ), e compara esse resultado com uma distribuição  $\chi^2$  com graus de liberdade definido pela diferença do número de parâmetros dos dois modelos, sob a hipótese nula  $H_0 : M_1 = M_2$ . Se a adição da interação reduziu a deviência significativamente (e.g.  $p < 0,05$ ), então ela foi incluída no modelo, e o procedimento se repetia para a próxima interação. Maiores detalhes sobre este procedimento utilizado aqui podem ser encontrados em Venables & Ripley (2002). Convém ressaltar que este procedimento não foi realizado para todos os modelos sob avaliação, apenas para simplificação da comparação entre modelos.

### 3.2.4 Diagnósticos

Os diagnósticos para MLGs consistem de uma série de análises complementares, como a análise dos resíduos do modelo, a verificação de pontos de alavanca e de pontos influentes.

Os resíduos ordinais expressam a diferença entre os valores observados  $y_i$ , e os valores ajustados pelo modelo  $\hat{\mu}_i$ . Portanto, cada resíduo  $r_i$  pode ser definido como

$$r_i = \text{sinal}(y_i - \hat{\mu}_i)$$

Quando se trata de MLGs onde a teoria dos modelos lineares clássicos pode ser simplesmente estendida (e.g. MLGs com erro normal e função de ligação identidade), as propriedades do comportamento dos resíduos são bem conhecidas. Nestes casos, os resíduos devem ser independentes, e apresentar uma distribuição próxima da normal padrão  $N(0, \sigma^2)$ , com média zero e variância constante.

Considerando-se MLGs com distribuições não-lineares, e com dados superdispersos, esta medida não possui as mesmas propriedades, e deve ser interpretada com mais cautela. No entanto, a utilização dos resíduos para diagnóstico nestes casos ainda é válida, desde que seja considerado o padrão

de distribuição, e não os valores absolutos dos resíduos (Hinde & Demétrio, 1998). Portanto, neste trabalho foram analisados os resíduos ordinais como uma forma de verificação do ajuste do modelo.

Hinde & Demétrio (1998) citam que uma técnica mais adequada para se examinar os resíduos é usando um gráfico de probabilidade normal, com a inclusão de um envelope simulado, que leva em conta a superdispersão. Este envelope serve como “bandas de confiança” por onde os resíduos plotados contra escores normais devem permanecer. Uma tentativa de construção desse gráfico com envelope simulado foi também realizada, baseada no algoritmo apresentado por Paula (2004, pgs. 46 e 230).

A construção de envelopes simulados consiste, em uma de suas etapas, na padronização dos resíduos ordinais, dada por

$$r_i^* = \frac{r_i}{\sqrt{(1-h_i)}}$$

onde  $h_i$  são os componentes do vetor  $\mathbf{h} = \text{diag}(\mathbf{H})$ . A matriz  $\mathbf{H}$  é a matriz de projeção, que mede a influência das mudanças em  $y_i$  dadas por  $\hat{\mu}_i$  em medidas padronizadas. Esta matriz é definida pela teoria dos modelos lineares clássicos, e sua extensão para MLGs é dada por

$$\mathbf{H} = \mathbf{W}^{\frac{1}{2}} \mathbf{X} (\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W}^{\frac{1}{2}}$$

onde  $\mathbf{X}$  é a matriz do modelo, e  $\mathbf{W}$  a matriz de pesos (McCullagh & Nelder, 1989). Uma propriedade interessante sobre essa matriz é que  $\text{posto}(\mathbf{H}) = \text{tr}(\mathbf{H}) = \sum_i h_{ii} = p$ , ou seja, o somatório dos elementos da diagonal principal se iguala ao número de parâmetros  $p$  (Paula, 2004). Esta matriz tem um papel fundamental para os diagnósticos em MLGs. Além de padronização dos resíduos, os elementos  $h_i$  também servem como uma medida da influência da  $i$ -ésima observação sobre o próprio valor ajustado.

### 3.2.5 Estimativas dos parâmetros

Os modelos descritos acima envolvem a especificação da distribuição de probabilidade dos dados de contagens. Nesses casos, os métodos de máxima

verossimilhança são apropriados para a estimativa dos parâmetros dos modelos (Ridout *et al.*, 1998).

No caso específico dos modelos de mistura, o algoritmo EM (*Expectation-Maximization*) pode ser utilizado para maximizar a verossimilhança. Uma das etapas desse algoritmo consiste em ajustar dois MLGs aos dados: uma regressão logística e uma regressão log-linear de Poisson ou BN. Detalhes específicos podem ser encontrados em Lambert (1992). Os parâmetros para os modelos condicionais são relativamente mais simples de serem estimados. A estimativa dos parâmetros pode ser feita através do tradicional algoritmo de Mínimos Quadrados Reponderados Iterativamente (MQRI), utilizado normalmente em MLGs. Nesse caso, a primeira parte do modelo pode ser calculada através de um MLG logístico tradicional, e a segunda parte, para o modelo de Poisson ou BN truncado, uma adaptação da rotina mencionada pode ser utilizada. Detalhes desta implementação podem ser encontradas em Welsh *et al.* (1996). As referências básicas que contemplam o uso de MLGs em geral são Nelder & Wedderburn (1972), McCullagh & Nelder (1989), e Dobson (2002).

Todas as análises foram realizadas através do *software* R 2.12.1 (R Development Core Team, 2010), com auxílio dos pacotes MASS (Venables & Ripley, 2002) para o ajuste dos modelos com a distribuição binomial negativa, e pscl (Zeileis *et al.*, 2008), para o ajuste dos modelos condicionais e de mistura. Os mapas foram construídos com adaptações de funções do pacote Lattice (Sarkar, 2008).



## 4 Resultados

### 4.1 Esforço de pesca

A frota japonesa de espinhel de superfície atuou praticamente em todas as regiões do oceano Atlântico durante o período de 1970–2009 (Figura 4). Durante o primeiro trimestre o esforço foi concentrado na região nordeste do oceano. No segundo trimestre, o esforço foi direcionado para águas mais ao sul. Já no terceiro trimestre houve uma menor concentração de esforço em áreas específicas, mas a área central do oceano, mais próxima à África apresentou uma maior densidade de pesca. Finalmente, no quarto trimestre, a mesma região citada para o trimestre anterior foi a que recebeu os maiores esforços.

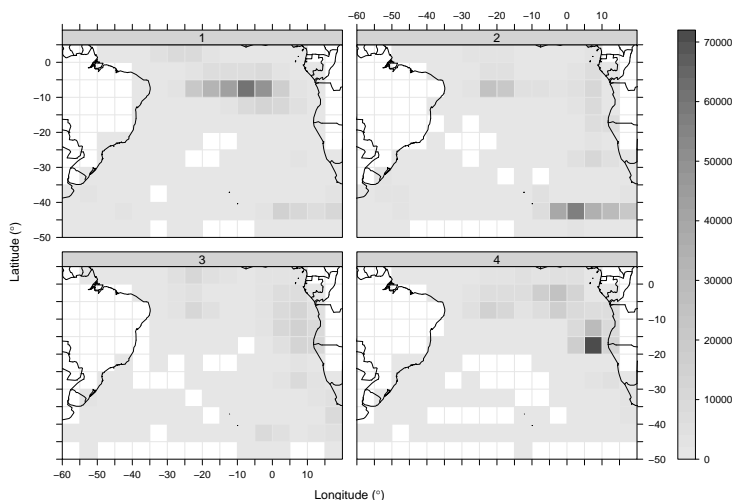


Figura 4: Esforço de pesca (número de anzóis  $\times$  1000) realizado pela frota japonesa de espinhel de superfície no oceano Atlântico, somado por trimestre para o período 1970–2009.

## 4.2 Agulhão branco

### 4.2.1 Análise descritiva

Considerando a Captura Por Unidade de Esforço (CPUE) trimestral do agulhão branco pela frota japonesa, percebe-se que maiores valores estão localizados à oeste do oceano Atlântico, próximos, inclusive, da costa do Brasil. Estes valores elevados são particularmente mais visíveis no primeiro e no quarto trimestres, ou seja, no verão austral. No segundo e no terceiro trimestres não existe um padrão claro da distribuição da CPUE. Nota-se também que as áreas em que a captura é zero estão localizadas mais ao sul do oceano, e algumas vezes também à leste/sudeste (Figura 5).

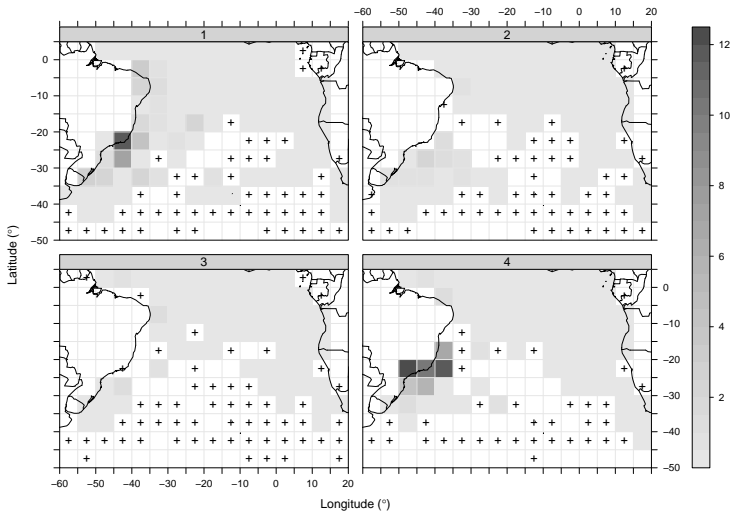


Figura 5: Captura por Unidade de Esforço (número de peixes capturados por lance com 1000 anzóis) trimestral do agulhão branco, obtida pela frota japonesa de espinhel de superfície no oceano Atlântico. Os quadrantes marcados com o símbolo + indicam o valor zero.

De fato, quando observada a distribuição das CPUEs em cada área (cf. Figura 3), fica clara a evidência de que as áreas CW (centro-oeste) e NW



(norte-oeste) apresentam os maiores valores. A área SW (sul-oeste) também aparece com CPUEs elevadas no primeiro e no quarto trimestres. As demais áreas apresentam uma grande variabilidade nos valores de CPUE, com a maior parte dos valores mais próximos de zero (Figura 6).

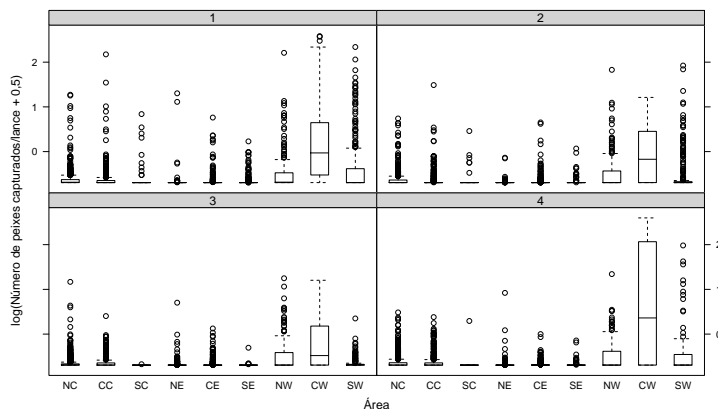


Figura 6: Distribuição da Captura por Unidade de Esforço (número de peixes capturados por lance com 1000 anzóis) trimestral do agulhão branco, obtida pela frota japonesa de espinhel de superfície em cada área no oceano Atlântico, entre 1970–2009. Os valores foram transformados para a escala  $\log(x + 0,5)$  para facilitar a visualização.

Quando analisada a série temporal da CPUE do agulhão branco por trimestre, duas situações ficam evidentes: (a) o primeiro e o quarto trimestre se destacam em praticamente todos os anos como os que apresentam as maiores CPUEs; (b) os valores do início da série, entre 1970 e 1972, claramente são superiores aos demais valores anuais. Além disso, CPUEs elevadas foram obtidas no período entre 1986 e 1991. No final da série temporal, a partir de 2001, parece haver uma tendência de declínio nas CPUEs (Figura 7).

#### 4.2.2 Modelagem

A proporção de zeros encontrada para a captura do agulhão branco no Atlântico Sul pela frota de barcos japoneses de espinhel de superfície foi de 70%. Os valores de captura variaram no intervalo  $[0, 843]$  (Figura 8). A média de captura ponderada pelo esforço foi de cerca de  $\bar{x} = 11$  peixes, com

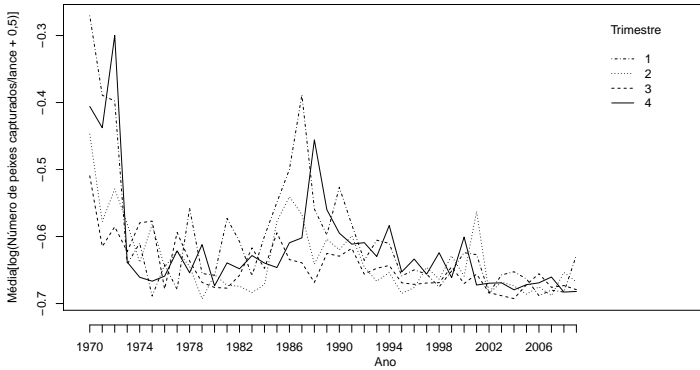


Figura 7: Média da Captura por Unidade de Esforço (número de peixes capturados por lance com 1000 anzóis) trimestral do agulhão branco, obtida pela frota japonesa de espinhel de superfície no oceano Atlântico, entre 1970–2009. Os valores foram transformados para a escala  $\log(x + 0,5)$  para facilitar a visualização.

variância  $s^2 = 490$  para o período analisado (1970–2009). Com isto, pode-se constatar que a variância é cerca de 43 vezes maior do que a média, o que indica uma superdispersão considerável nos dados.

Entre os modelos ajustados, o de Poisson foi o que apresentou os maiores valores de AIC e BIC, indicando o pior ajuste (Tabela 2). Os modelos de mistura e condicional ajustados com a distribuição de Poisson para os valores positivos tiveram resultados similares de acordo com os critérios de informação, e apresentaram um ganho pequeno em relação ao modelo de Poisson tradicional. Entre os modelos ajustados com a distribuição de Poisson, o modelo condicional foi o que apresentou o menor AIC/BIC. Os modelos ajustados com a distribuição BN (tradicional, de mistura e condicional) apresentaram uma grande redução nos valores de AIC/BIC em relação aos modelos anteriores (ajustados com a distribuição de Poisson), o que indica um melhor ajuste. De acordo com esses critérios, o modelo condicional BN foi o que apresentou o melhor ajuste entre todos (menor AIC/BIC). O número de zeros predito pelos modelos condicionais (Poisson e BN) foi exatamente igual ao número de zeros observado (*i.e.* 9037), fato evidente devido à própria formulação destes

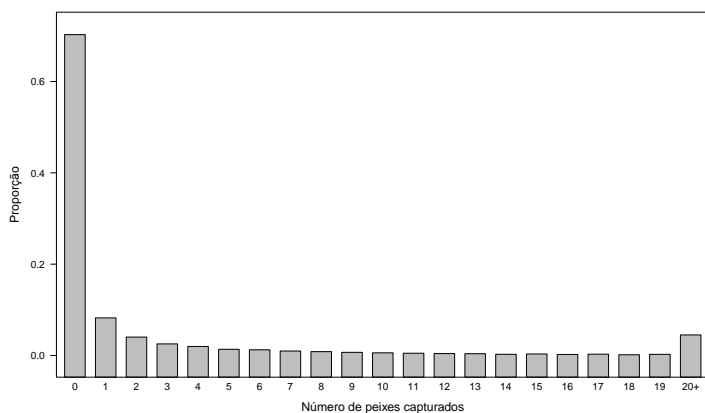


Figura 8: Proporção do número de agulhões branco capturados no Atlântico Sul no período 1970–2009. Como os valores de captura variaram no intervalo  $[0, 843]$ , aqueles maiores que 19 foram agregados na classe 20+.

modelos. O modelo de mistura BN, foi o segundo que se aproximou do número de zeros observado. O modelo de Poisson subestimou substancialmente o número de zeros.

Tabela 2: Número de parâmetros, logaritmo da versossimilhança ( $\log L$ ), critério de informação de Akaike (AIC), critério de informação bayesiano (BIC), e número de zeros preditos (o número de zeros observado foi de 9037) pelos diferentes modelos ajustados para a captura do agulhão branco.

	Poisson	Quase-Poisson	BN	Poisson (mistura)	BN (mistura)	Poisson (cond.)	BN (cond.)
N. parâm.	51	51	52	102	103	102	103
$\log L$	-55568	—	-17203	-47296	-16733	-45900	-16659
AIC	111238	—	34510	94797	33672	92004	33524
BIC	111619	—	34898	95558	34440	92765	34293
N. zeros	6579	—	8953	8272	8945	9037	9037

Através dos resultados da comparação entre modelos feita acima, notou-se portanto, que o “melhor” modelo, ou o modelo que melhor descreve os dados de captura do agulhão branco no Atlântico sul, foi o condicional utilizando a distribuição binomial negativa. Com esse modelo escolhido foi reali-

zado então o processo de inclusão de interações ao modelo com efeitos principais. Através de inclusão de cada interação possível (ano:área, ano:trimestre, e área:trimestre), constatou-se que apenas a interação área:trimestre foi significativa para o modelo (TRV:  $p < 2,2E - 16$ ). O algoritmo utilizado para a maximização da verossimilhança do modelo principal ajustado individualmente com cada uma das demais interações não atingiu a convergência necessária para o cálculo de todos os parâmetros, alegando que o sistema era exatamente singular<sup>1</sup>.

Os resultados das estimativas dos parâmetros, tanto da parte dos modelos ajustada com a distribuição BN quanto com a regressão logística (distribuição binomial), e seus respectivos erros e  $p$ -valores associados são apresentados na Tabela 3. De maneira geral, as estimativas dos coeficientes para ambas as partes foram significativamente diferentes de zero, como pode ser constatado pelos  $p$ -valores baixos (*e.g.*  $\Pr(>|z|) < 0,05$ ). Através do modelo logístico, percebe-se que as probabilidades de captura são influenciadas negativamente pelas áreas, ou seja, com exceção das áreas NW e CW, as demais áreas apresentaram probabilidades menores se comparadas com a área de referência NC. De fato, as áreas NW e CW apresentaram médias de captura maiores, como constatado pelos coeficientes do modelo para as contagens. Observando-se os coeficientes estimados por ano, percebe-se uma diminuição tanto na probabilidade quanto na média de captura do agulhão branco, quando comparados ao coeficiente estimado para 1970 (primeiro ano da série). Quanto aos trimestres, o segundo apresentou um aumento significativo nas médias de captura, enquanto que no terceiro trimestre a probabilidade de captura foi menor.

Através da parte de contagem do modelo BN, foi estimado também o parâmetro de superdispersão ( $\theta$ ) da distribuição binomial negativa (cf. equação 4). A estimativa para  $\theta$  foi  $\log(\theta) = -1,31 \Rightarrow \theta = 0,27$  (Tabela 3). O valor é significativamente diferente de zero, o que indica uma real superdispersão dos dados utilizados.

---

<sup>1</sup>Nesse caso, a matriz Hessiana foi identificada como singular, ou seja, tem determinante zero e não possui inversa. A inversa da matriz Hessiana é utilizada em uma das etapas do processo de maximização da verossimilhança, para a derivação numérica dos erros-padrão dos coeficientes.

Tabela 3: Estimativas dos parâmetros do modelo condicional ajustado com a distribuição binomial negativa para a captura do agulhão branco. As colunas “Contagem” se referem aos coeficientes do modelo BN para as contagens. As colunas “Zero” representam os coeficientes para o modelo logístico considerando a presença/ausência de zeros.

Termos	Contagem			Zero		
	Estimativa	EP	Pr(> z )	Estimativa	EP	Pr(> z )
(Intercepto)	-9,43	0,172	0,00	-9,19	0,169	0,00
ÁreaCC	0,405	0,158	0,0106	-0,0828	0,139	0,552
ÁreaSC	1,92	0,655	0,00334	-4,19	0,373	3,35E-29
ÁreaNE	0,413	0,473	0,383	-1,70	0,347	1,04E-06
ÁreaCE	0,0915	0,297	0,758	-1,07	0,206	2,11E-07
ÁreaSE	-0,460	0,307	0,134	-2,57	0,226	7,07E-30
ÁreaNW	1,35	0,166	3,83E-16	0,818	0,164	6,03E-07
ÁreaCW	2,70	0,285	3,08E-21	2,43	0,380	1,71E-10
ÁreaSW	3,01	0,214	7,58E-45	-0,200	0,190	0,293
Ano1971	0,868	0,221	8,83E-05	-0,516	0,207	0,0126
Ano1972	0,353	0,226	0,119	-0,478	0,215	0,0261
Ano1973	-0,273	0,251	0,277	-0,537	0,228	0,0182
Ano1974	0,571	0,391	0,144	-0,645	0,302	0,0328
Ano1975	0,435	0,294	0,139	-1,07	0,256	3,19E-05
Ano1976	-1,10	0,495	0,0260	-0,456	0,367	0,214
Ano1977	-0,281	0,359	0,433	-0,871	0,334	0,00902
Ano1978	-0,0832	0,355	0,815	-0,786	0,298	0,00836
Ano1979	-0,928	0,302	0,00211	-1,02	0,278	0,000239
Ano1980	-2,29	0,279	2,17E-16	-1,70	0,238	8,79E-13
Ano1981	-1,09	0,248	1,14E-05	-1,38	0,224	6,58E-10
Ano1982	-0,822	0,247	0,000853	-1,82	0,224	3,66E-16
Ano1983	-0,796	0,283	0,00487	-1,50	0,257	5,15E-09
Ano1984	-1,23	0,236	2,04E-07	-1,22	0,236	2,49E-07
Ano1985	-0,667	0,212	0,00164	-1,11	0,207	8,95E-08
Ano1986	-0,702	0,218	0,00126	-0,959	0,212	5,84E-06
Ano1987	-0,450	0,235	0,0553	-0,840	0,229	0,000251
Ano1988	-1,13	0,201	2,12E-08	-1,12	0,206	4,64E-08
Ano1989	-0,834	0,202	3,56E-05	-1,37	0,204	1,77E-11
Ano1990	-0,570	0,210	0,00656	-1,57	0,204	1,57E-14
Ano1991	-1,12	0,197	1,34E-08	-1,32	0,194	9,32E-12
Ano1992	-1,09	0,220	6,46E-07	-1,78	0,214	1,08E-16
Ano1993	-1,25	0,207	1,37E-09	-1,58	0,208	2,68E-14
Ano1994	-0,980	0,210	2,93E-06	-1,95	0,200	1,43E-22
Ano1995	-1,58	0,221	8,60E-13	-2,32	0,205	1,03E-29
Ano1996	-1,66	0,207	8,81E-16	-1,89	0,200	3,55E-21
Ano1997	-1,57	0,216	4,07E-13	-1,90	0,204	1,36E-20
Ano1998	-1,40	0,243	8,15E-09	-1,91	0,213	3,83E-19
Ano1999	-1,60	0,235	1,15E-11	-1,85	0,215	7,06E-18
Ano2000	-1,36	0,231	4,01E-09	-2,00	0,205	2,17E-22
Ano2001	-0,571	0,268	0,0333	-2,95	0,225	1,83E-39
Ano2002	-2,51	0,274	5,20E-20	-2,99	0,234	1,77E-37
Ano2003	-2,17	0,251	5,42E-18	-2,85	0,215	3,36E-40
Ano2004	-2,00	0,262	2,34E-14	-2,93	0,221	3,98E-40
Ano2005	-1,54	0,314	9,82E-07	-3,21	0,240	1,15E-40

Continua na próxima página ...

... continuação da página anterior.

Termos	Contagem			Zero		
	Estimativa	EP	Pr(> z )	Estimativa	EP	Pr(> z )
Ano2006	-1,79	0,262	9,74E-12	-2,50	0,218	1,53E-30
Ano2007	-2,13	0,249	1,55E-17	-2,52	0,220	2,31E-30
Ano2008	-2,35	0,246	1,13E-21	-2,50	0,208	3,90E-33
Ano2009	-1,93	0,242	1,59E-15	-2,48	0,210	2,70E-32
Trim2	0,481	0,120	6,32E-05	-0,139	0,112	0,214
Trim3	0,213	0,123	0,0829	-0,551	0,112	8,19E-07
Trim4	0,211	0,103	0,0402	-0,0807	0,103	0,433
ÁreaCC:Trim2	0,172	0,280	0,540	-0,487	0,211	0,0214
ÁreaSC:Trim2	-0,955	0,968	0,324	-1,94	0,565	0,000607
ÁreaNE:Trim2	-2,57	0,582	9,97E-06	0,0496	0,423	0,907
ÁreaCE:Trim2	-1,32	0,369	0,000364	-0,495	0,260	0,0565
ÁreaSE:Trim2	-0,224	0,521	0,667	-1,93	0,347	2,70E-08
ÁreaNW:Trim2	-0,250	0,251	0,319	0,304	0,241	0,207
ÁreaCW:Trim2	-0,886	0,506	0,0797	-0,123	0,633	0,846
ÁreaSW:Trim2	-1,28	0,318	5,96E-05	-0,762	0,265	0,00405
ÁreaCC:Trim3	-0,275	0,245	0,261	0,287	0,206	0,164
ÁreaSC:Trim3	-17,3	988	0,986	-1,41	1,09	0,193
ÁreaNE:Trim3	-1,58	0,538	0,00335	0,786	0,394	0,0459
ÁreaCE:Trim3	-0,786	0,360	0,0289	-0,188	0,252	0,456
ÁreaSE:Trim3	0,491	0,818	0,548	-1,49	0,516	0,00379
ÁreaNW:Trim3	0,0343	0,258	0,894	0,835	0,249	0,000795
ÁreaCW:Trim3	-0,676	0,548	0,217	-0,196	0,578	0,735
ÁreaSW:Trim3	-3,24	0,360	2,43E-19	-0,251	0,301	0,405
ÁreaCC:Trim4	-0,343	0,232	0,140	-1,56E-05	0,205	1,00
ÁreaSC:Trim4	-0,683	1,76	0,698	-0,887	1,12	0,427
ÁreaNE:Trim4	-0,600	0,548	0,273	0,779	0,400	0,0516
ÁreaCE:Trim4	-1,32	0,334	8,40E-05	-0,507	0,241	0,0356
ÁreaSE:Trim4	-1,06	0,543	0,0511	0,205	0,349	0,557
ÁreaNW:Trim4	-0,0517	0,244	0,832	0,284	0,236	0,229
ÁreaCW:Trim4	1,22	0,430	0,00442	-0,0747	0,538	0,890
ÁreaSW:Trim4	-0,812	0,407	0,0457	0,509	0,338	0,132
log( $\theta$ )	-1,31	0,0839	2,83E-55	—	—	—

A análise dos resíduos ordinais do modelo condicional BN, demonstra que os valores se distribuem em sua maior parte em torno de zero. No entanto, pode-se observar alguns valores que se afastam muito dessa tendência central (Figura 9). Uma análise mais aprofundada dos resíduos não foi possível pelo fato de que o cálculo da matriz  $\mathbf{H}$ , para a extração do vetor  $\mathbf{h} = \text{diag}(\mathbf{H})$ , não pôde ser concluída por problemas computacionais. O tamanho das matrizes envolvidas no cálculo matricial de  $\mathbf{H}$  tornou o processo inviável de ser realizado em computadores comuns<sup>2</sup>, principalmente pela disponibilidade de memória RAM. Com isso, a construção dos gráficos de probabilidade normal com envelopes simulados, também não pôde ser realizada.

Com os coeficientes do modelo BN para o agulhão branco, foram rea-

<sup>2</sup>Foi utilizado um computador com processador Intel Core2Duo 2.0 GHz, com 3 GB de memória RAM e sistema operacional Linux.

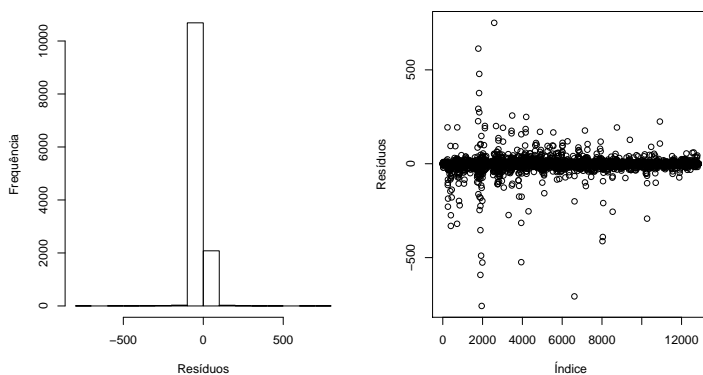


Figura 9: Histograma de frequência (esquerda) e gráfico de dispersão (direita) dos resíduos do modelo condicional BN para o agulhão branco.

lizadas predições para o número de peixes capturados por ano, trimestre e área. Esta predição foi realizada para um número fixo de 1000 anzóis como unidade de esforço. Dessa forma, obteve-se uma Captura Por Unidade de Esforço padronizada, predita pelo modelo. Para se obter a estimativa média anual de CPUE, foi realizada a média entre os valores de trimestre e área para cada ano.

De maneira geral, o modelo parece ter superestimado os valores de CPUE, se comparados aos mesmos valores observados em cada ano. No entanto, as tendências de variação parecem seguir o mesmo padrão. Se comparada com o primeiro ano, a tendência da CPUE padronizada apresenta uma tendência decrescente a partir do início da série. Até 1980 houve uma queda constante, quando então a CPUE aumentou novamente até o início da década de 1990. No entanto, esse discreto aumento esteve em patamares bem abaixo dos valores do início da série temporal analisada. A partir de meados de 1990 até 2009, a CPUE padronizada apresentou uma queda gradativa, demonstrando que valores mais baixos prevaleceram nos últimos anos, inclusive se comparados com os valores sem considerar os primeiros anos, que parecem estar em outro patamar (Figura 10).

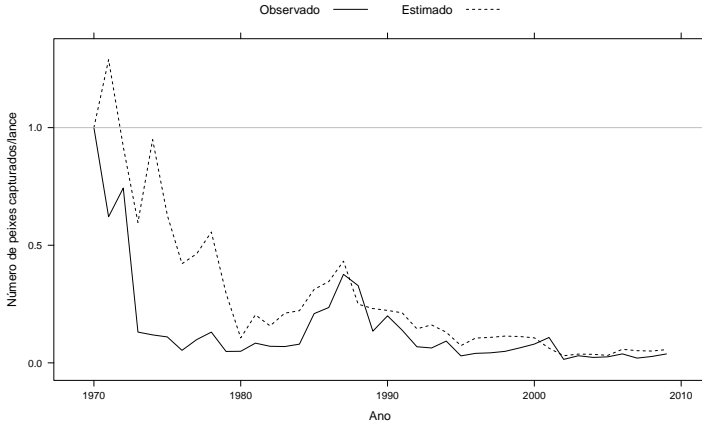


Figura 10: Captura por Unidade de Esforço (número de peixes capturados por lance com 1000 anzóis) do agulhão branco observada e estimada pelo modelo condicional BN. Os valores foram padronizados, tendo o primeiro ano como referência.

Através das previsões das probabilidades de captura por ano, trimestre e área, percebe-se claramente que a área CW, próxima à costa do Brasil, foi a que apresentou as maiores probabilidades, em todos os trimestres. Independente da área, parece haver uma diminuição das probabilidades ao longo dos anos (Figura 11).



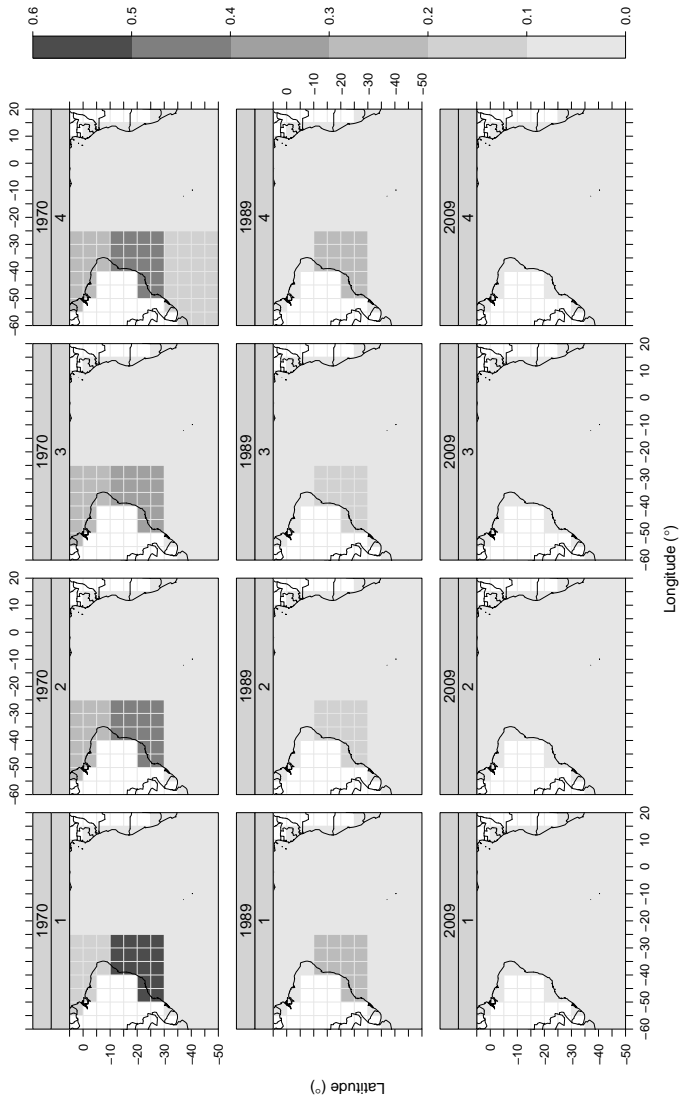


Figura 11: Probabilidade de captura trimestral por grande área do agulhão branco no oceano Atlântico Sul. São apresentados os resultados para o primeiro ano (1970), o ano central (1989) e o último ano (2009) da série temporal analisada (1970–2009).

## 4.3 Agulhão negro

### 4.3.1 Análise descritiva

A distribuição espaço-temporal da CPUE do agulhão negro mostrou que as áreas com os maiores valores estão no centro-oeste do Atlântico Sul, especialmente no primeiro e no quarto trimestres. No entanto, toda a parte tropical deste oceano demonstrou valores relativamente altos de CPUEs, independente do trimestre. Ainda pode ser notado que a região leste tropical, próxima à costa da África também apresenta um padrão com valores altos. As áreas onde a captura foi exclusivamente zero se concentram mais ao sul do oceano (Figura 12).

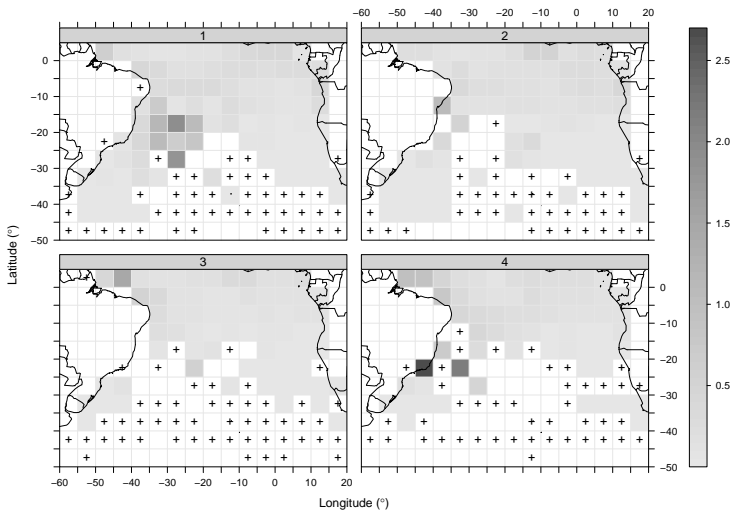


Figura 12: Captura por Unidade de Esforço (número de peixes capturados por lance com 1000 anzóis) trimestral do agulhão negro, obtida pela frota japonesa de espinhel de superfície no oceano Atlântico. Os quadrantes marcados com o símbolo + indicam o valor zero.

Através da distribuição das CPUEs por área e trimestre, observa-se que as áreas norte são as que apresentam os maiores valores. As áreas centrais,

particularmente CC e CW também correspondem à valores relativamente elevados. Nas áreas sul, foram observados os valores mais baixos de CPUE (Figura 13).

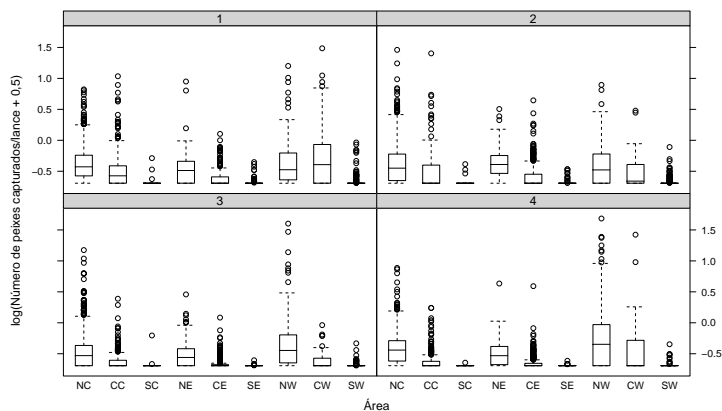


Figura 13: Distribuição da Captura por Unidade de Esforço (número de peixes capturados por lance com 1000 anzóis) trimestral do agulhão negro, obtida pela frota japonesa de espinhel de superfície em cada área no oceano Atlântico, entre 1970–2009. Os valores foram transformados para a escala  $\log(x + 0,5)$  para facilitar a visualização.

Através da análise da série temporal da CPUE do agulhão negro, observa-se que, no geral, houve um aumento a partir de 1979, permanecendo com valores mais altos até 2001. Após este ano, as CPUEs tenderam novamente a aumentar, embora com grandes oscilações. O primeiro trimestre de cada ano normalmente apresenta as maiores CPUEs (Figura 14).

### 4.3.2 Modelagem

A proporção de zeros encontrada para a captura do agulhão negro no Atlântico Sul pela frota de barcos japoneses de espinhel de superfície foi de 49%. Os valores de captura variaram no intervalo  $[0, 681]$  (Figura 15). A média de captura ponderada pelo esforço foi de cerca de  $\bar{x} = 50$  peixes, com variância  $s^2 = 2738$  para o período analisado (1970–2009). Com isto, pode-se constatar que a variância é cerca de 55 vezes maior do que a média,

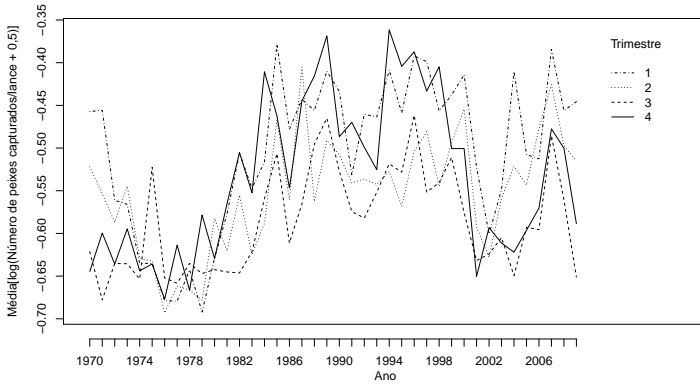


Figura 14: Média da Captura por Unidade de Esforço (número de peixes capturados por lance com 1000 anzóis) trimestral do agulhão branco, obtida pela frota japonesa de espinhel de superfície no oceano Atlântico, entre 1970–2009. Os valores foram transformados para a escala  $\log(x + 0,5)$  para facilitar a visualização.

indicando uma superdispersão considerável nos dados, em relação à distribuição de Poisson.

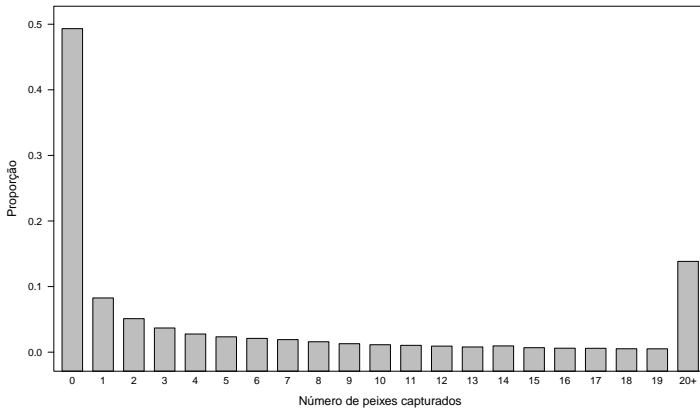


Figura 15: Proporção do número de agulhões negro capturados no Atlântico Sul no período 1970–2009. Como os valores de captura variaram no intervalo  $[0, 681]$ , aqueles maiores que 19 foram agregados na classe 20+.

Entre os modelos ajustados, o de Poisson foi o que apresentou os maiores valores de AIC e BIC, indicando o pior ajuste (Tabela 4). Os modelos de mistura e condicional ajustados com a distribuição de Poisson para os valores positivos tiveram resultados similares de acordo com os critérios de informação, e apresentaram um ganho pequeno em relação ao modelo de Poisson tradicional. Os modelos ajustados com a distribuição BN (tradicional, de mistura e condicional) apresentaram uma grande redução nos valores de AIC/BIC em relação aos modelos anteriores (ajustados com a distribuição de Poisson), o que indica um melhor ajuste. De acordo com esses critérios, o modelo condicional BN foi o que apresentou o melhor ajuste entre todos (menor AIC/BIC). O número de zeros predito pelos modelos condicionais (Poisson e BN) foi exatamente igual ao número de zeros observado (*i.e.* 6343). O modelo de mistura BN, foi o segundo que se aproximou do número de zeros observado. O modelo de Poisson subestimou substancialmente o número de zeros.

Tabela 4: Número de parâmetros, logaritmo da versossimilhança ( $\log L$ ), critério de informação de Akaike (AIC), critério de informação bayesiano (BIC), e número de zeros preditos (o número de zeros observados foi de 6343) pelos diferentes modelos ajustados para a captura do agulhão negro.

	Poisson	Quase- Poisson	BN	Poisson (mistura)	BN (mistura)	Poisson (cond.)	BN (cond.)
N. parâm.	51	51	52	102	103	102	103
$\log L$	-63292	—	-27010	-57940	-26708	-57008	-26519
AIC	126685	—	54125	116084	53622	114219	53244
BIC	127066	—	54513	116845	54390	114980	54013
N. zeros	4821	—	5948	5762	5998	6343	6343

Através dos resultados da comparação entre modelos acima, constatou-se portanto, que o modelo que melhor descreve os dados de captura do agulhão negro no Atlântico sul, foi o condicional utilizando a distribuição binomial negativa. Com esse modelo escolhido foi realizado então o processo de inclusão de interações. Assim como aconteceu no caso do agulhão branco, constatou-se que apenas a interação área:trimestre foi significativa para o modelo (TRV:  $p < 2, 2E - 16$ ), e as demais interações não puderam ser testadas devido à problemas de convergência do algoritmo de maximização da verossimilhança (pelos mesmos detalhes expostos para o modelo do agulhão branco

descrito acima).

Os resultados das estimativas dos parâmetros, tanto da parte dos modelos ajustada com a distribuição BN quanto com a regressão logística (distribuição binomial), e seus respectivos erros e  $p$ -valores associados são apresentados na Tabela 5. De maneira geral, a maioria das estimativas dos coeficientes para ambas as partes foram significativamente diferentes de zero, como pode ser constatado pelos  $p$ -valores. Através do modelo logístico, percebe-se que as probabilidades de captura são influenciadas negativamente pelas áreas, com exceção da área NW, onde esta mediada parece ser maior quando comparada com a área de referência NC. A área CW, de acordo com o coeficiente do modelo para as contagens, foi a única que apresentou aumento significativo do número médio de agulhões negro capturados. As estimativas dos coeficientes anuais, tanto da média de organismos capturados quanto da probabilidade de captura, aparentam normalmente uma diminuição quando comparados com o coeficiente de 1970 (ano de referência). Quanto aos trimestres, parece haver um aumento do número médio no segundo trimestre, e diminuição no terceiro e quarto trimestres. De fato, as probabilidades de captura aparentam ser menores nestes últimos dois períodos do ano.

O parâmetro de superdispersão ( $\theta$ ) da distribuição binomial negativa utilizada para os dados de contagem foi estimado em  $\log(\theta) = 0,265 \Rightarrow \theta = 1,3$  (Tabela 5). Esse valor elevado demonstra a grande superdispersão para os dados de captura do agulhão negro.

Tabela 5: Estimativas dos parâmetros do modelo condicional ajustado com a distribuição binomial negativa para a captura do agulhão negro. As colunas “Contagem” se referem aos coeficientes do modelo BN para as contagens. As colunas “Zero” representam os coeficientes para o modelo logístico considerando a presença/ausência de zeros.

Termos	Contagem			Zero		
	Estimativa	EP	Pr(> z )	Estimativa	EP	Pr(> z )
(Intercepto)	-8,09	0,0929	0,00	-8,05	0,191	0,00
ÁreaCC	-0,351	0,0656	8,62E-08	-0,788	0,156	4,36E-07
ÁreaSC	-1,79	0,886	0,0431	-7,44	0,617	2,15E-33
ÁreaNE	-0,182	0,128	0,155	-0,236	0,276	0,393
ÁreaCE	-0,958	0,117	3,24E-16	-2,06	0,190	2,91E-27
ÁreaSE	-4,32	0,636	1,14E-11	-5,58	0,275	3,16E-91
ÁreaNW	0,0289	0,0754	0,701	0,339	0,205	0,0985
ÁreaCW	0,962	0,171	1,70E-08	-0,151	0,335	0,652
ÁreaSW	-1,75	0,211	1,25E-16	-3,44	0,227	1,13E-51
Ano1971	-0,175	0,128	0,172	-0,324	0,227	0,152
Ano1972	-0,118	0,138	0,395	-0,152	0,235	0,519
Ano1973	-0,210	0,134	0,118	0,242	0,248	0,328
Ano1974	-0,323	0,197	0,102	0,0886	0,316	0,779
Ano1975	-0,636	0,151	2,64E-05	0,0210	0,268	0,938
Ano1976	-1,36	0,353	0,000111	-0,738	0,407	0,0699
Ano1977	-0,936	0,223	2,71E-05	-0,134	0,344	0,697
Ano1978	-0,617	0,226	0,00643	-0,220	0,318	0,490
Ano1979	-0,682	0,172	7,41E-05	-0,377	0,306	0,217
Ano1980	-0,907	0,138	4,22E-11	0,0123	0,254	0,961
Ano1981	-0,617	0,136	5,98E-06	-0,543	0,235	0,0207
Ano1982	-0,227	0,122	0,0635	-0,284	0,237	0,232
Ano1983	-0,357	0,140	0,0109	-0,210	0,276	0,448
Ano1984	-0,329	0,122	0,00699	0,357	0,273	0,191
Ano1985	-0,144	0,112	0,199	0,328	0,233	0,160
Ano1986	-0,383	0,120	0,00146	-0,245	0,230	0,287
Ano1987	0,0266	0,127	0,834	0,0550	0,245	0,822
Ano1988	-0,269	0,111	0,0152	0,206	0,235	0,381
Ano1989	-0,0843	0,109	0,437	-0,117	0,232	0,616
Ano1990	-0,418	0,109	0,000122	-0,187	0,230	0,417
Ano1991	-0,642	0,108	3,34E-09	-0,236	0,212	0,266
Ano1992	-0,555	0,115	1,51E-06	-0,631	0,226	0,00526
Ano1993	-0,633	0,111	1,03E-08	-0,136	0,229	0,553
Ano1994	-0,115	0,108	0,287	-0,414	0,220	0,0605
Ano1995	-0,135	0,110	0,220	-0,823	0,217	0,000154
Ano1996	0,0374	0,108	0,729	-0,440	0,220	0,0458
Ano1997	-0,241	0,110	0,0291	-0,537	0,219	0,0144
Ano1998	-0,00517	0,114	0,964	-0,221	0,226	0,329
Ano1999	-0,208	0,115	0,0711	-0,375	0,227	0,0982
Ano2000	-0,367	0,112	0,00107	-0,345	0,221	0,118
Ano2001	-0,904	0,129	2,25E-12	-1,97	0,226	2,78E-18
Ano2002	-1,31	0,129	4,92E-24	-1,82	0,235	7,97E-15
Ano2003	-0,701	0,117	2,36E-09	-1,62	0,217	7,21E-14
Ano2004	-0,374	0,121	0,00197	-1,99	0,224	6,50E-19
Ano2005	-0,609	0,124	8,42E-07	-1,65	0,225	2,44E-13

Continua na próxima página ...

... continuação da página anterior.

Termos	Contagem			Zero		
	Estimativa	EP	Pr(> z )	Estimativa	EP	Pr(> z )
Ano2006	-0,691	0,116	2,34E-09	-0,661	0,226	0,00339
Ano2007	-0,261	0,111	0,0192	-0,280	0,247	0,257
Ano2008	-0,625	0,108	8,18E-09	-0,0268	0,229	0,907
Ano2009	-0,907	0,111	3,45E-16	-0,457	0,220	0,0378
Trim2	0,172	0,0477	0,000304	-0,158	0,135	0,242
Trim3	-0,278	0,0494	1,80E-08	-0,597	0,128	3,16E-06
Trim4	-0,145	0,0438	0,000917	-0,392	0,129	0,00245
ÁreaCC:Trim2	-0,0864	0,106	0,413	-0,507	0,215	0,0183
ÁreaSC:Trim2	-0,278	1,45	0,848	-1,61	0,854	0,0588
ÁreaNE:Trim2	0,0427	0,156	0,784	0,640	0,391	0,101
ÁreaCE:Trim2	0,0995	0,142	0,483	-0,171	0,242	0,481
ÁreaSE:Trim2	2,10	0,733	0,00407	-1,54	0,411	0,000172
ÁreaNW:Trim2	-0,129	0,117	0,270	-0,309	0,279	0,269
ÁreaCW:Trim2	-1,18	0,388	0,00226	-0,371	0,587	0,527
ÁreaSW:Trim2	-0,632	0,270	0,0193	0,0326	0,300	0,914
ÁreaCC:Trim3	-0,676	0,114	3,28E-09	-0,822	0,214	0,000122
ÁreaSC:Trim3	-14,3	859	0,987	0,306	0,956	0,749
ÁreaNE:Trim3	-0,0642	0,151	0,670	-0,00134	0,334	0,997
ÁreaCE:Trim3	-0,555	0,148	0,000172	-0,758	0,233	0,00113
ÁreaSE:Trim3	0,142	0,814	0,862	-0,0318	0,428	0,941
ÁreaNW:Trim3	0,568	0,117	1,13E-06	0,555	0,290	0,0551
ÁreaCW:Trim3	-0,992	0,375	0,00818	-0,570	0,552	0,302
ÁreaSW:Trim3	-1,43	0,404	0,000421	0,432	0,345	0,210
ÁreaCC:Trim4	-1,19	0,117	4,62E-24	-0,827	0,219	0,000161
ÁreaSC:Trim4	-12,7	643	0,984	1,38	1,21	0,254
ÁreaNE:Trim4	-0,160	0,154	0,299	0,524	0,346	0,130
ÁreaCE:Trim4	-0,728	0,135	7,82E-08	-0,548	0,231	0,0177
ÁreaSE:Trim4	0,359	1,08	0,739	-0,174	0,545	0,750
ÁreaNW:Trim4	0,846	0,111	2,03E-14	0,0833	0,278	0,765
ÁreaCW:Trim4	-0,0646	0,281	0,818	-0,565	0,489	0,248
ÁreaSW:Trim4	-0,483	0,505	0,339	0,722	0,428	0,0914
log( $\theta$ )	0,265	0,0266	2,47E-23	—	—	—

A análise dos resíduos do modelo condicional BN mostrou que a maioria dos valores se concentram em torno de zero, com alguns valores extremos podendo ser identificados (Figura 16). Assim como na análise dos resíduos para o modelo do agulhão branco, aqui também não foi possível realizar uma análise mais aprofundada dos resíduos e de medidas de influência. Como a base de dados utilizada foi a mesma, o mesmo problema computacional para o cálculo de matrizes foi encontrado (mais detalhes foram descritos acima).

Com os valores de CPUE padronizados, observa-se uma leve tendência de declínio nas CPUEs, quando comparadas ao valor de 1970, sempre com grandes oscilações. Até 1983, pode-se verificar que a CPUE observada era mais baixa do que aquela que foi predita pelo modelo, quando, então, esse padrão se inverte até o final da série temporal. A partir de 1989, as CPUEs estimadas ficaram sempre abaixo do valor inicial de 1970. Houve uma grande queda a



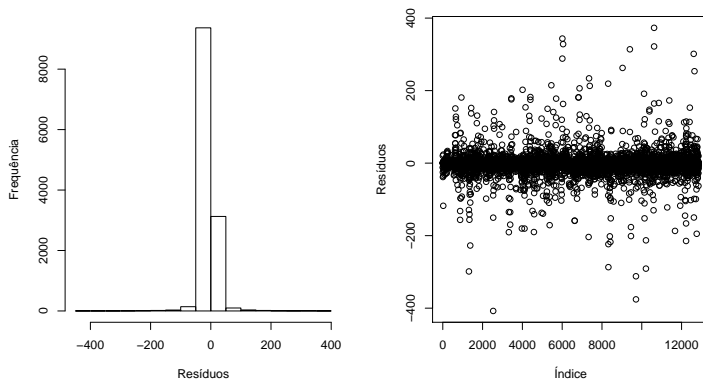


Figura 16: Histograma de frequência (esquerda) e gráfico de dispersão (direita) dos resíduos do modelo condicional BN para o agulhão negro.

partir de 2000, com um rápido crescimento a partir de 2004 (Figura 17).

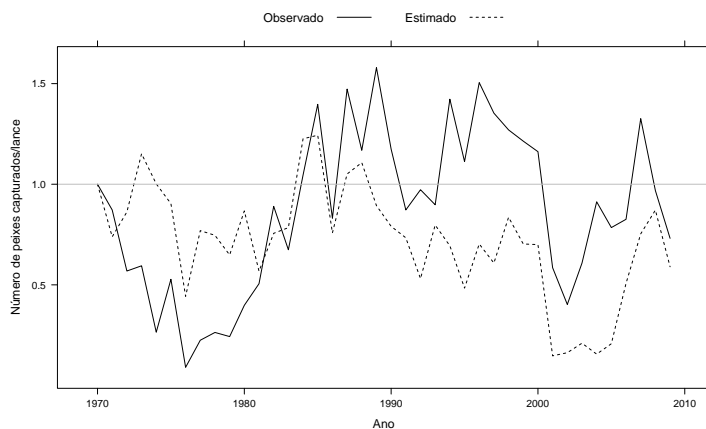


Figura 17: Captura por Unidade de Esforço (número de peixes capturados por lance com 1000 anzóis) do agulhão negro observada e estimada pelo modelo condicional BN. Os valores foram padronizados, tendo o primeiro ano como referência.

Através das predições das probabilidades de captura por ano, trimestre e área, percebe-se claramente que, independente do trimestre, as áreas NW, NC

e NE são as que apresentam as maiores probabilidades. Em particular, a área NW, próxima à costa norte do Brasil, tem maior destaque. Durante o primeiro e o segundo trimestre dos anos analisados, ainda pode-se constatar que, além das áreas já mencionadas, as áreas mais centrais e à oeste do oceano (CW e CC) também apresentam maiores probabilidades de captura do agulhão negro. A área NE, mais próxima à costa da África, também apresenta valores de probabilidade elevados em determinados períodos (Figura 18).

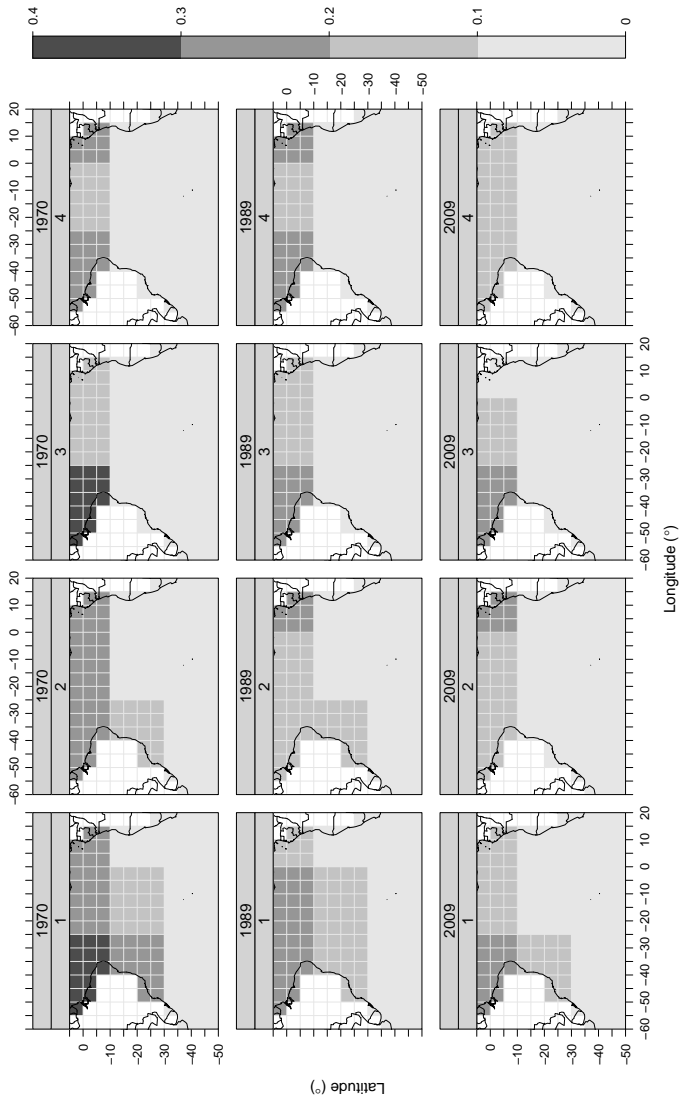


Figura 18: Probabilidade de captura trimestral por grande área do agulhão negro no oceano Atlântico Sul. São apresentados os resultados para o primeiro ano (1970), o ano central (1989) e o último ano (2009) da série temporal analisada (1970–2009).

## 5 Discussão

### 5.1 Modelos

A comparação entre os modelos mostra claramente que os modelos que consideraram a distribuição binomial negativa tiveram melhores resultados (se considerados os critérios de informação AIC/BIC) do que os demais modelos onde a distribuição de Poisson foi utilizada. Mesmo os modelos de mistura e condicional de Poisson, que levam em conta o excesso de zeros, não tiveram um bom desempenho, no sentido de tentar descrever os dados de captura dos agulhões no Atlântico sul. Isso se deve possivelmente ao fato de que a superdispersão, sendo extremamente elevada (note que a variância ponderada pelo esforço foi 43 vezes maior do que a média ponderada para o agulhão branco, e 55 vezes maior para o agulhão negro), faz com que mesmo que um modelo considere o excesso de zeros, ainda falta considerar essa superdispersão excessiva. É nesse ponto que o modelo BN leva vantagem, pois possui um parâmetro a mais ( $\theta$ ) que é responsável pela “acomodação” dessa grande variabilidade.

Considerando a estrutura dos modelos com excesso de zeros (mistura e condicional), percebe-se que os modelos condicionais levam vantagem em relação aos modelos de mistura (comparando mistura Poisson com condicional Poisson e mistura BN com condicional BN). Esse fato pode estar relacionado à fonte dos zeros no tipo de dado utilizado nesta análise. Em ecologia os zeros podem ser classificados em “zero verdadeiro” e “zero falso”. O primeiro ocorre quando uma espécie não é detectada devido a um processo ecológico real (*e.g.* habitat desfavorável). O último pode ocorrer de duas formas: (a) a espécie ocupa o habitat, mas não está presente durante a amostragem (nesse caso, a pescaria), e (b) além de ocupar o habitat, a espécie está presente mas o observador falha na sua detecção (nesse caso, os peixes podem não ser atraídos pelas iscas), o que é comum quando a espécie é rara. A fonte de captura zero em pescarias comerciais se encaixa nesta segunda definição, pois as pescarias ocorrem exclusivamente em locais (*i.e.* habitats) onde os pescadores sabem que determinadas espécies (ou a espécie-alvo) ocorrem. No caso do espinhel de superfície, a pescaria ocorre no ambiente pelágico, habitat tanto

das espécies-alvo (atuns e espadarte) quanto dos agulhões, e portanto, não ocorre a situação de se obter um “zero verdadeiro”.

Os resultados obtidos também demonstram que estimativas de abundância relativa de espécies capturadas incidentalmente devem ser feitas através de modelos que considerem tanto a superdispersão como o excesso de zeros. Tradicionalmente na ciência pesqueira, o procedimento padrão para a padronização da CPUE é através da aplicação de Modelos Lineares Generalizados com família de erro log-normal (*e.g.* Hilborn & Walters, 1992). Quando ocorre dos dados de uma determinada espécie apresentarem uma moderada porcentagem de zeros, o comum é se aplicar o método denominado “delta” (Lo *et al.*, 1992; Stefánsson, 1996). Este método consiste em modelar separadamente os dados positivos (com uma distribuição para dados contínuos como a normal ou log-normal), e a presença/ausência de zeros (através de uma regressão logística tradicional, com distribuição binomial). Posteriormente, é realizada uma combinação destes resultados para se obter a estimativa final. Esta metodologia é similar à dos modelos condicionais para dados de contagem apresentados aqui, e pode ser útil em casos onde o excesso de zeros é moderado e não ocorre superdispersão, uma situação pouco provável em dados de espécies capturadas incidentalmente. Portanto, o uso dos modelos de mistura e condicional, especialmente utilizando a distribuição binomial negativa que permite modelar a superdispersão, parece ser mais adequado para a modelagem de dados de espécies pouco frequentes em pescarias comerciais.

## 5.2 Aplicações à ecologia dos agulhões

Assumindo que os modelos utilizados são os mais apropriados para descrever a captura dos agulhões no oceano Atlântico sul, pode-se considerar que o estoque de ambas espécies analisadas apresentam uma tendência de declínio ao longo da série temporal. A padronização da CPUE do agulhão branco mostra que a situação atual dos padrões de abundância relativa do estoque é muito menor do que aquela observada no início da década de 1970 (Figura 10). No entanto, os elevados padrões de abundância desta espécie observados neste período, ainda é uma questão sendo debatida no âmbito da

ICCAT. Muitos pesquisadores consideram que os dados de espinhel de superfície da frota japonesa podem conter informações não-consolidadas no início da série histórica desta frota. A tendência de diminuição do estoque para o agulhão negro (Figura 17) é mais discreta se comparada ao agulhão branco. No entanto, ainda é uma questão que deve ser considerada nas decisões de manejo futuras desta espécie.

Considerando as probabilidades de captura, pode ser concluído que as regiões tropicais do Atlântico sul são áreas de captura elevadas para ambas espécies (Figuras 11 e 18). As áreas de maior probabilidade para o agulhão branco são mais restritas ao oeste do oceano, principalmente na região central-oeste, próxima à costa do Brasil (Figura 11). Já para o agulhão negro, probabilidades de captura maiores são menos restritas, e podem ser observadas em todo Atlântico sul tropical, e em áreas mais centrais e do centro-oeste do oceano, principalmente no primeiro e segundo trimestre (Figura 18).

Recentemente, a FAO divulgou um documento com sugestões de medidas para o manejo de espécies capturadas incidentalmente e para a redução dos descartes destas espécies (FAO, 2010). Este documento cita o Código de Conduta para a Pesca Responsável, de 1995 (elaborada pela própria FAO), que já recomendava “o uso sustentável dos ecossistemas aquáticos”, e “que as pescarias sejam conduzidas considerando o impacto ambiental”. Neste Código ainda consta que “devem ser promovidos meios de manutenção e preservação da biodiversidade dos ecossistemas, minimizando o impacto das pescarias em espécies que não são alvo, e nos ecossistemas em geral”. Apesar de todas estas recomendações do Código da FAO, atualmente há uma crescente preocupação sobre o aumento dos níveis de mortalidade por pesca como resultado de capturas incidentais e dos descartes, comprometendo a sustentabilidade a longo prazo de muitas pescarias, e afetando a manutenção da biodiversidade em diversas regiões (FAO, 2010). Particularmente, a captura dos agulhões branco e negro já vêm sofrendo esse tipo de pressão, como reportado anteriormente pela ICCAT (ICCAT, 2007a). Tanto a FAO quanto a ICCAT estão também estimulando estudos que levem à redução ou à eliminação das capturas incidentais em todo o mundo (FAO, 2010).

Os resultados obtidos neste trabalho visam dar apoio a estas medidas que

vêm sendo solicitadas por organizações internacionais. De uma forma ainda inicial e discreta, os resultados apresentados aqui podem servir de apoio, ou como uma primeira tentativa de se identificar áreas no oceano Atlântico sul, onde as probabilidades de captura sejam altas. No documento da FAO supracitado, entre muitas medidas sugeridas para a redução de capturas incidentais, estão ações que consideram medidas espaço-temporais (FAO, 2010, item 7.6). Estas recomendações se baseiam principalmente no fechamento de áreas para a pesca (ou tipo de pescaria) em determinados períodos do ano. Para isso, uma alternativa foi apresentada aqui. Com a identificação de áreas com probabilidade alta de captura dos agulhões, estas áreas poderiam ser fechadas em trimestres específicos, com a intenção de se reduzir a captura incidental destas espécies.

As medidas atuais tomadas pelo governo brasileiro, que envolvem a proibição da comercialização, e a devolução dos espécimes de agulhão capturados vivos, ainda são pouco efetivas no sentido de que nenhuma ação para a redução de fato das capturas foi realizada. Somente com ações que possam minimizar efetivamente as capturas, o objetivo de redução de *bycatch* e descartes pode ser atingido.

### **5.3 Sugestões para trabalhos futuros**

Uma primeira dificuldade encontrada quando se consideram dados superdispersos é ainda na seleção de modelos utilizando diferentes distribuições de probabilidade. Os diferentes critérios de seleção utilizados (*e.g.* teste de razão de verossimilhança, AIC, BIC) poderiam ter uma performance inferior caso os dados não fossem superdispersos. Fitzmaurice (1997) propõe modificações para estes critérios, com a intenção de incorporar a superdispersão. Estas modificações devem ser consideradas em trabalhos futuros.

Uma consideração importante quando se analisam dados de pescarias com estruturas espaciais, é levar em conta esta espacialidade na estrutura do modelo. Por exemplo, Pereira *et al.* (2009) consideraram ferramentas geoestatísticas bayesianas para incorporar a questão espacial na modelagem de diferentes índices de CPUEs da albacora-lage (onde não se considerou haver

superdispersão nem excesso de zeros). Fora da literatura pesqueira, Agarwal *et al.* (2002) e Gschlöbl & Czado (2008) utilizaram também ferramentas geoestatísticas e bayesianas, mas na modelagem de dados com superdispersão e excesso de zeros. A estrutura utilizada por estes últimos autores pode ser um objeto de aplicação em dados de captura superdispersos e com excesso de zeros, como os utilizados neste trabalho.

Shono (2008) propôs o uso da distribuição Tweedie para modelar a captura de espécies com excesso de zeros. A distribuição Tweedie apresenta uma propriedade importante, que permite a modelagem de valores positivos e zeros ao mesmo tempo, sem a necessidade do uso de modelos de dois estágios, como os modelos de mistura e condicionais. Apesar de já ter sido utilizada por Shono (2008) em dados com excesso de zeros, seu uso em dados que apresentam tanto excesso de zeros quanto superdispersão ainda deve ser estudado.

Finalmente, um trabalho interessante e bastante completo seria a comparação dos diversos modelos e estruturas de análise citadas acima, além dos modelos de mistura e condicionais utilizados neste trabalho. Neste estudo poderia ser utilizado dados de diferentes espécies, que apresentam ou não superdispersão e/ou excesso de zeros, com a finalidade de se tentar determinar classes de modelos mais adequados para cada tipo de dado utilizado.



## Referências

- Agarwal, D. K.; Gelfand, A. E. & Citron-Pousty, S. 2002. Zero-inflated models with application to spatial count data. *Environmental and Ecological Statistics*, 9: 341–355.
- Akaike, H. 1974. A new look at the statistical model identification. *IEEE Trans. Automat. Contr.* 19(6).
- Amorim, A. F. & Arfelli, C. A. 2001. Analysis of the Santos fleet from São Paulo, southern Brazil (1971–1999). *Col. Vol. Sci. Pap. ICCAT*. 53: 263–271.
- Amorim, A. F. & Arfelli, C. A. 2003. Review of white marlin (*Tetrapturus albidus*) fishery biology off southern brazilian coast. *Col. Vol. Sci. Pap. ICCAT*. 55(2): 467–474.
- Andrade, H. A. 2007. Estimation of the relative abundance of Atlantic billfish: effects of three approaches to cope with catches equal to zero. *Col. Vol. Sci. Pap. ICCAT*. 60(5): 1707–1719.
- Arfelli, C. A. 1996. Estudo da pesca e aspectos da dinâmica populacional de espadarte, *Xiphias gladius*, L. 1758, no Atlântico sul. 175 f. Tese de Doutorado – Instituto de Biociências, Universidade Estadual Paulista, Rio Claro (SP).
- Borgatto, A. F.; Demétrio, C. G. B. & Leandro, R. A. 2006. Modelos para proporções com superdispersão e excesso de zeros — um procedimento bayesiano. *Revista de Matemática e Estatística*. 4(2): 121–131.
- Cunningham, R. B. & Lindenmayer, D. B. 2005. Modeling count data of rare species: some statistical issues. *Ecology*. 86(5): 1135–1142.
- Dobson, A. J. 2002. *An introduction to Generalized Linear Models*. Chapman & Hall/CRC, London. 225 p.
- FAO. 2010. Report of the technical consultation to develop international guidelines on bycatch management and reduction of discards. Rome, 6–10 December 2010. *FAO Fisheries and Aquaculture Report 957*. Rome, FAO, 32 p.
- Fitzmaurice, G. M. 1997. Model selection with overdispersed data. *J. R. Statist. Soc. D*. 46(1): 81–91.

- Fletcher, D.; Mackenzie, D. & Villouta, E. 2005. Modeling skewed data with many zeros: A simple approach combining ordinary and logistic regression. *Environmental and Ecological Statistics*. 12: 45–54.
- Grogger, J. T. & Carson, R. T. 1991. Models for truncated counts. *Journal of Applied Econometrics*. 6: 225–238.
- Gschlöbl, S. & Czado, C. 2008. Modelling count data with overdispersion and spatial effects. *Statistical Papers*, 49: 531–552.
- Gulland, J. A. 1983. *Fish stock assessment: a manual of basic methods*. John Wiley & Sons.
- Hilborn, R. & Walters, C. J. 1992. *Quantitative fisheries stock assessment – Choice, dynamics & uncertainty*. Chapman and Hall, New York. 570 p.
- Hinde, J. & Demétrio, C. G. B. 1998. Overdispersion: Models and estimation. *Computational Statistics & Data Analysis*. 27: 151–170.
- ICCAT. 2003. Report of the 2002 ICCAT white marlin stock assessment meeting. Col. Vol. Sci. Pap. ICCAT. 55(2): 350–452.
- ICCAT. 2007a. Report of the 2006 ICCAT billfish stock assessment. Col. Vol. Sci. Pap. ICCAT. 60(5): 1431–1546.
- ICCAT. 2007b. Report for biennial period 2006-07. Part I (2006) – Vol. 2. Madrid, Spain. 244 p.
- ICCAT. 2008. Report for biennial period 2006-07. Part II (2007) – Vol. 2. Madrid, Spain. 266 p.
- Lambert, D. 1992. Zero-Inflated Poisson regression, with application to defects in manufacturing. *Technometrics*. 34(1): 1–14.
- Lo, N. C. H.; Jacobson, L. D. & Squire, J. L. 1992. Indices of relative abundance from fish spotter data based on delta-lognormal models. *Can. J. Fish. Aquat. Sci.* 49: 2515-2526.
- 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.

- McCulagh, P. & Nelder, J. A. 1989. *Generalized Linear Models*. Chapman & Hall, London. 513 p.
- Mullahy, J. 1997. Heterogeneity, excess zeros, and the structure of count data models. *Journal of Applied Econometrics*. 12: 337–350.
- Nakamura, I. 1985. *FAO species catalogue*. Vol. 5. Billfishes of the world. An annotated and illustrated catalogue of marlins, sailfishes, spearfishes and swordfishes known to date. *FAO Fish. Synop.* 125(5): 65 p.
- Nelder, J. A. & Wedderburn, R. W. M. 1972. *Generalized Linear Models*. *J. R. Statist. Soc. A.* 135(3): 370–384.
- Paula, G. A. 2004. *Modelos de regressão com apoio computacional*. Disponível em: [www.ime.usp.br/~giapaula/mlgs.html](http://www.ime.usp.br/~giapaula/mlgs.html).
- Pereira, J.; Leandro, R. A.; Petrere Jr., M. & Nishida, T. 2009. Comparing three indices of catch per unit effort using Bayesian geostatistics. *Fisheries Research*. 100: 200–209.
- Potts, J. M. & Elith, J. 2006. Comparing species abundance models. *Ecological Modelling*. 199: 153–163.
- R Development Core Team. 2010. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL [www.R-project.org](http://www.R-project.org).
- Ridout, M; Demétrio, C. G. B. & Hinde, J. 1998. Models for count data with many zeros. In: *International Biometric Conference, 1998, Cape Town. Proceedings...*, Cape Town: IBC, 1998, p. 179–192.
- Ridout, M; Hinde, J. & Demétrio, C. G. B. 2001. A score test for testing a Zero-Inflated Poisson regression model against Zero-Inflated Negative Binomial alternatives. *Biometrics*. 57: 219–223.
- Saito, H. & Yokawa, K. 2006. Use of pop-up tags to estimate vertical distribution of Atlantic blue marlin (*Makaira nigricans*) released from the commercial and research longline cruise during 2002 and 2003. *Col. Vol. Sci. Pap. ICCAT*. 59(1): 252–264.
- Sarkar, D. 2008. *Lattice: multivariate data visualization with R*. Springer, Nova Iorque. 265 p.
- Schwarz, G. 1978. Estimating the dimension of a model. *Annals of Statistics*, 6(2): 461–464.

- Shono, H. 2008. Application of the Tweedie distribution to zero-catch data in CPUE analysis. *Fisheries Research*, 93: 154–162.
- Sileshi, G. 2008. The excess-zero problem in soil animal count data and choice of appropriate models for statistical inference. *Pedobiologia*. 1–14.
- Stefánsson, G. 1996. Analysis of groundfish survey abundance data: combining the GLM and delta approaches. *ICES J. Mar. Sci.* 53: 577-588.
- Sund, P.; Blackburn, M. & Willians, F. 1981. Tunas and their environment in the Pacific ocean: a review. *Oceanogr. Mar. Biol. Ann. Rev.* 19: 443–512.
- Tukey, J. W. 1977. *Exploratory data analysis*. Addison-Wesley Publishing Company, Reading, Massachusetts. 688 p.
- Venables, W. N.; Ripley, B. D. 2002. *Modern applied statistics with S*. Springer, Nova Iorque, 495 p.
- Welsh, A. H.; Cunningham, R. B.; Donnelly, C. F. & Lindenmayer, D. B. 1996. Modelling the abundance of rare species: statistical models for counts with extra zeros. *Ecological Modelling*. 88: 297–308.
- Welsh, A. H.; Cunningham, R. B.; Chambers, R. L. 2000. Methodology for estimating the abundance of rare animals: Seabird nesting on North East Herald Cay. *Biometrics*. 56: 22–30.
- Zeileis, A.; Kleiber, C.; Jackman, S. 2008. Regression models for count data in R. *Journal of Statistical Software*. 27(8): 1–25.