

ALEXANDRO ALVES VIEIRA

**Modelagem da taxa de captura para o espadarte capturado pela frota
espinheira brasileira no Atlântico Sul: enfoque transversal versus
longitudinal**

RECIFE-PE - ABRIL/2015



UNIVERSIDADE FEDERAL RURAL DE PERNAMBUCO
PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM BIOMETRIA E ESTATÍSTICA APLICADA

**Modelagem da taxa de captura para o espadarte capturado pela frota
espinheira brasileira no Atlântico Sul: enfoque transversal versus
longitudinal**

Tese apresentada ao Programa de Pós-Graduação em Biometria e Estatística Aplicada como exigência parcial à obtenção do título de Doutor.

Área de Concentração: Biometria e Estatística Aplicada

Orientador: Prof. Dr. Humber Agreli Andrade

RECIFE-PE - ABRIL/2015.

UNIVERSIDADE FEDERAL RURAL DE PERNAMBUCO
PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM BIOMETRIA E ESTATÍSTICA APLICADA

**Modelagem da taxa de captura para o espadarte capturado pela frota
espinheira brasileira no Atlântico Sul: enfoque transversal versus
longitudinal**

Alexandro Alves Vieira

Tese julgada adequada para obtenção do título de Doutor em Biometria e Estatística Aplicada, defendida e aprovada por unanimidade em 27/02/2015 pela Banca Examinadora.

Orientador:

Prof. Dr. Humber Agrelli Andrade
UFRPE-DEPAq

Banca Examinadora:

Prof. Dr. Paulo José Duarte Neto
DEINFO-UFRPE

Prof. Dr. Francisco José Victor de Castro
UAE-UFCG

Prof. Dr. Gilberto da Silva Matos
UAME-UFCG

Prof. Dr. Lázaro de Souto Araújo
DECAT-UFS

"Dedico este trabalho à minha mãe

Eneida Alves Vieira"

(in memoriam)

Agradecimentos

Agradeço primeiramente a Deus, por estar comigo em todos os momentos de minha vida, me mostrando o caminho a seguir nas horas difíceis.

Ao Prof. Humber Agrelli Andrade a orientação, os momentos compartilhados e por acreditar sempre em mim.

Aos professores membros da banca examinadora, Paulo José Duarte Neto, Gilberto da Silva Matos, Francisco José Victor de Castro e Lázaro de Souto Araújo pelas valiosas sugestões para este trabalho.

Aos professores do departamento de Pesca e Aquicultura da UFRPE, Paulo Travassos e Rosângela Lessa pelas contribuições no período de qualificação.

Aos demais professores do Programa de Pós Graduação em Biometria e Estatística Aplicada da UFRPE, pelos conhecimentos transmitidos, atenção e amizade.

Aos funcionários do Programa de Pós Graduação em Biometria e Estatística Aplicada da UFRPE, em particular o secretário Marco, a dedicação e atenção nos momentos solicitados.

Aos amigos do Programa de Pós Graduação em Biometria e Estatística Aplicada da UFRPE, durante os anos de doutorado convivi com muitas pessoas especiais, com as quais pude compartilhar momentos de estudo, ansiedade e boas conversas.

Ao amigo Chico e seu filho Bruno por me acomodarem em sua casa em Recife durante a realização do doutorado.

Aos meus queridos amigos, pelas valiosas discussões e, principalmente, pela amizade tornando os momentos de convívio muito mais agradáveis e divertidos.

A meu pai Onaci e sua esposa Lucia, pelo carinho e por sempre compreender a minha ausência ou cansaço em algumas confraternizações, sempre incentivadores de meu crescimento pessoal e profissional.

Às minhas irmãs Adiene e Adriana por sempre me incentivar.

À minha esposa Alane, por ter enfrentado esta batalha comigo, pelo apoio e pela

alegria em tê-la sempre ao meu lado, principalmente nos momentos mais difíceis.

À minha filha Lara e minha sogra Hilda, por me desejarem boa-sorte a cada viagem e estarem sempre dispostas a ajudar no que fosse necessário.

Enfim, a todas as pessoas que de forma direta ou indireta, apoiaram-me nessa decisão de abdicar de tantas coisas em minha vida, para me dedicar ao doutorado e que torceram para tudo dar certo.

"A mente que se abre a uma nova ideia jamais volta ao seu tamanho original."

Albert Einstein

Resumo

No presente trabalho são apresentados métodos estatísticos de regressão para obtenção de séries padronizadas da captura por unidade de esforço (CPUE) do espadarte (*Xiphias gladius*) capturado pela frota atuneira brasileira no Atlântico Sul, de 1998 a 2006. Tais séries, são fundamentais para a avaliação do estoque, e a gestão sustentável da pescaria. De início executamos uma abordagem transversal fundamentada na suposição de independência entre as observações, inclusive, as advindas de uma mesma embarcação em ocasiões distintas. Para tanto, foram ajustados diversos modelos pertencentes à classe de Modelos Lineares Generalizados (GLM) para dados de contagens, levando em conta a suposta sobredispersão e o excesso de zeros presente nos dados, sendo revelada a superioridade do modelo binomial negativo com função de ligação logarítmica. A proporção da deviência explicada pelo modelo foi de aproximadamente 54,5% (pseudo- r^2). As variáveis ano e número de anzóis entre boias explicaram 63% do total da deviência que foi explicada pelo modelo. Apesar das oscilações e da falta de precisão da estimativa em 1998, a tendência da série da CPUE padronizada revelou um padrão de estabilidade da biomassa, sugerindo que a pesca espinheira não esteja afetando fortemente o estoque de espadarte no Atlântico Sul, resultado que concorda com as conclusões e recomendações da ICCAT quanto ao manejo do referido estoque. Em seguida, foi realizada uma análise sob o enfoque longitudinal, em que se considerou a possibilidade de correlação entre as observações de uma mesma embarcação ao longo do tempo, com o uso da teoria das Equações de Estimção Generalizadas (GEE), para estender os GLMs aos dados longitudinais. Para esta análise, foram utilizados dados relativos a eventos de pesca realizados durante os meses de maio, junho e julho de 2001 (92 dias), por oito embarcações arrendadas da Espanha. O modelo GEE com distribuição de Poisson e estrutura de correlação auto-regressiva de ordem 1 apresentou melhor ajuste e foi utilizado na padronização da CPUE de espadarte, em escala mensal. O ajuste deste modelo revelou a influencia significativa das variáveis número de anzóis entre boias, dia e mês sobre a CPUE de espadarte. Ao se ajustar, a partir da mesma base de dados, um modelo GLM com mesma distribuição, preditor linear e função de ligação utilizadas nos modelos GEE, foi verificado que as estimativas dos parâmetros diferem bastante e que o ajuste via GLM subestima os erros padrões, alterando as-

sim as significâncias dos parâmetros. Quanto à padronização da CPUE através dos modelos GLM e GEE, verificou-se que, apesar de ambas as séries apresentarem em comum uma tendência decrescente nos índices relativos de abundância ao longo dos três meses considerados, os índices padronizados via GLM apresentam decréscimo mais acentuado, decrescendo 58% do primeiro para o terceiro mês, contra 46% relativo ao GEE. Esta diferença na declividade das séries pode promover uma avaliação incorreta do estoque da espécie e medidas regulatórias equivocadas.

Palavras-chave: CPUE; espinhel de superfície; GLM; dados longitudinais; GEE.

Abstract

In this paper we present statistical regression methods to obtain standardized series of catch per unit effort (CPUE) of swordfish (*Xiphias gladius*) caught by Brazilian tuna fleet in the South Atlantic, from 1998 to 2006. These series are essential for evaluating the stock, and the sustainable management of the fishery. At first we performed a cross-cutting approach based on the assumption of independence between observations, including those arising from the same vessel on different occasions. To this end, were adjusted several models in Class Generalized Linear Models (GLM) to counts of data, taking into account the alleged overdispersion and excess zeros in this data, and revealed the superiority of the negative binomial model with logarithmic link function. The proportion of deviance explained by the model was approximately 54,5% (pseudo- r^2). Variables year and number of hooks between buoys explained 63% of the total deviance which was explained by the model. Despite the fluctuations and lack of accuracy of the estimate in 1998, the trend of the standardized CPUE series revealed a pattern of stability biomass, suggesting that longline fishing is not strongly affecting the stock of swordfish in the South Atlantic, a result that agrees with the conclusions and recommendations of ICCAT regarding the management of that stock. Then, an analysis from the longitudinal approach was conducted, it was considered a possibility the correlation between observations of the same vessel over time with the use of the theory of Generalised Estimating Equations (GEE) to extend GLMs the longitudinal data. For this analysis, we used data relating to fishing events held during the months of May, June and July 2001 (92 days) for eight leased ships from Spain. The GEE model with Poisson distribution and structure of correlation autoregressive of order 1 showed better fit and was used in the standardization of swordfish CPUE in monthly scale. The setting of this model revealed the significant influence of varying number of hooks between buoys, day and month on the swordfish CPUE. When set, from the same data base, a GLM model with the same distribution, linear predictor and binding function models used in the greenhouse, it was found that the parameter estimates are very different, and the adjust means GLM underestimates the standard errors, thus changing the significance of the parameters. As for the standardization of CPUE through the GLM and TG models, it was found that although both series present in

common a decreasing trend in relative abundance levels over the three months under standardized rates via GLM have more pronounced decrease, decreasing 58% from the first to the third month, to 46% for the GEE. This difference in slope of the series can promote an incorrect assessment of the stock of the species and misguided regulatory measures.

Keywords: CPUE; longline; GLM; longitudinal data; GEE.

Lista de Figuras

| | | |
|----|---|------|
| 1 | Esquema básico do espinhel de superfície. <i>Fonte: Fisheries Research e Development Corporation (FRDC)</i> | p.2 |
| 2 | Ilustração de um espadarte adulto (<i>Xiphias gladius</i>). <i>Fonte: Abid and Idrissi (2010)</i> | p.15 |
| 3 | Delimitações geográficas dos estoques de Espadarte no Oceano Atlântico. <i>Fonte: ICCAT (2014)</i> | p.18 |
| 4 | Ilustração das oito direções radiais utilizadas no cálculo das distâncias do local da pescaria à costa | p.28 |
| 5 | Histograma do número de espadartes capturados por eventos de pesca. | p.32 |
| 6 | Distribuição geográfica da CPUE nominal de espadarte (número de peixes/1000 anzóis). | p.33 |
| 7 | Diagramas de caixa relacionando a CPUE nominal de espadarte (número de peixes/1000 anzóis) e os fatores: (A) turno (am=antes do meio dia, pm=após meio dia), (B) trimestre e (C) ano. | p.33 |
| 8 | Diagramas de dispersão relacionando a CPUE nominal (número de espadartes/1000 anzóis) versus covariáveis: (A) número de anzóis entre boias, (B) latitude, (C) distância mínima à costa e (D) tempo de imersão do espinhel na água. A linha cinza é uma linha suavizadora (<i>Locally Weighted Scatterplot Smoother-LOWESS</i>), que mostra a tendência da CPUE. | p.34 |
| 9 | Envelope simulado para os Componentes do Desvio no modelo de Poisson. | p.36 |
| 10 | Envelope simulado para os Componentes do Desvio no modelo GLM binomial Negativo. | p.36 |
| 11 | Envelope simulado para os resíduos de Pearson Studentizado no modelo ZIP. | p.37 |

| | | |
|----|--|-------|
| 12 | Envelope simulado para os resíduos de Pearson Studentizado no modelo ZINB. | p. 38 |
| 13 | Gráficos de diagnósticos referentes ao modelo binomial negativo utilizado na padronização da CPUE de espadarte. (a) Componente do Desvio versus valores ajustados e (b) Distância de Cook versus a ordem das observações. | p. 39 |
| 14 | Efeito isolado dos termos principais do modelo binomial negativo sobre captura por unidade de esforço (CPUE) padronizada de espadarte (nº de peixes/1000 anzóis): A) distância mínima à costa (x1000 km), B) trimestre, C) latitude, D) tempo de imersão (em horas), E) turno e F) nº de anzóis entre boias. | p. 43 |
| 15 | Efeito da interação ano:distância mínima à costa sobre captura por unidade de esforço (CPUE) padronizada de espadarte (nº de peixes/1000 anzóis), derivado de um modelo binomial negativo. | p. 44 |
| 16 | Efeito da interação ano:trimestre sobre captura por unidade de esforço (CPUE) padronizada de espadarte (nº de peixes/1000 anzóis), derivado de um modelo binomial negativo. | p. 44 |
| 17 | Efeito da interação ano:turno sobre captura por unidade de esforço (CPUE) padronizada de espadarte (nº de peixes/1000 anzóis), derivado de um modelo binomial negativo. | p. 45 |
| 18 | Efeito da interação ano:nº de anzóis entre boias sobre captura por unidade de esforço (CPUE) padronizada de espadarte (nº de peixes/1000 anzóis), derivado de um modelo binomial negativo. | p. 45 |
| 19 | Efeito da interação distância mínima à costa:latitude sobre captura por unidade de esforço (CPUE) padronizada de espadarte (nº de peixes/1000 anzóis), derivado de um modelo binomial negativo. | p. 46 |
| 20 | Efeito da interação distância mínima à costa:tempo de imersão do espinhel sobre captura por unidade de esforço (CPUE) padronizada de espadarte (nº de peixes/1000 anzóis), derivado de um modelo binomial negativo. | p. 46 |
| 21 | Efeito da interação distância mínima à costa:turno sobre captura por unidade de esforço (CPUE) padronizada de espadarte (nº de peixes/1000 anzóis), derivado de um modelo binomial negativo. | p. 47 |

| | | |
|----|--|-------|
| 22 | Efeito da interação distância mínima à costa:nº de anzóis entre boias sobre captura por unidade de esforço (CPUE) padronizada de espadarte (nº de peixes/1000 anzóis), derivado de um modelo binomial negativo. | p. 47 |
| 23 | Efeito da interação trimestre:latitude sobre captura por unidade de esforço (CPUE) padronizada de espadarte (nº de peixes/1000 anzóis), derivado de um modelo binomial negativo. | p. 48 |
| 24 | Efeito da interação latitude:nº de anzóis entre boias sobre captura por unidade de esforço (CPUE) padronizada de espadarte (nº de peixes/1000 anzóis), derivado de um modelo binomial negativo. | p. 48 |
| 25 | Captura por unidade de esforço (CPUE) nominal e padronizada obtida através do modelo binomial negativo para o espadarte capturado pela frota atuneira brasileira no Atlântico Sul, de 1998 a 2006. | p. 49 |
| 26 | Comparação entre CPUEs padronizadas para o espadarte capturado pela frota atuneira brasileira no Atlântico Sul, de 1998 a 2006. | p. 50 |
| 27 | Histograma do número de espadartes capturados por eventos de pesca. | p. 65 |
| 28 | Distribuição geográfica da CPUE nominal de espadarte (número de peixes/1000 anzóis). | p. 66 |
| 29 | Perfis individuais dos barcos por instante de pesca, relativos à CPUE nominal de espadarte (nº de peixes capturados/1000 anzóis). Os instantes de 1 a 31 correspondem ao mês de maio, de 32 a 61 a junho e 61 a 92 a julho. A linha vermelha é uma linha suavizada (<i>Locally Weighted Scatterplot Smoother</i> - LOWESS), que mostra a tendência da CPUE. | p. 66 |
| 30 | Perfil médio dos barcos relativo ao esforço de pesca (nº de anzóis) (A), e relativo à CPUE (nº de espadartes/1000 anzóis) (B). | p. 67 |
| 31 | Gráfico do resíduo de Pearson referente ao modelo GEE de Poisson com estrutura de correlação AR(1) ajustado aos dados de capturas de espadarte. | p. 69 |

- 32 Captura por unidade de esforço (CPUE) nominal e padronizada escalonadas (nº de peixes/1000 anzóis) obtida através do modelo GEE de Poisson para o espadarte capturado pela frota atuneira brasileira no Atlântico Sul, de maio a julho, 2001. p. 70
- 33 Séries de CPUEs padronizadas escalonadas (nº de peixes/1000 anzóis), obtidas através dos modelos GEE e GLM de Poisson para o espadarte capturado pela frota atuneira brasileira no Atlântico Sul, de maio a julho, 2001. p. 72

Lista de Tabelas

- 1 Número de eventos de pesca por embarcação, 1998-2006. p.35
- 2 Comparativo dos modelos ajustados quanto ao logaritmo da verossimilhança (logLik), critério de informação de Akaike (AIC), critério de informação bayesiano (BIC), e o número de zeros preditos, sendo o número de zeros observados igual a 1110. p.38
- 3 Análise de deviância para o modelo binomial negativo, utilizado na padronização da CPUE de espadarte capturado pela frota espinheira brasileira, de 1998 a 2006. Para cada termo é apresentado o decréscimo provocado na deviância residual, graus de liberdade e o p-valor do teste. Siglas utilizadas: min - distância mínima à costa (x1000km), lat - latitude, hpb - nº de anzóis entre boias, tp.ime - tempo de imersão (horas) e significâncias: (***) - 0,001, (**) - 0,01. p.40
- 4 Estimativa dos coeficientes dos parâmetros do modelo binomial negativo, utilizado na padronização da CPUE de espadarte capturado pela frota espinheira brasileira. Siglas utilizadas: min - distância mínima à costa (x1000km), lat - latitude, hpb - nº de anzóis entre boias, tp.ime - tempo de imersão (horas) e significâncias: (***) - 0,001, (**) - 0,01, (*) - 0,05 e (.) - 0,1. p.41
- 5 Continuação da tabela 4. Estimativa dos coeficientes dos parâmetros do modelo binomial negativo, utilizado na padronização da CPUE de espadarte capturado pela frota espinheira brasileira. Siglas utilizadas: min - distância mínima à costa (x1000km), lat - latitude, hpb - nº de anzóis entre boias, tp.ime - tempo de imersão (horas) e significância: 0 ***, 0.001 **, 0.01 *, 0.05 p.42
- 6 Estrutura de dados básica para estudos com desenho longitudinal. . . p.55

- 7 Estimativas, erros padrão naive e p-valores relativo ao teste de Wald dos parâmetros do modelo GEE de Poisson com estrutura de correlação ar1. Siglas utilizadas: min - distância mínima à costa (x1000km), lat - latitude, hpb - n° de anzóis entre boias, tp.ime - tempo de imersão (horas) e significância: 0 ***, 0.001 **, 0.01 *, 0.05 : p. 68
- 8 Estimativas, erros padrão e p-valores relativo aos parâmetros do modelo GLM de Poisson, ignorando as correlações entre as observações intrabarcos. Siglas utilizadas: min - distância mínima à costa (x1000km), lat - latitude, hpb - n° de anzóis entre boias, tp.ime - tempo de imersão (horas) e significância: 0 ***, 0.001 **, 0.01 *, 0.05 : p. 71

Sumário

| | | |
|----------|---|-------|
| 1 | Introdução | p. 1 |
| 2 | Revisão da Literatura | p. 7 |
| 2.1 | Métodos de padronização da CPUE | p. 7 |
| 2.1.1 | Modelo linear generalizado (GLM) | p. 9 |
| 2.1.1.1 | GLM lognormal | p. 12 |
| 2.1.1.2 | Delta GLM | p. 12 |
| 2.1.2 | Modelo aditivo generalizado (GAM) | p. 13 |
| 2.1.3 | Modelo linear generalizado misto (GLMM) | p. 14 |
| 2.2 | Espécie Analisada | p. 15 |
| 3 | Modelagem da CPUE via GLMs para dados de contagens | p. 19 |
| 3.1 | Modelo de Poisson | p. 19 |
| 3.1.1 | Sobredispersão | p. 20 |
| 3.1.2 | Teste para Sobredispersão no Modelo de Poisson | p. 21 |
| 3.2 | Modelo Binomial Negativo | p. 22 |
| 3.3 | Modelos Inflacionados de Zeros | p. 23 |
| 3.3.1 | Excesso de Zeros | p. 23 |
| 3.3.2 | Modelo de Poisson Inflacionado de Zeros | p. 23 |
| 3.3.3 | Modelo Binomial Negativo Inflacionado de Zeros | p. 24 |
| 3.3.4 | Teste Comparativo entre os Modelos Inflacionados de Zeros e suas Versões Usuais | p. 25 |
| 3.4 | Materiais e Métodos | p. 26 |

| | | |
|----------|--|--------------|
| 3.4.1 | Descrição dos dados | p. 26 |
| 3.4.2 | Variável Resposta | p. 27 |
| 3.4.3 | Variáveis explicativas | p. 27 |
| 3.4.4 | Ajuste dos Modelos | p. 29 |
| 3.4.4.1 | Inclusão das variáveis explicativas e simplificação dos modelos | p. 29 |
| 3.4.4.2 | Métodos de Diagnóstico | p. 30 |
| 3.4.4.3 | Padronização da CPUE | p. 31 |
| 3.4.4.4 | Procedimento Computacional | p. 31 |
| 3.5 | Resultados | p. 32 |
| 3.5.1 | Análise Exploratória | p. 32 |
| 3.5.2 | Ajuste dos Modelos | p. 34 |
| 3.5.3 | Avaliação dos Modelos | p. 35 |
| 3.5.4 | Modelo utilizado na padronização da CPUE do espadarte . . . | p. 37 |
| 3.5.5 | Gráficos de diagnóstico do modelo binomial negativo | p. 38 |
| 3.5.6 | Análise de deviância e estimativas dos parâmetros do modelo binomial negativo | p. 39 |
| 3.5.7 | CPUE padronizada | p. 48 |
| 3.6 | Discussão | p. 49 |
| 3.7 | Conclusão | p. 52 |
| 4 | Modelagem da CPUE de espadarte sob o enfoque longitudinal | p. 53 |
| 4.1 | Introdução | p. 53 |
| 4.2 | Dados Longitudinais | p. 54 |
| 4.3 | Equações de Estimação Generalizadas | p. 56 |
| 4.3.1 | Estimação dos parâmetros | p. 59 |
| 4.4 | Materiais e Métodos | p. 60 |

| | | |
|----------|---|------|
| 4.4.1 | Descrição dos dados | p.60 |
| 4.4.2 | Variáveis consideradas | p.60 |
| 4.4.3 | Modelos GEE propostos | p.61 |
| 4.4.4 | Critério para seleção da matriz de correlação de trabalho . . . | p.62 |
| 4.4.5 | Seleção e ajuste dos modelos | p.62 |
| 4.4.6 | Técnicas de Diagnóstico | p.63 |
| 4.4.7 | Procedimento Computacional | p.64 |
| 4.4.8 | Tratamento de dados ausentes | p.64 |
| 4.4.9 | Padronização da CPUE | p.64 |
| 4.5 | Resultados | p.65 |
| 4.5.1 | Análise Exploratória | p.65 |
| 4.5.2 | Modelo GEE utilizado na padronização da CPUE do espadarte | p.67 |
| 4.5.3 | Análise dos Resíduos | p.69 |
| 4.5.4 | CPUE padronizada via GEE | p.70 |
| 4.5.5 | Modelagem via GLM | p.71 |
| 4.5.6 | CPUE padronizada via GEE e GLM | p.72 |
| 4.6 | Discussão | p.73 |
| 4.7 | Conclusão | p.75 |
| 5 | Considerações Finais | p.76 |
| | Referências | p.78 |

1 Introdução

A pesca extrativa marinha desempenha um papel essencial no sustento de milhões de pessoas em todo o mundo. Ao longo dos anos de 2006-2011, a produção mundial de pescado, proveniente da pesca extrativa marinha, oscilou entre 77 e 80 milhões de toneladas por ano, representando, em 2011, cerca de 51,2% do total produzido de 154 milhões de toneladas, incluindo as quatro modalidades de pesca: extrativa marinha, extrativa continental, aquicultura marinha e aquicultura continental (FAO, 2012). O Brasil contribuiu com 553.670 t em 2011, representando 0,7% da pesca extrativa marinha, ocupando a 23^o colocação no ranking mundial de produção de pescados por pesca extrativa (MPA, 2013). No ano de 2011, o Brasil apresentou exportações de US\$ 271.193.147 milhões e importações de US\$ 1.262.888.212, ou seja, um déficit de aproximadamente US\$ 991 milhões. No mundo, o consumo médio global aumentou de uma média de 9,9 kg (peso vivo equivalente) na década de 1960 para 18,8 kg em 2010, e as previsões para 2030 apontam que seja alcançado um valor médio de 22,5 kg/ano, o que vai gerar uma demanda adicional de 100 milhões de toneladas ao ano para consumo (FAO, 2012).

No Brasil, nos 8500 km de extensão de costa e 3,5 milhões de quilômetros quadrados de mar da zona econômica exclusiva (ZEE), as informações sobre a dinâmica espaço-temporal das embarcações de pesca permissionadas pelo Ministério da Pesca e Aquicultura (MPA), nas categorias artesanal ou industrial (nacional ou arrendada), são controladas pelo Programa Nacional de Rastreamento de Embarcações Pesqueiras por Satélite (PREPS). Sendo que a pesca artesanal, limitada à áreas costeiras (rios, lagos, lagoas, estuários, enseadas, baías e praias), é realizada com embarcações de pequeno porte (menores que 8 metros), e não ultrapassa 20 metros de profundidade. A extração do recurso pesqueiro fica limitada localmente. Quanto à pesca industrial, as embarcações utilizadas são de médio e grande porte (maiores que 8 metros), com grande autonomia de mar, possuindo a bordo sistemas de conservação do pescado, podendo ficar alguns dias a meses em busca do recurso pesqueiro, e se utilizar de diversas

áreas da plataforma e talude continental para extração do mesmo. Quando o pescado é desembarcado, podem ser distribuídos a diversas áreas de comércio, inclusive para a exportação (MPA, 2012).

Uma das modalidades pesqueiras de maior importância econômica é a de espinhel de superfície (*long-line*) (Figura 1), direcionada à captura de peixes oceânicos de grande porte, tais como os atuns e espécies afins (i.e. espadarte, agulhões e tubarões) (MPA, 2012). De maneira geral, o espinhel de superfície consiste em uma linha de náilon monofilamento de alta resistência, chamada de linha principal, que se estende por quilômetros abaixo da superfície e é suspensa por boias. Na linha principal também são fixadas através de grampos de metal ("snaps"), 800 a 1200 linhas secundárias com anzóis iscados. O conjunto afunda lentamente até profundidades que variam de 45 a 300 metros (KOTAS *et. al.*, 2005 e DOMINGO *et. al.*, 2014).

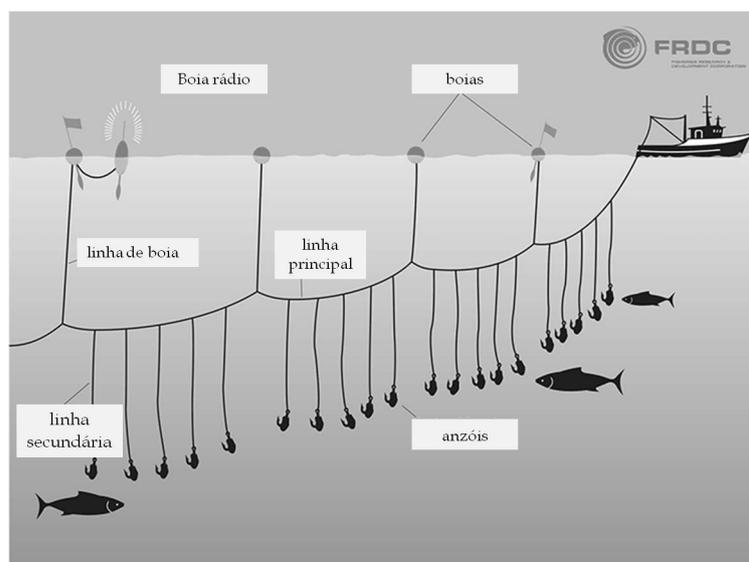


Figura 1: Esquema básico do espinhel de superfície. Fonte: Fisheries Research e Development Corporation (FRDC)

Devido à ampla área de distribuição e a capacidade de migração, os grandes peixes pelágicos são capturados por diversos países, com diferentes artes de pesca, o que faz com que a avaliação e o manejo dos estoques só sejam possíveis através de esforços conjuntos de várias nações. Para assegurar a produtividade dos recursos pesqueiros, a biomassa dos estoques tem que ser mantida nos níveis bioecológicos satisfatórios, que evitem riscos de sobrepesca. Neste sentido, para as pescarias que atuam sobre estoques pesqueiros que ocorrem nas águas jurisdicionais (zona econômica exclusiva-ZEE) de vários países e, inclusive, na zona de alto mar, as medidas de regulamentação para a conservação e ordenamento destes, são estabelecidas no âmbito de comissões

internacionais de pesca, que congregam todos os países com interesses nas pescarias.

O espadarte (*Xiphias gladius*), também conhecido como meka, é um dos peixes pelágicos mais visados pela indústria pesqueira mundial, devido à sua carne de ótima qualidade, à crescente demanda no cenário internacional, e ao seu alto valor de mercado. No Oceano Atlântico, sua exploração é controlada pela Comissão Internacional para a Conservação dos Atuns do Atlântico (International Commission for the Conservation of Atlantic Tunas - ICCAT). Os países signatários, incluindo o Brasil, são responsáveis pela coleta, análise e apresentação dos dados referentes às capturas e aos esforços de pesca em sua área de atuação. A hipótese mais aceita atualmente é a de que há três estoques distintos, um no Atlântico Norte, outro no Atlântico Sul, separados na latitude 5°N, e um terceiro no Mediterrâneo (ICCAT, 2014). Anualmente, a ICCAT faz recomendações, com o intuito de assegurar a sustentabilidade dos recursos explorados, incluindo a definição de tamanhos mínimos para a captura, e o estabelecimento de quotas de captura para os países membros. No Atlântico Sul as capturas de espadarte atingiram seu ápice em 1995 com 21.930 toneladas, apresentando a partir daí, uma tendência de queda, atingindo 14.027 toneladas em 1998, 12.633 toneladas em 2003 e 10.178 em 2012, sendo 98% destas, proveniente de espinhéis de superfície (ICCAT, 2014a). Esta redução foi parcialmente atribuída a medidas regulatórias da ICCAT, a mudança de frotas para outros oceanos, e a mudanças nas espécies-alvo (ICCAT, 2014). Devido ao crescente interesse comercial sobre o espadarte, e ao aumento dos esforços de pesca, após uma avaliação técnica a ICCAT recomendou em 2009, uma redução da quota de espadarte no Atlântico Sul de 17.000 t/ano para 15.000 t/ano, que, segundo a comissão, deve manter a biomassa do estoque em um nível saudável. Esta quota foi mantida até 2013 e postergada até 2016, segundo as recomendações da ICCAT [Rec. 12-01] e [Rec. 13-03], respectivamente (ICCAT, 2014). O Brasil tem quota de 3.940 t/ano desde 2012, a segunda maior no Atlântico Sul, entretanto, não tem capturado mais que 3.000 t (ICCAT, 2014a).

As medidas de ordenamento e recomendações científicas para o manejo das pescarias são baseadas em análises de avaliação de estoques, comumente feitas a partir de índices relativos de abundância populacional, calculados com base em dados de captura e esforço de pesca, obtidos de pescarias comerciais (HILBORN e WALTERS, 1992). A taxa calculada pela razão entre a captura e esforço, é então denominada de captura por unidade de esforço (CPUE). Um dos pressupostos mais comuns na ciência pesqueira clássica é que a CPUE, utilizada como índice relativo de abundância, é proporcional à abundância absoluta (N) (i.e, número de peixes) a partir de uma equação que na forma

mais simples é linear e é dada por $CPUE = q \times N$, em que q é denominado de coeficiente de capturabilidade. Este termo que expressa a proporcionalidade entre a CPUE e a abundância (N), pode ser afetado por diversos fatores relacionados à eficiência de pesca, como: distribuição espacial e temporal da espécie e do esforço de pesca, variações na experiência da tripulação, estratégia de pesca, e configuração do aparelho de pesca (HILBORN e WALTERS, 1992; ARREGUÍN-SANCHES, 1996). Assim, alterações na CPUE podem estar relacionadas tanto a mudanças na abundância do estoque quanto a mudanças no coeficiente de capturabilidade. Portanto, para tornar possível a utilização da CPUE como índice relativo de abundância confiável, faz-se necessário a identificação e remoção das variações na CPUE decorrentes de fatores que não estejam diretamente relacionados com flutuações na abundância real do estoque. A este processo dá-se o nome de "padronização da CPUE" (MAUNDER e PUNT, 2004).

Vários métodos estatísticos de regressão têm sido desenvolvidos para obtenção de séries de CPUEs padronizadas, úteis na avaliação da biomassa, e para a tomada de decisões para o manejo e sustentabilidade dos estoques pesqueiros. No entanto, nas últimas décadas, os modelos lineares generalizados (*Generalized Linear Models* - GLM) (MCCULLAGH e NELDER, 1989; DOBSON, 2002), têm sido os mais amplamente utilizados. A aplicação dos GLMs possibilita a separação dos efeitos de muitos fatores associados ao coeficiente de capturabilidade (q) que causam variabilidade na CPUE, permitindo a remoção dos mesmos, e a produção de séries de CPUE padronizadas, que refletiriam as variações temporais (i.e, anuais ou mensais) da abundância (MAUNDER e PUNT, 2004).

Na aplicação do GLM assume-se que a variável resposta (número de peixes capturados ou taxa de captura) segue uma distribuição de probabilidade da família exponencial, e que a amostra é composta por observações independentes. Contudo, os dados obtidos de pescarias comerciais, utilizados na padronização da CPUE, são advindos das mesmas embarcações em diferentes eventos de pesca realizados ao longo do tempo, caracterizando assim um cenário longitudinal, e embora as embarcações possam ser consideradas independentes entre si, espera-se que exista uma correlação não nula entre as observações aninhadas em uma mesma embarcação (i.e, intrabarcos) em ocasiões distintas, levando à violação da suposição de independência. DIGGLE *et al.* (2002) e BALLINGER (2004) mostraram que, ao modelar dados longitudinais com o GLM, em que supõe-se a independência entre observações, tem-se como consequência, inferências incorretas dos parâmetros da regressão devido a erros padrões subestimados e estimadores ineficientes. . Estes resultados evidenciam a necessidade de que modelos

mais adequados que o GLM sejam utilizados na análise de dados longitudinais. A análise de dados longitudinais tem uma longa história na literatura estatística e tem sido objeto de inúmeros textos, dentre os quais podemos destacar LIANG e ZEGER (1986), ZEGER e LIANG (1986), VONESH e CHINCHILLI (1997), SINGER e ANDRADE (2000), DIGGLE *et al.* (2002), MOLENBERGHS e VERBEKE (2005), WEISS (2005), HEDEKER e GIBBONS (2006), VENEZUELA *et al.* (2007) e FITZMAURICE, *et al.* (2009).

Entretanto, em estudos envolvendo capturas de peixes pelágicos por espinhéis de superfície, como a espécie espadarte, a modelagem sob o enfoque longitudinal tem sido negligenciada. Provavelmente, isto se deve a algumas particularidades: Trata-se de um fenômeno observacional e não experimental, em que as repetições ocorrem em tempos não controlados, gerando dados não balanceados e a existência de observações omissas (*missing*), visto que as embarcações, após realização de cada viagem de pesca, realizando lances de pesca diários, retornam ao porto de origem para descarregar o pescado e se preparar para uma nova viagem, interrompendo a atividade durante vários dias. Além disso, muitas das variáveis explicativas, que influenciam no rendimento das pescarias, assumem valores que podem mudar com o tempo (tempo-dependentes), ou com a embarcação. Diante deste cenário peculiar, utilizamos a teoria das Equações de Estimção Generalizadas (Generalized Estimating Equations - GEE), proposta por LIANG e ZEGER (1986) e ZEGER e LIANG (1986), para estender os GLMs aos dados longitudinais de pesca e incorporar a suposta correlação existente entre as observações intrabarcos, realizadas ao longo do tempo, na análise. Nos trabalhos mais recentes envolvendo a padronização de CPUE para o espadarte do Atlântico Sul, utilizaram-se o GLM delta lognormal (HAZIN *et al.*, 2007), o lognormal (MOURATO *et al.*, 2007), o quasi-poisson (HAZIN *et al.*, 2008), o quasi-poisson e Tweedie (HAZIN *et al.*, 2010), e modelos lineares generalizados mistos, usando o método delta lognormal (HAZIN *et al.*, 2014). Apesar dos diversos estudos, considerando os conflitos entre as CPUEs padronizadas disponíveis para o espadarte (tendências diferentes e alta variabilidade), recomenda-se que mais pesquisas sejam realizadas para diminuir a elevada incerteza no cálculo dos índices relativos de abundância do estoque do Atlântico Sul (ICCAT, 2014). Portanto, é de suma importância que o Brasil cumpra com o seu papel de signatário da ICCAT e desenvolva trabalhos científicos para correção desses desvios, buscando fornecer séries históricas de CPUE livres dos vícios que causam variabilidade não relacionada com a abundância das diferentes espécies controladas pela Comissão. Em razão do exposto, a partir do Banco Nacional de Dados de Atuns e Afins (BNDA),

que contém informações dos mapas e de observadores de bordo das embarcações espinheiras brasileiras (nacional e arrendada) atuantes no Atlântico Sul, objetiva-se com este trabalho:

1. Estimar a abundância relativa do espadarte, com a aplicação de GLMs para dados de contagens, levando em conta a possível presença de sobredispersão e excesso de zeros nos dados;
2. Produzir uma série de CPUE padronizada para o estoque de espadarte do Atlântico Sul, a partir do melhor dentre os GLMs propostos, e confrontá-la com as séries obtidas nos trabalhos realizados anteriormente;
3. Utilizar a teoria das equações de estimação generalizadas (GEE) para estender os GLMs e incorporar a correlação existente entre as observações intrabarcos, realizadas ao longo do tempo;
4. Produzir uma série de CPUE padronizada para o espadarte a partir dos modelos GEE e confrontar os resultados com os obtidos com o GLM;

A expectativa a partir do desenvolvimento deste trabalho, é de que as informações geradas possam colaborar com a ICCAT na diminuição de incertezas sobre os índices de abundância, na avaliação da biomassa e para a sustentabilidade da exploração do estoque de espadarte no Atlântico Sul. Convém ressaltar que, apesar deste estudo ter como foco o espadarte, os métodos aqui desenvolvidos podem ser utilizados para produzir índices de relativos de abundância para outras espécies capturadas por espinhéis de superfície.

2 Revisão da Literatura

2.1 Métodos de padronização da CPUE

As medidas de ordenamento e recomendações científicas para o manejo da pesca são baseadas nos resultados de algumas técnicas de avaliação de estoques, geralmente feita a partir de índices de abundância baseados em dados de captura e esforço de pesca obtidos de pescarias comerciais (HILBORN e WALTERS, 1992). Considera-se que a mortalidade gerada pelo esforço de pesca é proporcional à intensidade do seu uso, de modo que, a captura (em número ou peso) dos peixes retirados por unidade do esforço de pesca, ou seja, a CPUE, é uma medida relativa da abundância aparente do estoque.

Um dos pressupostos mais comuns na ciência pesqueira clássica é que a abundância relativa de um estoque pode ser estimada a partir do conhecimento da CPUE a que um recurso está submetido em uma dada área e período, através da equação $CPUE = qN$, onde N representa uma medida de abundância, como, por exemplo a densidade de um estoque e q é o coeficiente de capturabilidade.

A partir de uma ampla revisão da literatura sobre capturabilidade, ARREGUIN-SANCHEZ (1996) concluiu que o coeficiente de capturabilidade pode ser afetado por diversos fatores relacionados a eficiência de pesca, como:

- dinâmica espacial e temporal das embarcações;
- padrão de comportamento das espécies em relação aos fatores abióticos (ambientais);
- variações na experiência da tripulação;
- qualidade e quantidade do esforço de pesca, e, estratégia de pesca.

Segue-se que as alterações na CPUE podem estar relacionadas tanto a mudanças na densidade do estoque quanto a mudanças no coeficiente de capturabilidade q . Portanto, para tornar possível a utilização da CPUE como índice relativo de abundância

confiável, faz-se necessário a identificação e remoção das variações na CPUE decorrentes de fatores que não esteja diretamente relacionada com flutuações na abundância real do estoque. A este processo dá-se o nome de "padronização da CPUE".

Os perigos associados à utilização de CPUEs nominais (não padronizadas) nas avaliações de estoques pesqueiros já são conhecidos há muitos anos, e vários métodos para a padronização de dados de captura e esforço foram desenvolvidos. BEVERTON e HOLT (1957) desenvolveram um método que envolve a seleção de uma embarcação-padrão para determinação do poder de pesca relativo de todas as outras embarcações que compõem uma dada frota, sendo definido por:

$$PPR_i = \frac{C_i/E_i}{C_s/E_s}, \quad (2.1)$$

onde PPR_i é o poder de pesca relativo de uma embarcação i , C_i o total de capturas da embarcação i , C_s o total de capturas da embarcação padrão, ambos em um dado período em que embarcação i esteve presente com a embarcação padrão na pescaria, E_i o esforço de pesca da embarcação i e E_s o esforço de pesca da embarcação padrão. Assim, a CPUE padronizada para o ano t , I_t , foi definida por

$$I_t = \frac{\sum_i C_{t,i}}{\sum_i (PPR_i E_{t,i})}, \quad (2.2)$$

onde $C_{t,i}$ é a captura da embarcação i no ano t e $E_{t,i}$ o esforço de pesca da embarcação i no ano t . Embora simples de aplicar, segundo MAUNDER E PUNT (2004), tal abordagem não é facilmente generalizada para lidar com múltiplos fatores como mês e área, e quando não há nenhuma embarcação que esteja em atividade durante muitos anos e que possa ser utilizada como embarcação padrão. Além disso, não é fácil determinar a precisão das estimativas das CPUEs padronizadas, necessárias ao se aplicar muitos dos métodos de avaliação de estoque.

No início da década de 1980, foram desenvolvidos outros métodos para a padronização envolvendo o ajuste de modelos estatísticos para dados de captura e esforço de pesca, quantificando e removendo o efeito de variáveis que influenciam as taxas de captura. Os primeiros exemplos desses métodos foram desenvolvidos por GAVARIS (1980) e KIMURA (1981). GAVARIS (1980) estendeu o modelo multiplicativo de ROBSON (1966) aplicando um modelo de análise de variância para o logaritmo natural da CPUE, utilizando somente variáveis explicativas categóricas e assumindo erro aleatório com distribuição lognormal com variância constante e independência entre as observações. KIMURA (1981) estendeu a abordagem de GAVARIS (1980) incluindo no modelo variáveis explicativas contínuas além de categóricas. A partir de então, têm-se

visto uma proliferação de novos métodos de padronização da CPUE, sendo a maioria, de certa forma, uma extensão destes (por exemplo, COOKE e BEDDINGTON, 1984; RICHARDS e SCHNUTE, 1986). A escolha dentre tais métodos deve ser baseada numa avaliação dos pressupostos subjacentes aos modelos e uso de testes estatísticos e diagnósticos apropriados.

Em geral, o uso de modelos lineares clássicos não é apropriado, visto que é bastante comum encontrar distribuições de CPUE com desvio padrão aproximadamente igual à média e distribuições com assimetria positiva, assim como distribuições em que a variância da CPUE aumenta proporcionalmente com a média, caracterizando tipos de distribuição que não são normalmente distribuídas. Em contra-partida, a técnica de modelos lineares generalizados (*Generalized linear models* - GLMs), desenvolvida por NELDER e WEDDERBURN (1972), tem permitido a generalização ou flexibilização dos modelos lineares clássicos, estendendo sua aplicação a um conjunto mais heterogêneo de dados que inclui tanto as variáveis contínuas como as variáveis discretas, variando desde a distribuição de probabilidade normal à distribuição binomial negativa. A aplicação dos GLMs pode separar os efeitos de muitos fatores que causam variabilidade na CPUE, sendo úteis para a produção de séries de CPUE padronizadas (QUINN e DERISO, 1999 e VENABLES e DITCHMOND, 2004).

2.1.1 Modelo linear generalizado (GLM)

Os modelos lineares generalizados são ferramentas poderosas na análise de dados onde o interesse é o estudo da relação entre uma variável resposta, medida em escala contínua ou discreta, em função de diferentes variáveis preditoras (quantitativas e/ou qualitativas). De maneira geral, os GLMs podem ser descritos da seguinte forma:

$$Y_i \sim f(y_i; \theta_i, \phi), \quad i = 1, \dots, n, \quad (2.3)$$

$$E(Y_i | x_i) = \mu_i = g^{-1}(\mathbf{x}_i^T \boldsymbol{\beta}), \quad (2.4)$$

sendo Y_i variáveis aleatórias independentes que compartilham da mesma distribuição probabilística pertencente família exponencial na forma

$$f(y_i; \theta_i, \phi) = \exp \{ a(\phi)^{-1} [y_i \theta_i - b(\theta_i)] + c(y_i, \phi) \}, \quad (2.5)$$

para funções $a(\cdot)$, $b(\cdot)$ e $c(\cdot)$ conhecidas, onde $E(Y_i) = \mu_i = b'(\theta_i)$, $Var(Y_i) = \phi b''(\theta_i) = \phi V_i$, em que $V_i = V(\mu_i) = \frac{d\mu_i}{d\theta_i}$ é denominada de função de variância e depende unicamente da média μ_i , sendo $\phi > 0$ um parâmetro de dispersão e θ_i o parâmetro canônico

da função (2.5); $\eta_i = \mathbf{x}_i^T \boldsymbol{\beta}$ é o preditor linear, onde $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^T$, $p < n$, é um vetor de parâmetros desconhecidos a serem estimados, $\mathbf{x}_i = (x_{i1}, \dots, x_{ip})^T$ representa os valores de p variáveis explicativas e $g(\cdot)$ é uma função monótona e diferenciável, denominada função de ligação, que liga a média da variável resposta ao preditor linear η_i . Usando a notação $\boldsymbol{\eta} = \mathbf{X}\boldsymbol{\beta}$, $\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_n)^T$ seria a matriz de delineamento do modelo (MCCULLAGH e NELDER, 1989). Se um parâmetro tem valor conhecido, o termo correspondente na estrutura linear é chamado offset (DOBSON, 2002).

O Algoritmo de ajuste de um modelo linear generalizado é o dos mínimos quadrados iterados ponderados (Iteratively Reweighted Least-Squares IRLS), mediante o qual são obtidos os estimadores de verossimilhança dos parâmetros $\boldsymbol{\beta}$ do preditor linear. NELDER e WEDDERBURN (1972), propuseram também um processo iterativo para a estimação dos parâmetros e introduziram o conceito de deviência que tem sido largamente utilizado na avaliação da qualidade do ajuste dos GLM, bem como no desenvolvimento de resíduos e medidas de diagnóstico. Daí em diante, surgiram novas abordagens e modelos para os diversos cenários de dados.

MAUNDER e PUNT (2004) apresentam uma revisão dos métodos e procedimentos utilizados na padronização da captura e do esforço de pesca e ressaltam que a escolha de distribuição amostral para a variável resposta dentre as distribuições da família exponencial (entre as quais se incluem as distribuições normal, binomial, Poisson, binomial negativa e Gamma), deve-se levar em conta a natureza do processo que gerou os dados que está sendo modelado. Por exemplo, uma distribuição discreta, tal como a de Poisson ou binomial negativa, pode ser a distribuição mais adequada se a captura é registrada em número de peixes. No entanto, se a variável resposta é a captura em peso ou a taxa de captura uma distribuição contínua pode ser mais apropriada.

Comumente, mais de um modelo pode ser utilizado para descrever um mesmo fenômeno. Deste modo, é necessário fazer a seleção do "melhor" modelo, dentre aqueles que foram ajustados, para explicar o fenômeno sob estudo. Como não há modelo verdadeiro, comparamos os modelos em termos relativos. As medidas comumente utilizadas na comparação e seleção de modelos GLM, são:

- **Deviância** - Traduzida como desvio por CORDEIRO (1986), a deviência é uma medida de discrepância entre os valores observados (modelos saturado) e os valores estimados pelo modelo corrente, proposta por NELDER e WEDDERBURN (1972), para avaliar a qualidade do ajustamento de um modelo, com expressão

dada por

$$D(\mathbf{y}; \hat{\boldsymbol{\mu}}) = 2\{l(\mathbf{y}; \mathbf{y}) - l(\hat{\boldsymbol{\mu}}; \mathbf{y})\} \quad (2.6)$$

onde $l(\hat{\boldsymbol{\mu}}; \mathbf{y})$ é o logaritmo da função de verossimilhança do modelo sob investigação, com p parâmetros, avaliado na estimativa de máxima verossimilhança $\hat{\boldsymbol{\beta}}$; e $l(\mathbf{y}; \mathbf{y})$ é o logaritmo da função de verossimilhança do modelo saturado (que contém tantos parâmetros quanto observações, ou seja, n parâmetros). O melhor modelo ajustado é aquele com a máxima $l(\boldsymbol{\mu}; \mathbf{y})$, ou seja, com menor deviência. Um valor pequeno para a deviência indica que, para um número menor de parâmetros, obtemos um ajuste tão bom quanto o ajuste com o modelo saturado. Sob condições de regularidade satisfeitas pela família exponencial, a deviência $D(\mathbf{y}; \hat{\boldsymbol{\mu}})$, converge para uma distribuição qui-quadrado com $(n - p)$ graus de liberdade. Portanto, se o valor da deviência estiver próximo dos graus de liberdade, o modelo pode ser considerado como adequado. Contudo, a deviência só pode ser usada para comparar modelos encaixados, ou seja, com mesma distribuição de probabilidade e mesma função de ligação, mas o componente linear do modelo mais simples é um caso especial do componente linear do modelo mais geral.

- **Critério de Informação de Akaike (AIC)** - O AIC (AKAIKE, 1973; BURNHAM e ANDERSON, 2002) utiliza a verossimilhança como medida de discrepância relativa dos modelos. No entanto, para balancear o viés e a variância dos modelos, o AIC penaliza a verossimilhança pelo número de parâmetros dos mesmos, e é dado por

$$AIC = -2l(\hat{\boldsymbol{\beta}}) + 2p, \quad (2.7)$$

em que p denota o número de parâmetros do modelo investigado e $l(\hat{\boldsymbol{\beta}})$ o logaritmo da sua função de verossimilhança avaliada no seu máximo. Nesse caso, o critério consiste em encontrarmos o modelo que tenha o menor AIC, ou seja, um modelo parcimonioso, ou em outras palavras, que esteja bem ajustado e tenha um número reduzido de parâmetros.

- **Critério de Informação Bayesiano (BIC)**- Segundo WAGENMAKERS e FARRELL (2004), quando o número de observações n cresce muito, a probabilidade de que o AIC recupere um verdadeiro modelo com baixa dimensão, não se aproxima da unidade. Neste caso, uma alternativa para o AIC é a utilização do

BIC, proposto por SCHWARZ (1978) que é dado por:

$$BIC = -2l(\hat{\beta}) + p\log(n), \quad (2.8)$$

em que p denota o número de parâmetros do modelo investigado e $l(\hat{\beta})$ o logaritmo da sua função de verossimilhança avaliada no seu máximo e n é o número de observações. Tanto o AIC como o BIC não possuem nenhuma distribuição associada para proceder um teste formal e podem ser utilizados na comparação entre não encaixados (KUHA, 2004).

2.1.1.1 GLM lognormal

Uma abordagem aplicada frequentemente nos estudos de padronização das taxas de captura tem sido a de se considerar GLM lognormais, em que a variável resposta (CPUE), depois que é submetida a uma transformação logarítmica, atende às suposições do modelo de regressão linear, seguindo uma distribuição normal, com variância igual e constante. O procedimento usual consiste em adicionar uma constante às taxas de captura, antes de se aplicar a transformação logarítmica, evitando-se assim o problema de se ter logaritmos indefinidos para as taxas de captura nulas.

Em diversos estudos sobre padronização de dados de CPUE que incluíram os registros com captura zero este método tem sido utilizado com sucesso para se conseguir normalizar a CPUE e estabilizar a variância, como por exemplo, na análise realizada por ALLEN e PUNSLY (1984) com dados de CPUE da albacora lage, capturada com redes de cerco no oceano Pacífico oriental. No entanto, outros estudos têm demonstrado que tal transformação nem sempre é suficiente para se atingir a normalidade dos dados. Neste sentido, PUNSLY e NAKANO (1992) realizaram análise de padronização da CPUE da albacora lage da pesca com espinhel do oceano Pacífico oriental concluindo que a alta frequência de capturas zero foi um dos principais motivos que resultaram na falta de ajuste entre o logaritmo da CPUE e a distribuição normal.

2.1.1.2 Delta GLM

Uma alternativa bastante utilizada, quando se tem uma elevada proporção de capturas nulas, é a utilização do modelo Delta (LO et al., 1992). Nesse modelo constrói-se, separadamente, um modelo GLM para as capturas positivas assumindo-se uma distribuição lognormal ou gamma dos erros, e pondera-se o resultado pela proporção de capturas positivas, assumindo-se uma distribuição binomial. O índice de CPUE pa-

dronizado é o produto dos coeficientes estimados desses dois modelos.

ORTIZ et al. (2000) ao aplicarem o GLM lognormal para estimar as taxas de captura de espécies de peixes que ocorrem como by catch, concluíram que o modelo gera estimativas dependentes da seleção da constante utilizada, devido ao fato de que os dados das CPUEs observadas das espécies não cumpriam com as suposições associadas ao modelo, em particular, em função da elevada proporção de capturas nulas nos dados analisados. Contudo, ao aplicar o modelo delta lognormal aos mesmos dados observaram que este modelo se ajustava melhor a estrutura dos dados, além de se evitar o problema da definição da constante a adicionar à CPUE antes da sua transformação logarítmica.

2.1.2 Modelo aditivo generalizado (GAM)

Embora a grande maioria dos trabalhos científicos de padronização de captura e esforço pesqueiro se baseie no GLM, outras técnicas também têm sido aplicadas. Mais recentemente o modelo aditivo generalizado (*Generalized additive model* - GAM) (HASTIE e TIBSHIRANI, 1990) tem sido usado na ciência pesqueira (por exemplo, BIGELOW et al., 1999; BELLIDO et al., 2001; WALSH e KLEIBER, 2001; WALSH et al., 2005).

Os GLMs assumem que a relação da variável resposta com as variáveis explicativas é de natureza linear. No entanto, tal pressuposto pode falhar quando a relação de uma variável explicativa com a CPUE ocorre de forma não-linear, sendo os GAMs ferramentas estatísticas úteis para mensurar os efeitos não-lineares entre as diversas variáveis com a CPUE. Os GAMs correspondem a generalizações de modelos de regressão não-paramétricos e são menos restritivos quanto à distribuição da variável resposta (HASTIE et al., 2001). Sua estrutura geral é formada por: a) a variável resposta (normalmente, mas não exclusivamente, a CPUE); b) uma função suavizadora (*smoother function*) como *loess* ou *cubic spline* que ajusta o efeito não linear das variáveis sobre a variável resposta e c) o erro do modelo. A principal diferença entre os GAMs e os modelos de regressão convencionais está na incorporação de efeitos não-lineares complexos de múltiplas fontes.

2.1.3 Modelo linear generalizado misto (GLMM)

Nos modelos lineares generalizados considera-se apenas o estudo de variáveis com efeitos fixos, modelando-se apenas a média dos dados. Uma extensão natural são modelos que se ajustem a dados obtidos a partir de fenômenos em que os níveis de um fator foram selecionados de uma população de níveis, isto é, são aleatórios. O Modelo linear generalizado misto (*Generalized linear mixed model* - GLMM), tem a habilidade de modelar não somente a média dos dados, ma também sua variância, introduzindo termos adicionais na relação entre a resposta média e as variáveis explicativas, que sejam quantidades aleatórias ao invés de parâmetros fixos (BRESLOW e CLAYTON, 1993; MCCULLOCH e SEARLE, 2001). Os GLMMs assumem a forma:

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i + \boldsymbol{\epsilon}_i, \quad \text{para } i = 1, 2, \dots, N \quad (2.9)$$

onde \mathbf{y}_i representa um vetor ($k \times 1$) de respostas do i -ésimo indivíduo com densidade pertencente à família exponencial, e \mathbf{X}_i é uma matriz ($k \times p$) de especificação dos efeitos fixos (conhecida), $\boldsymbol{\beta}$ é o vetor ($p \times 1$) de parâmetros (efeitos fixos), \mathbf{Z}_i é uma matriz ($k \times q$) de especificação dos efeitos aleatórios (desconhecida), \mathbf{b}_i é o vetor ($q \times 1$) de parâmetros (efeitos aleatórios), tal que $\mathbf{b}_i \sim N(\mathbf{0}, \mathbf{G}_i)$ e $\boldsymbol{\epsilon}_i$ é o vetor ($k \times 1$) de erros aleatórios normalmente distribuídos e independentes dos efeitos aleatórios, tal que $\boldsymbol{\epsilon}_i \sim N(\mathbf{0}, \mathbf{R}_i)$. Supõe-se que os erros residuais, bem como os efeitos aleatórios são homocedásticos e são não correlacionados, de modo que, suas respectivas matrizes de variância-covariância são matrizes diagonais dadas por: $\mathbf{R} = \sigma_\epsilon^2 \mathbf{I}$ e $\mathbf{G} = \sigma_b^2 \mathbf{I}$.

Vários trabalhos sobre padronização de CPUE de espécies de atuns e afins através do GLMM tratam alguns dos parâmetros como efeitos aleatórios (por exemplo, CHANG, 2003; RODRIGUES-MARÍN et al., 2003; BRANDÃO et al., 2004; ORTIZ e AROCHA, 2004; HAZIN et al., 2014). Em geral, os efeitos aleatórios foram introduzidos em modelos para lidar com interações entre ano e outras variáveis categóricas. Também foram aplicados GLMM considerando-se embarcação como um efeito aleatório e área e tempo como efeitos fixos (HELSER et al., 2004).

2.2 Espécie Analisada

O espadarte, também conhecido como meka, cujo nome científico é dado por *Xiphias gladius*, é o único representante da família Xiphiidae e do gênero Xiphias. Este peixe tem como principal característica, e que lhe dá nome, o prolongamento do maxilar superior, como se fosse uma longa espada, podendo atingir mais de 1/3 do comprimento total do corpo (NAKAMURA, 1985). O padrão de siglas da ICCAT é definido a partir do nome vulgar da espécie em inglês, para o espadarte a sigla é SWO (do inglês, swordfish). Para simplificação do texto, SWO será também utilizada de agora em diante para fazer referência ao espadarte.

Quanto à sua descrição, o formato longo e fusiforme do seu corpo lhe permite ser um dos mais velozes nadadores oceânicos (NAKAMURA, 1985). Segundo ABID e IDRISSE (2010), as barbatanas dorsal e anal são ambas feitas de duas partes muito distantes em adultos, onde a primeira é bem maior que a segunda, mas estas são contínuas em espécies jovens; as barbatanas peitorais são um pouco rígida, e cada uma está situada na parte inferior dos dois flancos (Figura 2); a linha lateral está ausente na fase adulta, mas é visível em exemplares de até 1 metro de comprimento. A coloração no dorso e flancos pode variar de azul-escuro a roxo (quando vivos) e preto a castanho (30 minutos após sua morte), mudando gradualmente de tonalidade para castanho-claro na parte ventral. A nadadeira dorsal é preta ou castanho-escuro. Os juvenis são geralmente azulados na parte superior. Apenas os indivíduos jovens apresentam escamas, bastante diferentes, que desaparecem gradualmente com a idade (PALKO et al., 1981; ARFELLI, 1996).

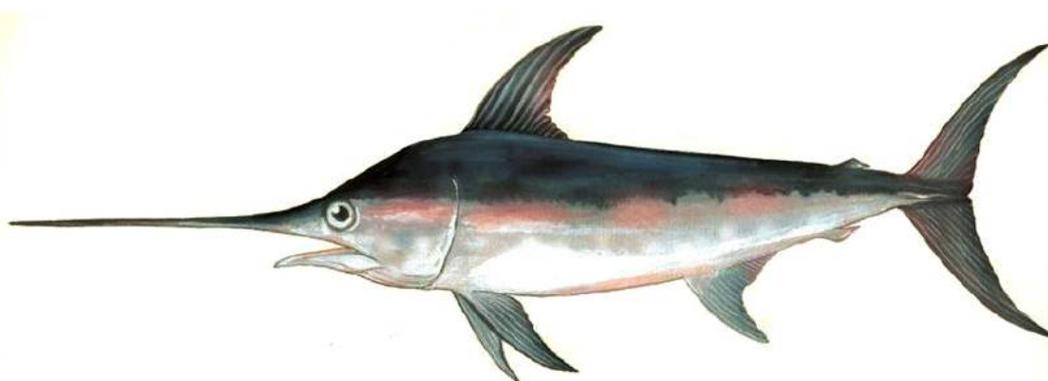


Figura 2: Ilustração de um espadarte adulto (*Xiphias gladius*). Fonte: Abid and Idrissi (2010)

O maior espadarte registrado foi de 455 cm de comprimento da extremidade da mandíbula Inferior até a furca caudal (*Lower Jaw to Fork Length* - LJFL), e pesava um total de cerca de 537 quilogramas (ABID e IDRISSE, 2010). As fêmeas crescem mais rápido que os machos e também atingem um tamanho máximo maior.

Quanto ao habitat e migração, o SWO é considerado uma espécie pelágico-oceânica, vivendo desde a superfície até as águas profundas, geralmente acima das termoclinas, mas podem ser encontrados em águas costeiras (PALKO et al., 1981). Realiza migração vertical diária, encontrando-se em águas mais profundas durante o dia e na superfície durante a noite (CAREY e ROBISON, 1981). Apresenta um comportamento migratório horizontal, entre 45°N e 45°S, relacionado com as flutuações sazonais na temperatura da água do mar, ou seja, deslocando-se para águas temperadas no verão e retornando para águas mais quentes no outono (COLLETTE, 1995). Resultados dos programas de etiquetagem realizados no Atlântico Norte e do Sul indicam que o espadarte movimenta-se significativamente entre as águas subtropicais relativamente quentes e as águas temperadas, tanto do Norte, quanto do Sul do Atlântico (ICCAT, 2007). Sua desova ocorre em latitudes tropicais e subtropicais, e geralmente em profundidades maiores que 200 metros (PALKO et al., 1981).

É a espécie de peixe de bico com a maior tolerância de temperatura (5 a 27°C), mas é frequentemente encontrado em águas de superfície a temperaturas superiores a 13°C (NAKAMURA, 1985). Quando jovens, são mais termicamente dependentes e gostam de temperaturas mais elevadas (MEJUTO e HOEY, 1991). Possuem um sistema de aquecimento altamente especializado que aquece especificamente os olhos e o cérebro até 10 – 15°C acima da temperatura da água do ambiente (CAREY, 1981 e FRISCHES et al., 2005). É um predador oportunista, alimentando-se principalmente de peixes, lulas e crustáceos (MELLO, 1992).

Quanto ao tempo de vida, crescimento e maturação, a máxima idade da espécie foi estimada em 10 anos no Mediterrâneo, porém alguns estudos tem mostrado que alguns espadartes podem viver até 15 anos (ABID e IDRISSE, 2010). São considerados juvenis os SWO com comprimento inferior a 120 cm de comprimento (LJFL) (ICCAT, 1996). Em geral, atinge 140 cm de LJFL em três anos e após isto a taxa de crescimento diminui. Para que esta espécie de grande porte e com alta taxa metabólica cresça rapidamente, é necessária a ingestão de grandes quantias de alimento. O espadarte com 8 anos já atinge de 80 - 130 kg de peso (BERKELEY e HOULE, 1983). Quanto à reprodutividade, segundo HAZIN et al. (2002), a primeira maturação sexual do SWO no Atlântico Sul só é atingida em torno de 156 cm de comprimento (LJFL). A ICCAT

limita o comprimento e peso das espécies de interesse econômico, evitando assim a captura de indivíduos imaturos, para o SWO, o comprimento mínimo de desembarque é de 125 cm (LJFL), no entanto, este comprimento mínimo, não são aplicados para o espadarte capturado acidentalmente, até ao limite de 15%, expresso em número de espadartes, das quantidades desembarcadas (ICCAT, 1996 e 2014a). Apesar das restrições quanto ao tamanho mínimo, 28% dos SWO capturados entre os anos de 2000 a 2002 no norte do Brasil, tinham LJFL menores que 125 cm (ASANO-FILHO et al., 2004). Segundo DALLAGNOLO e ANDRADE (2006), a média anual de comprimento dos espadartes capturados na Costa Oeste do Atlântico, entre 1992 e 2003, apresentou-se estável entre 1993 e 2002, com valor mínimo em 2003 (140 cm) e máximo em 1998 (149 cm), ou seja, indivíduos na fase de maturação sexual. A captura de espécimes que ainda não contribuíram com a reposição da biomassa da população, por não terem gerado descendentes, pode causar uma gradativa diminuição no estoque da espécie, tornando a atividade pesqueira insustentável para este recurso.

Quanto aos estoques, a hipótese mais aceita atualmente é a de que há três distintos para o espadarte (figura 3), um no Atlântico Norte (ATN), outro no Atlântico Sul (ATS), separados na latitude 5°N, e um terceiro no Mediterrâneo (MED), embora haja discussões sobre esta distinção, por exemplo, quanto a demarcação da linha divisória na latitude 5°N entre os estoques Sul e Norte, estudos recentes apoiam esta distinção, visto que foram verificada a existência de diferenças significativas entre as espécies dos estoques, quanto ao tamanho inicial da maturidade sexual, parâmetros de crescimento, área de desova, estrutura genética, dentre outras (ICCAT, 2013). Entretanto, segundo o relatório do Comitê Permanente de Pesquisa e Estatística da ICCAT (ICCAT, 2014) as avaliações de estoques realizadas ainda apresentam um indesejável elevado grau de incerteza em função da carência de dados, particularmente em relação ao Atlântico Sul. Segundo HAZIN *et al.* (2002), a diversidade de parâmetros reprodutivos do espadarte, em diferentes áreas do oceano Atlântico parecem adicionar mais complexidade no entendimento da estrutura populacional desta espécie.

Quanto às capturas e do esforço de pesca, o desenvolvimento da pesca de atuns e afins com espinhel de superfície no Brasil, desde a sua introdução em 1956 (quando uma embarcação japonesa começou a operar baseada no porto de Recife-PE), sofreu grandes oscilações no que diz respeito ao número de embarcações componentes da frota, aos avanços tecnológicos do petrecho de pesca e ao direcionamento das capturas. Até os anos 90 os principais componentes das capturas eram atuns e cações e o espadarte era capturado ocasionalmente como fauna acompanhante (ARFELLI e AMORIN, 1988).

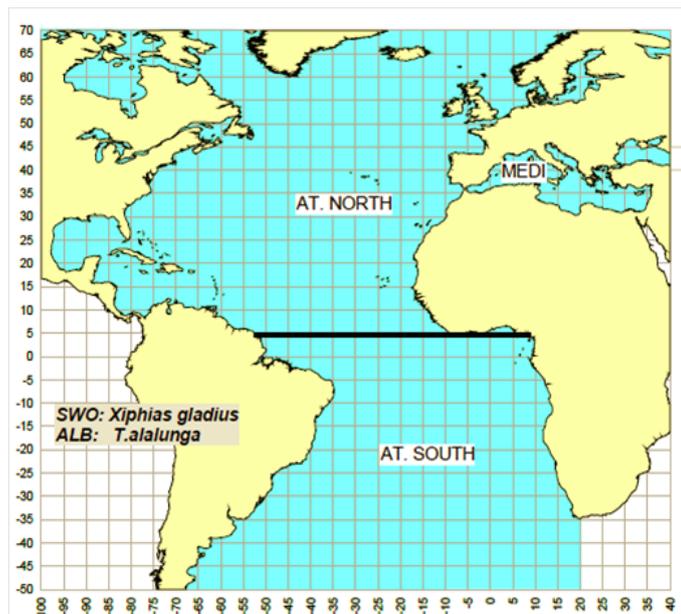


Figura 3: Delimitações geográficas dos estoques de Espadarte no Oceano Atlântico.
 Fonte: ICCAT (2014)

A partir de 1991, devido à crescente demanda do mercado internacional pelo espadarte, as embarcações brasileiras e algumas arrendadas (embarcações estrangeiras contratadas por empresas de pesca brasileiras) começaram a direcionar as capturas para esta espécie, com o uso o espinhel de nylon monofilamento e de atrativos luminosos (ARFELLI, 1996) Além disso, o espinhel passou a ser lançado no início da noite, iniciou-se a utilização de lulas como isca, e o lançamento começou a ser feito para propiciar que o anzol ficasse entre 30 e 200 m de profundidade. Tais mudanças ocorreram devido ao fato de que o espadarte costuma alimentar-se no período noturno e em águas mais rasas (ARFELLI e AMORIN, 2000).

3 Modelagem da CPUE via GLMs para dados de contagens

A maioria dos métodos usados para modelar dados de capturas e esforço de pesca, em particular de espadarte (*Xiphias gladius*), são baseados na aplicação de GLMs com distribuições contínuas, utilizando como variável resposta a captura em peso ou a taxa de captura. Visando contribuir para a diminuição da elevada incerteza no cálculo dos índices relativos de abundância do estoque do Atlântico Sul relatada pela ICCAT (ICCAT, 2014), utilizamos alguns GLMs para dados de contagens, tomando como variável resposta, ao invés da taxa de captura, o número de peixes capturados e adicionando ao preditor linear o esforço de pesca (número de anzóis) como um *offset*.

De início, avaliamos o modelo com distribuição de Poisson com função de ligação log, que surge como primeira abordagem possível quando lidamos com dados de contagens, entretanto, devido a possível presença da sobredispersão e excesso de zeros, propomos também o modelo com distribuição binomial negativa (HINDE e DEMÉTRIO, 1998) e os modelos inflacionados de zeros que exigem em sua estrutura conhecimentos sobre mistura de modelos e são obtidos através da média ponderada de duas distribuições, uma degenerada no ponto zero e outra que se adequaria aos dados, caso não houvesse zeros excessivos (RIDOUT et al., 1998).

3.1 Modelo de Poisson

Seja Y_1, \dots, Y_n , variáveis aleatórias independentes com Y_i denotando o "número de espadartes capturados na i -ésima pescaria". Então, segundo a distribuição de Poisson, a probabilidade de Y_i assumir um valor qualquer y_i , de peixes capturados, é dada por

$$f(y_i; \lambda_i) = \frac{e^{-\lambda_i} \lambda_i^{y_i}}{y_i!}, \quad y_i = 0, 1, 2, \dots, \quad (3.1)$$

onde o parâmetro λ_i ($\lambda_i > 0$) representa o número médio de capturas. Uma propriedade da Poisson é que $E(y_i) = \lambda_i = Var(y_i)$, e por isso, o parâmetro de dispersão é fixo em $\phi^{-1} = 1$. Consideremos agora Y_i denotando o número de peixes capturados na i -ésima pescaria, para um dado número de anzóis n_i . Nesse caso, o valor esperado de Y_i pode ser escrito como

$$E(Y_i) = \lambda_i = n_i \theta_i. \quad (3.2)$$

Se θ_i depende de variáveis explicativas, então pode ser modelado por $\theta_i = e^{\mathbf{x}_i^T \boldsymbol{\beta}}$. Portanto, o GLM proposto, será dado por $E(Y_i) = \lambda_i = n_i e^{\mathbf{x}_i^T \boldsymbol{\beta}}$. Usando como função de ligação a função logarítmica, teremos como modelo resultante

$$\log \lambda_i = \log n_i + \mathbf{x}_i^T \boldsymbol{\beta}, \quad (3.3)$$

onde os termos x_i e $\boldsymbol{\beta}$, descreve a covariável padrão e os parâmetros. O modelo (3.3) difere da especificação tradicional do preditor linear, devido à inclusão do termo $\log n_i$ (Dobson, 2002). Este termo é uma constante compensatória (offset) nos casos onde o número de capturas é utilizada como variável resposta em vez da taxa de capturas, neste caso, a medida do esforço deve ser incluída no modelo como uma compensação. Embora este modelo seja o mais natural para modelar dados de contagens, ele não "lida" com excesso de zeros nem com sobredispersão dos dados. A partir do modelo (3.3), teremos como CPUE estimada

$$\begin{aligned} \log \left(\frac{\lambda_i}{n_i} \right) &= \mathbf{x}_i^T \boldsymbol{\beta}, \\ \log \left(\widehat{CPUE} \right) &= \mathbf{x}_i^T \hat{\boldsymbol{\beta}}, \\ \widehat{CPUE} &= e^{\mathbf{x}_i^T \hat{\boldsymbol{\beta}}} \end{aligned} \quad (3.4)$$

3.1.1 Sobredispersão

A distribuição de Poisson tem apenas um parâmetro e por isso não é muito flexível para fins de ajustes empíricos. Podemos generalizar, permitindo-lhes a inserção de um parâmetro de dispersão. Sobre ou subdispersão pode ocorrer em modelos de Poisson. Sendo assim, podemos introduzir um parâmetro de dispersão ϕ , tal que $Var(Y) = \lambda \phi$. Para $\phi = 1$ temos um modelo de Poisson usual, no entanto, se $\phi > 1$ teremos o caso da sobredispersão e $\phi < 1$ o caso da subdispersão. Sob a hipótese de que o modelo de Poisson fornece um bom ajuste aos dados, podemos estimar o parâmetro de dispersão ϕ por meio da estimativa $\hat{\phi} = \frac{X^2}{(n-p)}$, onde $X^2 = \frac{\sum (y_i - \hat{\mu}_i)^2}{\hat{\mu}_i}$ é a estatística de Pearson

(SMYTH, 2003 e FARAWAY, 2006). Neste caso, se $\hat{\phi}$ for maior que a unidade, podemos estar diante do problema da sobredispersão ou variabilidade extra-Poisson, cuja ocorrência é caracterizada quando a variabilidade observada excede aquela assumida pelo modelo de Poisson, isto é, $Var(Y_i) > \lambda_i$ (BRESLOW, 1984, LAWLESS, 1987 e PAULA, 2013).

Segundo HINDE e DEMÉTRIO (1998), isso ocorre devido a vários motivos, tais como: variabilidade das unidades observacionais, maior incidência de contagens iguais a zero e correlação entre respostas individuais, por exemplo, em dados de pesca, as capturas feitas pelo mesmo barco em momentos diferentes provavelmente são correlacionadas. Entretanto, em geral, é difícil saber qual é a causa exata da sobredispersão, mas é preciso estar atento para sua ocorrência, pois não considerá-la na análise dos dados pode levar a sérios problemas de subestimação dos erros-padrão e, conseqüentemente, uma avaliação incorreta da significância dos parâmetros. Além disso, pode provocar alterações na devião, dificultando a seleção do melhor modelo. Portanto, reconhecer dados com sobredispersão é primordial para que sejam tomadas precauções e medidas capazes de garantir uma estimação com menor margem de erro, ou seja, com maior segurança.

BANNEROT e AUSTIN (1983) e PUNT et al. (2000), afirmam que em estudos sobre a padronização de dados de captura e esforço, quando se utiliza como variável resposta o número de peixes capturados, normalmente se tem dados sobredispersos em relação ao modelo de Poisson, e recomendam que seja mais adequado modelar os dados usando a distribuição binomial negativa, que permite uma relação quadrática entre a média e a variância.

3.1.2 Teste para Sobredispersão no Modelo de Poisson

LAWLESS (1987), desenvolveu um teste para verificar a presença de sobredispersão baseado na estatística de dispersão padronizada

$$S = \frac{\sum_{i=1}^n \{(y_i - \hat{\mu}_i)^2 - \bar{Y}\}}{\left(2 \sum_{i=1}^n \hat{\mu}_i^2\right)^{1/2}}. \quad (3.5)$$

Sob a hipótese de que $Y_i \sim Poisson(\lambda_i)$, para $n > 50$, S é assintoticamente normal padrão onde, valores positivos grandes de S , ($p < 0.05$), indica sobredispersão relativa

ao modelo de Poisson, sugerindo assim o Modelo Binomial Negativo, visto a seguir. Testes assintoticamente semelhantes foram desenvolvidos por DEAN (1992) e GREENE (2002). Convém destacar que a grande maioria dos testes disponíveis na literatura confrontam somente modelos aninhados.

3.2 Modelo Binomial Negativo

Verificada a presença de sobredispersão em dados de contagem, podemos proceder à modelagem dos dados assumindo uma distribuição Binomial Negativa (*Negative Binomial-NB*) para os mesmos, a qual surge de uma mistura das distribuições Poisson e Gama. Para tanto, substitui-se o valor médio da Poisson por uma variável aleatória Z , tal que, $Y|z \sim P(z)$ e $Z \sim Gama(\mu, \phi)$, onde ϕ independe de μ . Neste caso, $E(Z) = \mu$ e $Var(Z) = \frac{\mu^2}{\phi}$. A combinação destas duas distribuições é a Binomial Negativa com parâmetros média μ e parâmetro de dispersão ϕ ($\phi \geq 0$) (veja Dobson, 2002). Denotamos $Y \sim BN(\mu, \phi)$.

Suponhamos então que Y_1, \dots, Y_n , variáveis aleatórias independentes com Y_i denotando o "número de peixes capturados na i -ésima pescaria", tais que Y_i segue uma Binomial Negativa com parâmetros μ e ϕ , denotado por $BN(\mu, \phi)$. Desta forma, a função de probabilidades de Y_i , segundo a parametrização de NELDER e WEDDERBURN (1972), em que $p = \frac{\mu}{\mu + \phi}$, $0 < p < 1$ e ϕ^{-1} é o parâmetro de dispersão ($\phi > 0$), fica dada por

$$P(Y_i = y_i) = \frac{\Gamma(y_i + \phi)}{\Gamma(\phi) \cdot y_i!} \left(\frac{\mu_i}{\mu_i + \phi} \right)^{y_i} \left(1 - \frac{\mu_i}{\mu_i + \phi} \right)^{\phi}, \quad y_i = 0, 1, 2, \dots, \quad (3.6)$$

em que $\Gamma(\cdot)$ é a função gama, isto é, $\Gamma(x) = \int_0^{\infty} t^{x-1} e^{-t} dt$, $x > 0$.

Segue que a média desta distribuição é dada por $E(Y_i) = \mu_i$ e a variância é dada por $Var(Y_i) = \mu_i + \frac{\mu_i^2}{\phi}$ (DOBSON, 2002). Observe que, comparativamente a distribuição de Poisson, a variância da binomial negativa é maior que a média, sendo bastante útil no ajuste de dados de contagem com sobredispersão. Além disso, quando ϕ^{-1} tende a zero, a distribuição Binomial Negativa se aproxima da Poisson (CAMERON e TRIVEDI, 1998). Informações mais detalhadas sobre o modelo Binomial Negativo, podem ser encontradas em LAWLESS (1987).

3.3 Modelos Inflacionados de Zeros

3.3.1 Excesso de Zeros

Em muitas ocasiões práticas, onde se trabalha com dados de contagem, não raramente depara-se com a ocorrência de excesso de zeros, o que supostamente, entende-se como uma das causas do fenômeno de sobredispersão, que é o caso dos dados de contagem relativos a estudo de abundancia de espécies raras (POTTS e ELITH, 2006; WELSH *et al.*, 1996). Nesses casos, a suposição de distribuição de Poisson e binomial negativo para a resposta pode ser inadequada, por não se saber com precisão a quantidade de observações nulas que esses modelos suportam, sendo comumente necessário o uso de modelos alternativos que acomodem o excesso de zeros (MCCULLAGH e NELDER, 1989).

Em se tratando de dados de captura de peixes da espécie espadarte, objeto de estudo deste trabalho, assumimos que o número de unidades capturadas de espadarte é o resultado de um processo de duas partes independentes: a primeira decorrente de observações que assumem o valor zero como única opção (distribuição degenerada no ponto zero), indicando uma baixa frequência de ocorrência da espécie ou sua total ausência (zeros estruturais ou verdadeiros) e a segunda proveniente de um processo que gera contagens aleatórias provenientes da distribuição discreta de probabilidade, indicando que a ocorrência de zero se trata de um valor nulo acidental (zero amostral), ou seja, que a espécie existe no local, mas as unidades da mesma não foram capturadas pelos anzóis. Diferenças entre essas duas classificações de zeros, são tratadas em RIDOUT *et al.* (1998).

Diante deste contexto, propomos o ajuste dos dados de captura do espadarte, por meio das versões inflacionadas de zeros dos modelos usuais de Poisson e Binomial Negativo, denominadas por, Modelo Poisson Inflacionado de Zeros (LAMBERT, 1992; HALL, 2000) e Modelo Binomial Inflacionado de Zeros (YAU *et al.* (2003)), respectivamente.

3.3.2 Modelo de Poisson Inflacionado de Zeros

A partir do Modelo Poisson Inflacionado de Zeros (*Zero Inflated Poisson-ZIP*), segundo nomenclatura utilizada por RIDOUT *et al.* (1998), tem-se a hipótese de que, com probabilidade p a variável resposta Y , números de unidades capturadas, assume o valor zero e com probabilidade $(1 - p)$ assume que as contagens aleatórias,

seguem um modelo de Poisson, com parâmetro $\lambda > 0$. Então, segundo o modelo ZIP, a probabilidade de Y_i assumir um valor qualquer y_i , de peixes capturados, é dada por

$$P(Y_i = y_i) = \begin{cases} p_i + (1 - p_i)e^{-\lambda} & \text{se } y_i = 0 \\ (1 - p_i)\frac{e^{-\lambda_i}\lambda_i^{y_i}}{y_i!} & \text{se } y_i = 1, 2, \dots, \end{cases} \quad (3.7)$$

em que o parâmetro p_i (com a restrição, $0 < p_i < 1$), representa a proporção de ocorrências de zeros estruturais, e o valor esperado e a variância de Y_i são, respectivamente:

$$E(Y_i) = \lambda_i(1 - p_i) \quad \text{e} \quad Var(Y_i) = \lambda_i(1 - p_i)(1 + p_i\lambda_i). \quad (3.8)$$

Neste caso, tem-se também, $\phi = 1$. Note que a variância da mistura é maior que a média da distribuição; e, quanto maior a probabilidade de zeros, p_i , maior a variância da variável, por outro lado, à medida que p_i se aproxima de zero, a variância se aproxima de μ_i , retornamos a distribuição Poisson usual.

Além disso, no ajuste de um modelo ZIP com covariáveis e parâmetros $\mathbf{p} = (p_1, \dots, p_n)^T$ e $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_n)^T$, via teoria dos GLM, faz-se necessário a definição das funções de ligação logística (para a parte inflacionada de zeros) e logarítmica (para a parte não inflacionada de zeros), dadas, respectivamente, por:

$$\log\left(\frac{p_i}{1 - p_i}\right) = \mathbf{G}_i^T \boldsymbol{\gamma} \quad \text{e} \quad \log\{E(Y_i)\} = \mathbf{Z}_i^T \boldsymbol{\beta} \quad (3.9)$$

em que \mathbf{G} e \mathbf{Z} são as matrizes associadas às covariáveis; e, $\boldsymbol{\gamma}$ e $\boldsymbol{\beta}$ os vetores de parâmetros desconhecidos das covariáveis associadas a cada parte do modelo;

De acordo com RIDOUT *et. al.* (1998), as matrizes \mathbf{G} e \mathbf{Z} podem ou não serem coincidentes. Caso elas sejam, pode-se pensar em \mathbf{p} como uma função de $\boldsymbol{\lambda}$, acarretando em uma possível redução no número de parâmetros a serem estimados e em uma simplificação do modelo. Assim, caso tenhamos as funções de ligações definidas anteriormente, teríamos o chamado modelo ZIP(τ), onde $\mathbf{G}\boldsymbol{\gamma} = -\tau\mathbf{Z}\boldsymbol{\beta}$, onde τ é um particular parâmetro escalar, o que implica em $p_i = (1 + \lambda_i^\tau)^{-1}$ (LAMBERT, 1992). A obtenção dos estimadores de \mathbf{p} e $\boldsymbol{\lambda}$, para o modelo ZIP, pode ser feita por meio do método da máxima verossimilhança (MMV), cujas soluções numéricas são obtidas por processos iterativos do tipo Newton-Raphson (XIE *et. al.*, 2001).

3.3.3 Modelo Binomial Negativo Inflacionado de Zeros

O modelo Binomial Negativo Inflacionado de Zeros (*Zero Inflated Negative Binomial-ZINB*) funciona de uma forma similar ao ZIP, bastando substituir a distribuição

de Poisson pela binomial negativa. Segundo LEWSEY e THOMSON (2004), a distribuição ZINB surge como uma boa alternativa para dados de contagens com excessos de zeros, visto que a sobredispersão neste cenário pode acarretar sérios problemas ao modelo ZIP, como a subestimação dos erros padrão dos estimadores e o consequente aumento do p-valor associado aos parâmetros do modelo, produzindo inferências pouco confiáveis. Segundo parametrização de YAU *et al.* (2003), o modelo ZINB pode ser escrito da forma:

$$P(Y_i = y_i) = \begin{cases} p_i + (1 - p_i) \left(\frac{\phi}{\mu_i + \phi} \right)^\phi & \text{se } y_i = 0 \\ (1 - p_i) \frac{\Gamma(y_i + \phi)}{\Gamma(\phi) \cdot y_i!} \left(\frac{\phi}{\mu_i + \phi} \right)^\phi \left(1 - \frac{\phi}{\mu_i + \phi} \right)^{y_i} & \text{se } y_i = 1, 2, \dots, \end{cases} \quad (3.10)$$

em que $0 < p_i < 1$, $\mu_i > 0$ e $\phi > 0$, onde $(1 - p_i)$ representa a probabilidade de zeros amostrais. O valor esperado e a variância de Y_i , são, respectivamente: $E(Y_i) = (1 - p_i)\mu_i$ e $Var(Y_i) = (1 - p_i) \left(1 + \frac{\mu_i}{\phi} + p_i\mu_i \right)$. A distribuição ZINB aproxima-se da ZIP quando $\phi \rightarrow \infty$ e da binomial negativa quando $p_i \rightarrow \infty$. Se ambos $\frac{1}{\phi}$ e $p_i \approx 0$ então a distribuição ZINB é reduzida à Poisson. Semelhante ao modelo ZIP, os parâmetros $\mathbf{p} = (p_1, \dots, p_N)^T$ e $\boldsymbol{\lambda} = (\mu_1, \dots, \mu_n)^T$ são modelados pelas funções de ligação $\ln \left(\frac{p_i}{1 - p_i} \right) = \mathbf{G}_i \boldsymbol{\gamma}$ e $\ln(\mu_i) = \mathbf{Z}_i \boldsymbol{\beta}$, respectivamente (MONTROYA, 2009).

O modelo ZIP é preferido na modelagem de dados de contagens onde as variações adicionais são causadas exclusivamente pelos zeros extras. Para dados de contagem sobredispersos onde a variabilidade extra é causada por excesso de zeros e pela heterogeneidade também não observada, recomenda-se o uso dos modelos ZINB (PHANG e LOH, 2013).

3.3.4 Teste Comparativo entre os Modelos Inflacionados de Zeros e suas Versões Usuais

Tendo em vista que, em se tratando de dados de contagens, dependendo da quantidade de dados amostrais e do número de observações nulas, os modelos de Poisson e Binomial Negativo podem ter ajuste, tão bom quanto, ou melhor, do que suas versões infladas de zeros. Para resolver este empasse, VUONG (1989) propôs uma estatística para testar se os modelos inflados ZIP ou ZINB se ajustam melhor a um conjunto de dados com excesso de zeros, que suas versões usuais (encaixados ou não encaixadas), sem comprometer a inferência estatística (PHANG e LOH, 2013). Seja $P_u(y_i|x_i)$ a probabilidade prevista de uma contagem observada para o caso i do modelo u ($u = 1, 2$),

define-se $m_i = \log \left(\frac{P_1(y_i|x_i)}{P_2(y_i|x_i)} \right)$. Assim, a estatística do teste de Vuong

$$V = \frac{\sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n m_i \right)}{\sqrt{\frac{1}{n} \sum_{i=1}^n (m_i - \bar{m})^2}}, \quad (3.11)$$

sob a hipótese nula de que os modelos são indistinguíveis, $E(m_i) = 0$, é assintoticamente distribuída $N(0, 1)$. Ao nível de significância de 5%, o primeiro modelo é preferido se $V > 1.96$; Se $V < -1.96$, então o segundo é preferível. Entretanto, quando $|V| < 1.96$ os modelos são equivalentes (MOUATASSIM e EZZAHID, 2012). No software R, o teste de Vuong está implementado no pacote `pscl` e para comparar dois modelos, por exemplo, `mod1` e `mod2`, utiliza-se o código `vuong(mod1, mod2)`. Convém destacar que, assim como os testes de sobredispersão, a grande maioria dos testes disponíveis na literatura confrontam somente modelos encaixados, onde um é um caso especial do outro.

3.4 Materiais e Métodos

3.4.1 Descrição dos dados

Foram analisadas as informações do Banco Nacional de Dados de Atuns e Afins (BNDA), mantido pelo Subcomitê Científico de Atuns e Afins (SCC de Atuns e Afins) vinculado ao Ministério de Pesca e Aquicultura (MPA). O BNDA trás informações dos mapas e de observadores de bordo sobre a produção e esforço pesqueiro em cada evento de pesca das embarcações pertencentes a frota atuneira brasileira (nacional ou arrendadas) que atuam com espinhel de superfície no oceano Atlântico. Após alguns dias de viagem, realizando eventos de pesca diários (geralmente consecutivos), as embarcações retornam ao porto de origem para descarregar a produção, retomando à atividade dias depois.

Após sucessivas análises exploratórias dos dados, constatou-se que apenas alguns anos tinham dados suficientes para uma análise temporal das capturas e, que a maioria das embarcações apresentava informações em intervalos irregulares de tempo, permanecendo inativas durante meses ou anos. Diante deste cenário, baseamos nossa análise nos registros que ocorreram em um intervalo de tempo com um maior número de embarcações trabalhando em anos coincidentes e com menor período de inatividade. Com

estes critérios, selecionamos 5579 eventos de pesca realizados durante o período de 1998 a 2006, referentes a 10 embarcações pertencentes às frotas nacional (BRA) e arrendadas da Espanha (BRA-ESP), Honduras (BRA-HND), Portugal (BRA-PRT), Estados Unidos (BRA-USA) e São Vicente e Granadinas (BRA-VCT). A área de estudo deste trabalho, compreendeu a porção oeste do Atlântico Sul, situada entre as latitudes de 5°N e 45°S e entre 0°E de longitude e a costa da América do Sul.

O espinhel de superfície é composto de uma linha principal suspensa na água através de boias com linhas de flutuação distribuídas em intervalos regulares. Na linha principal também são fixadas através de grampos de metal, linhas secundárias com atratores e anzóis iscados nas extremidades. As características dos materiais, dimensões, tipos de boias e de anzóis, bem como a configuração das linhas são bastante variáveis, dependendo principalmente da origem das frotas, da área de pesca e da espécie-alvo. Por exemplo, tanto o número de anzóis entre boias, quanto a distância entre os mesmos, pode variar caso deseje-se que o aparelho de pesca atue em menores ou maiores profundidades (DOMINGO et al., 2014).

No banco de dados do BNDA, além da origem das embarcações, há informações de caráter temporal (hora, dia, mês e ano do lançamento e da retirada do espinhel), espacial (latitude, longitude) e referentes às operações de pesca como, o número total de anzóis e o número de anzóis entre boias utilizados no espinhel, assim como, o desempenho quantitativo (número e peso, para as diferentes espécies capturadas).

3.4.2 Variável Resposta

Para o ajuste dos GLMs propostos, o número de espadartes capturados foi utilizado como variável resposta, adicionando ao preditor linear o logaritmo do número de anzóis como *offset*. Além disso, a captura por unidade de esforço (CPUE), para cada evento de pesca, foi definida pela equação

$$CPUE = \left(\frac{\text{número de espadartes capturados}}{\text{número de anzóis}} \right) \times 1000. \quad (3.12)$$

3.4.3 Variáveis explicativas

Para a análise, foram consideradas as seguintes variáveis explicativas com possíveis influências na capturabilidade do espadarte: número de anzóis entre boias (hpb), tempo de imersão do espinhel (tp.ime), trimestre, ano, turno e a latitude (lat) em que ocorreu

o evento de pesca. Além destas, consideramos também, as distâncias entre o local onde a pescaria foi realizada para diferentes posições na linha da costa. Foram calculadas oito distâncias radiais, gerando o vetor $\mathbf{d} = \{d_1, d_2, d_3, d_4, d_5, d_6, d_7, d_8\}$ (Figura 4), sob os respectivos ângulos, $0^\circ, 45^\circ, 90^\circ, 135^\circ, 180^\circ, 225^\circ, 270^\circ$ e 315° graus.

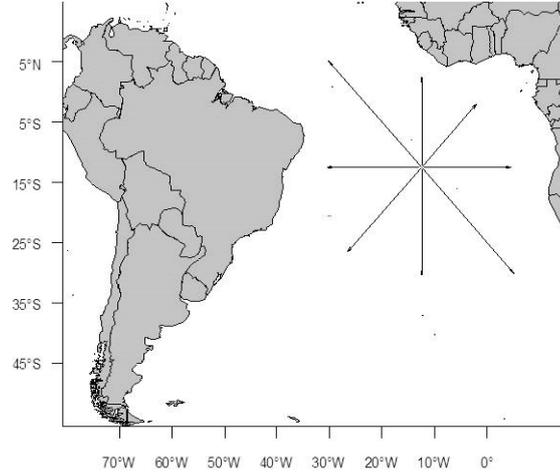


Figura 4: Ilustração das oito direções radiais utilizadas no cálculo das distâncias do local da pescaria à costa

É importante ressaltar que as trajetórias das radiais que não correspondem a ângulos retos sofrem distorções em virtude da curvatura da Terra. As equações que permitem o cálculo das trajetórias com o ajuste para as distorções são:

$$la_j = \arcsen(\sen(la_i)\cos(d/R) + \cos(la_i)\sen(d/R)\cos(\theta)) \quad (3.13)$$

$$lo_j = lo_i + \arctan^2(\sen(\theta)\sen(d/R)\cos(la_i)\cos(d/R)\sen(la_i)\sen(la_j)) \quad (3.14)$$

onde lo_i, lo_j, la_i e la_j são, respectivamente, as longitudes e latitudes (em radianos) da i -ésima e j -ésima localização geográfica. A função $\arctan^2(x, y)$ calcula o ângulo entre o eixo x e o vetor da origem até a posição definida pelas coordenadas (x, y) , \mathbf{d} representa a distância da posição $(lo_i; la_i)$ a $(lo_j; la_j)$, R representa o raio médio da Terra ($\approx 6,371km$) e θ é o ângulo da posição de partida.

Desta forma, novas variáveis explicativas puderam ser calculados com base no vetor de distâncias \mathbf{d} : min (distância mínima à costa), $s_{(k)}$ (média das k menores distâncias da costa), $k = 2, 3$ (nomeadas, respectivamente por: s_2 e s_3) e transformações potenciais das mesmas, min^p e $s_{(k)}^p$, com $p = 1/2, 2, 3$ (nomeadas, respectivamente por: min.a, min.b, min.c, s2.a, s2.b, s2.c, s3.a, s3.b e s3.c). Tais transformações permitem avaliar a existência de relações não lineares entre a variável resposta e essas variáveis explicativas.

3.4.4 Ajuste dos Modelos

3.4.4.1 Inclusão das variáveis explicativas e simplificação dos modelos

Como em estudos observacionais, os termos de GLMs 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 e NELDER, 1989). Neste sentido, para se definir a ordem de entrada nos modelos investigados, cada variável explicativa foi ajustada sozinha com a mesma variável resposta e, em seguida, calculou-se a deviança de cada modelo. Selecionamos inicialmente a variável explicativa do modelo com menor deviança ou, equivalentemente, maior log-verossimilhança, significando maior redução na deviança. Em seguida, os modelos foram ajustados considerando a variável previamente selecionada, além de todas as outras uma de cada vez, levando em conta sempre o modelo com a menor deviança. O processo continuou até a obtenção do modelo base com todas as variáveis explicativas (efeitos principais). Após obtenção do modelo base (com todas as variáveis explicativas, especificadas e previamente ordenadas), interações de primeira ordem foram incluídas uma a uma no modelo e mantidas, desde que promovessem uma redução significativa na deviança ($\alpha = 5\%$) em testes da razão de verossimilhança (TRV) entre os modelos com e sem cada termo. Neste contexto, o TRV foi utilizado para comparar a qualidade do ajuste (*goodness of fit*) de dois modelos encaixados M_0 e M_1 , sendo M_0 um caso particular de M_1 . Assim, dadas as hipóteses

$$H_0 : \boldsymbol{\beta} = \boldsymbol{\beta}_0 = \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_q \end{bmatrix} \quad H_1 : \boldsymbol{\beta} = \boldsymbol{\beta}_1 = \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_p \end{bmatrix}$$

com $q < p < N$. Para testar H_0 contra H_1 , o TRV utiliza como estatística $2[l(\mathbf{b}_1; \mathbf{y}) - l(\mathbf{b}_0; \mathbf{y})] \sim \chi_{(n-p)}^2$, onde l é a log-verossimilhança do modelo e \mathbf{b}_0 e \mathbf{b}_1 são, respectivamente, as estimativas de máxima verossimilhança de $\boldsymbol{\beta}_0$ e $\boldsymbol{\beta}_1$ (DOBSON, 2002).

Os resultados dos testes foram sumarizados na forma de tabelas de análise de deviança que mostram a sequência de modelos encaixados e as diferenças entre devianças residuais promovidas pela adição sucessiva das variáveis explicativas.

Posteriormente, buscando obter um modelo parcimonioso, bem ajustado e com um número reduzido de parâmetros, utilizou-se o método stepwise. O método foi iniciado com o modelo sob avaliação, com efeitos principais e interações significativas, e a seleção dos termos baseou-se no Critério de Informação de Akaike (Akaike Information

Criterion AIC) (AKAIKE, 1973). Para se ter uma medida do poder explicativo dos modelos, foi calculado o coeficiente de determinação Pseudo R^2 como a fração da deviança total explicada.

3.4.4.2 Métodos de Diagnóstico

O processo de escolha do modelo adequado, está relacionada com a qualidade do ajuste dos modelos, ou seja, com a verificação de possíveis afastamentos das suposições feitas para o modelo. Tal etapa, conhecida como análise de diagnóstico, se baseia na análise de medidas que expressam a discrepâncias entre os valores observados y_i e seus valores ajustados $\hat{\mu}_i$, denominadas resíduos (COX e SNELL, 1968). Segundo CORDEIRO e LIMA NETO (2006), os resíduos são usados para explorar a adequação do modelo ajustado com respeito à escolha da função de variância, da função de ligação e de termos no preditor linear. Além disso, eles também são úteis na identificação de observações atípicas que destoam do conjunto de dados. Segundo DOBSON (2002), essas observações são classificadas em três grupos: aberrantes (mal ajustadas com resíduos altos), de alavanca (com influência desproporcional no próprio valor ajustado) e influentes (com influência desproporcional nas estimativas dos coeficientes). Em se tratando de GLMs para dados de contagem, objeto de estudo deste trabalho, as formas mais usuais de resíduos, são os resíduos ordinários dados por $r_i = y_i - \hat{\mu}_i$ e os resíduos de Pearson Studentizado, dados por:

$$t_{S_i} = \frac{\phi^{1/2}(y_i - \hat{\mu}_i)}{\sqrt{\hat{V}_i(1 - \hat{h}_{ii})}}, \quad (3.15)$$

em que h_{ii} é o i -ésimo elemento da diagonal da matriz de projeção \mathbf{H} , definida para os GLMs, por $\mathbf{H} = \mathbf{W}^{1/2}\mathbf{X}(\mathbf{X}^T\mathbf{W}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{W}^{1/2}$, onde \mathbf{X} é a matriz do modelo e \mathbf{W} a matriz de pesos (MCCULLAGH e NELDER, 1989). Esta matriz tem um papel fundamental para os diagnósticos em GLMs. Além de padronização dos resíduos, os elementos h_{ii} também servem como uma medida da influência da i -ésima observação sobre o próprio valor ajustado. Um outro resíduo bastante utilizado dentro da teoria dos GLMs, é o resíduo componente do desvio padronizado

$$t_{D_i} = \frac{\phi^{1/2}d(y_i; \hat{\mu}_i)}{\sqrt{(1 - h_{ii})}}, \quad (3.16)$$

em que $d(y_i; \hat{\mu}_i)$ é a raiz quadrada da função desvio do modelo ajustado, cujo sinal é o mesmo de $y_i - \hat{\mu}_i$. Este resíduo representa uma distância entre y_i e seu valor ajustado $\hat{\mu}_i$, medida na escala logarítmica da função de verossimilhança. WILLIAMS (1987)

verificou através de simulações de Monte Carlo que a distribuição de t_{D_i} tende a estar mais próxima da normalidade do que as distribuições dos demais resíduos. Para se realizar a análise de diagnóstico e a verificação da existência de observações atípicas, recorreremos aos seguintes gráficos de diagnósticos, recomendados por PAULA (2013):

- gráfico de t_{D_i} ou t_{S_i} contra a ordem das observações, para verificar a existência de pontos aberrantes;
- gráfico de t_{D_i} ou t_{S_i} contra \hat{y}_i (valores ajustados), para verificar a falta de homogeneidade da variância;
- gráfico de h_{ii} contra a ordem das observações ou valores ajustados, para verificar possíveis pontos de alavanca;
- gráfico normal de probabilidades para t_{D_i} com banda de confiança (envelope), para verificarmos afastamentos da distribuição suposta; e
- gráfico de $LD_i = \frac{\hat{h}_{ii}}{(1-\hat{h}_{ii})^2} t_{S_i}$ (distância de Cook aproximada, COOK e WEISBERG (1982)) contra a ordem das observações, para detectarmos possíveis pontos influentes.

3.4.4.3 Padronização da CPUE

A padronização da CPUE de espadarte será realizada a partir do modelo GLM com melhor ajuste, assumindo que os valores preditos calculados exclusivamente com os coeficientes da variável "ano", após remoção das variações na CPUE decorrentes das variáveis explicativas incluídas no modelo, não sofram influência de variações causadas por outros fatores relacionados à eficiência de pesca que não foram considerados. Tais valores podem refletir mudanças anuais da biomassa, e foram utilizados como índices relativos de abundância anual (Maunder e Punt, 2004).

3.4.4.4 Procedimento Computacional

Todas as análises foram realizadas com o programa R 3.0.3 (R Development CORE TEAM, 2014), com auxílio da função `glm()` do pacote `stats`, da função `glm.nb()` do pacote `MASS` (VENABLES e RIPLEY, 2002) para o ajuste dos modelos com a distribuição binomial negativa, e da função `zeroinfl()` do pacote `pscl` (JACKMAN, 2008), para o ajuste dos modelos ZIP e ZINB. Os mapas foram construídos com adaptações de funções do pacote `Lattice` (SARKAR, 2008). Uma descrição mais detalhada

sobre modelos de contagem no R pode ser encontrada em ZEILEIS et al. (2008). Quanto a geração dos envelopes utilizados na análise de diagnóstico dos modelos, utilizamos as sub rotinas criadas pelo professor Gilberto Alvarenga Paula (IME/USP) disponibilizadas em sua página <http://www.ime.usp.br/giapaula/cursospos.htm>.

3.5 Resultados

3.5.1 Análise Exploratória

Para um melhor entendimento da relação entre a variável resposta e as variáveis explicativas, realizamos uma análise exploratória dos dados, onde foram construídos sumários e gráficos estatísticos das variáveis resposta e explicativas. Nos eventos de pesca analisados, ocorreram variações entre 0 e 125 unidades capturadas de espadarte. Sendo que 1100 dos 5579 eventos tiveram capturas iguais a zeros, correspondendo a aproximadamente 20% dos dados, como mostra o histograma (Figura 5). Tal histograma mostra uma distribuição assimétrica a direita, sendo a média de espadartes capturados igual a 9,8 unidades capturadas, com variância 175,2. Havendo, portanto, indício de presença de sobredispersão nos dados.

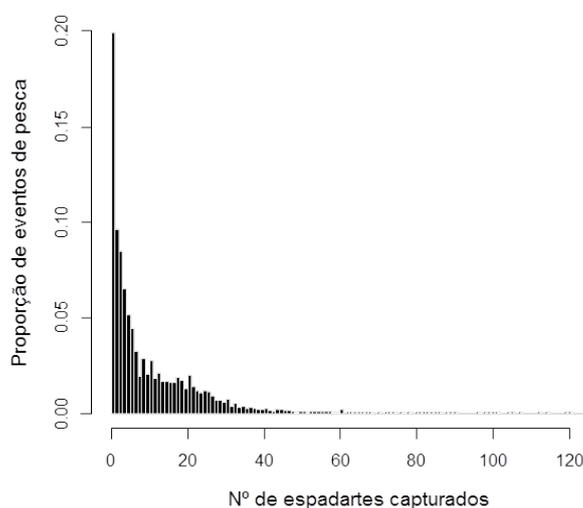


Figura 5: Histograma do número de espadartes capturados por eventos de pesca.

Os dados de captura e esforço de pesca foram agregados em blocos de $5^\circ \times 5^\circ$ de latitude e longitude e a CPUE nominal de espadarte no b -ésimo bloco foi assim calculada: $CPUE_b = (\sum C_b) / (\sum f_b) \times 1000$, em que C é o nº de espadartes capturados e f é o esforço de pesca nominal (número de anzóis). As capturas de espadarte ocorreram na porção oeste do Atlântico Sul, porém com CPUEs maiores nas áreas de alto mar

entre $20^{\circ} - 30^{\circ}\text{S}/10 - 15^{\circ}\text{W}$ e $25^{\circ} - 30^{\circ}\text{S}/25 - 30^{\circ}\text{W}$ (Figura 6).

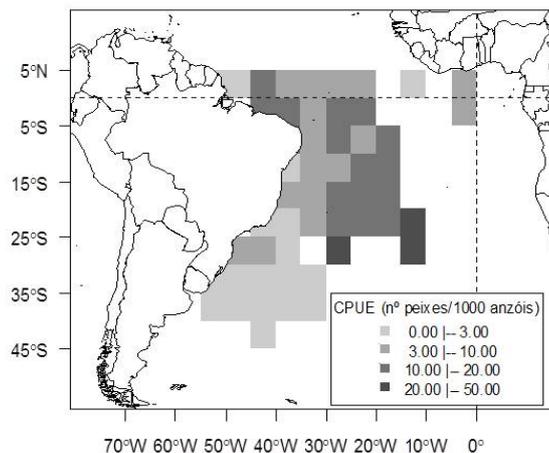


Figura 6: Distribuição geográfica da CPUE nominal de espadarte (número de peixes/1000 anzóis).

Os diagramas de caixa da Figura (7), apresentam as seguintes evidências: (A) as pescarias realizadas após meio dia apresentam maiores CPUEs; (B) não parece haver diferença entre as CPUEs trimestrais; (C) de modo geral o período 2003-2006 apresentam as maiores CPUEs, e percebe-se a presença valores discrepantes nos demais períodos e grande dispersão das CPUEs em 1998

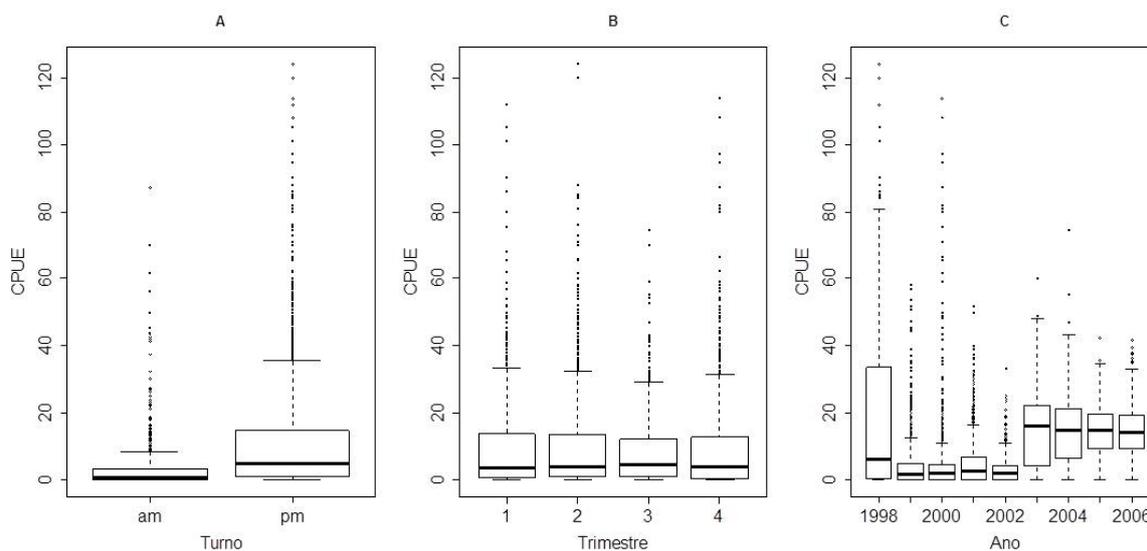


Figura 7: Diagramas de caixa relacionando a CPUE nominal de espadarte (número de peixes/1000 anzóis) e os fatores: (A) turno (am=antes do meio dia, pm=após meio dia), (B) trimestre e (C) ano.

Em se tratando da relação entre a CPUE e as covariáveis, os gráficos de dispersão da Figura (8) trazem algumas suspeitas: (A) a CPUE parece ser maior quando se utiliza 4

anzóis entre boias, havendo um decrescimento após isso; (B) a CPUE apresenta maiores valores próximo a linha do Equador e na latitude 20°S, diminuindo após esta; (C) a CPUE apresenta crescimento discreto até a distância mínima de 1000 km da costa, entretanto, após esta distância as capturas são raras (D) quanto ao tempo de imersão do espinhel na água, as maiores CPUEs ocorrem próximas a 15 horas de imersão.

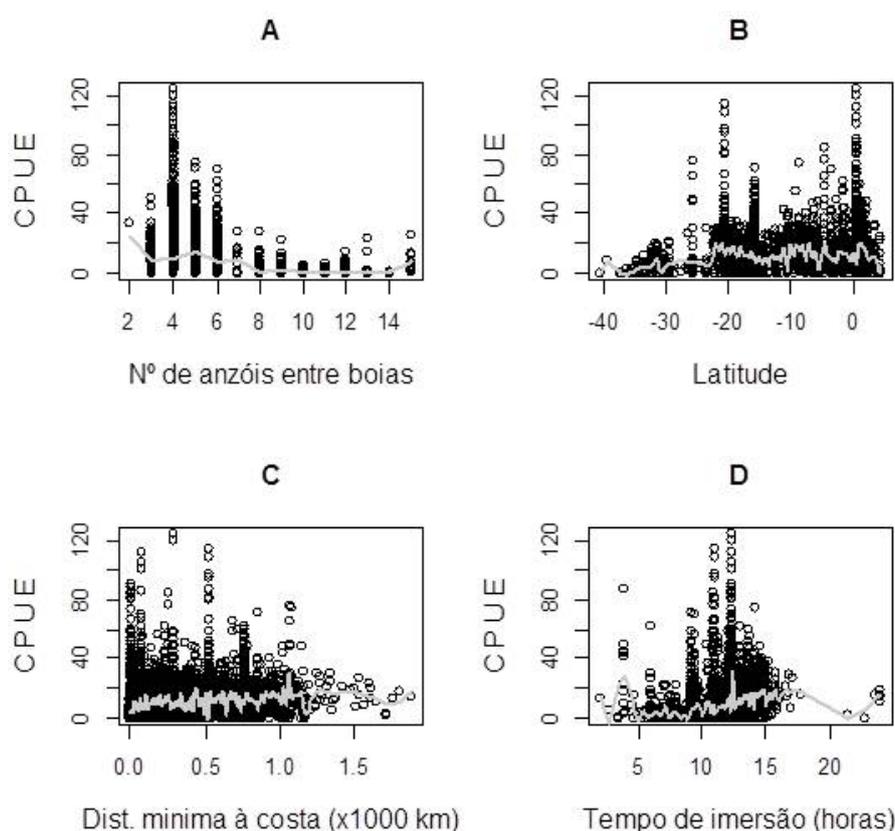


Figura 8: Diagramas de dispersão relacionando a CPUE nominal (número de espadartes/1000 anzóis) versus covariáveis: (A) número de anzóis entre boias, (B) latitude, (C) distância mínima à costa e (D) tempo de imersão do espinhel na água. A linha cinza é uma linha suavizadora (*Locally Weighted Scatterplot Smoother-LOWESS*), que mostra a tendência da CPUE.

3.5.2 Ajuste dos Modelos

Neste contexto, começamos propondo o modelo de Poisson, candidato natural em se tratando de dados de contagens, e posteriormente, levando em conta a sobredispersão e o excesso de zeros, propomos os modelos Binomial Negativo (NB), o Poisson Inflacionado de Zeros (ZIP) e o Binomial Negativo Inflacionado de Zeros (ZINB), todos utilizando como *offset* o logaritmo do número total de anzóis utilizados pelas embarcações. A partir da análise da deviança, testes da razão de verossimilhança, critério

AIC e gráficos de diagnósticos, foram escolhidos os melhores modelos. Uma peculiaridade dos dados sob análise é o desbalanceamento dos eventos de pesca e a ausência de observações em muitos dos cruzamentos de variáveis explicativas, como por exemplo, mostra a Tabela 1, onde percebe-se a inatividade de algumas embarcações durante alguns anos ao longo do período considerado. Tais ausências dificultaram a convergência de algumas funções utilizadas na obtenção dos modelos.

Tabela 1: Número de eventos de pesca por embarcação, 1998-2006.

| | 1998 | 1999 | 2000 | 2001 | 2002 | 2003 | 2004 | 2005 | 2006 |
|----|------|------|------|------|------|------|------|------|------|
| 1 | 6 | 0 | 55 | 30 | 6 | 0 | 16 | 12 | 0 |
| 2 | 204 | 236 | 185 | 252 | 2 | 5 | 47 | 191 | 165 |
| 3 | 0 | 141 | 158 | 88 | 10 | 2 | 54 | 0 | 0 |
| 4 | 0 | 0 | 162 | 207 | 30 | 5 | 94 | 87 | 101 |
| 5 | 73 | 131 | 56 | 254 | 36 | 0 | 0 | 0 | 0 |
| 6 | 0 | 143 | 204 | 100 | 0 | 89 | 58 | 188 | 175 |
| 7 | 116 | 120 | 85 | 61 | 25 | 1 | 13 | 1 | 0 |
| 8 | 0 | 115 | 96 | 87 | 19 | 4 | 22 | 27 | 63 |
| 9 | 16 | 112 | 28 | 55 | 12 | 39 | 53 | 0 | 0 |
| 10 | 59 | 96 | 64 | 74 | 58 | 0 | 0 | 0 | 0 |

3.5.3 Avaliação dos Modelos

O ajuste do modelo de Poisson forneceu desvio residual de 33360 (5424 graus de liberdade) e estimativa do parâmetro de dispersão igual a 8.12 ($\gg 1$), indicando forte indícios de sobredispersão, o que é comprovado pelo teste de Lawless 3.5, onde $S=258.53$ ($p=0.0$). Além disso, o nível descritivo para este desvio residual, $p<0.05$, indica que o modelo não tem ajuste satisfatório. O envelope simulado, mostra grandes discrepâncias entre a distribuição dos resíduos e a distribuição normal padrão (Figura 9). Além disso, este modelo subestimou substancialmente a ocorrência de zeros. Entre os modelos ajustados, este foi o pior modelo, apresentando os maiores valores de $AIC=52956$ e $BIC=53638$ (Tabela 2). Afim de melhor acomodar a extra variação presente nos dados, um modelo Binomial Negativo foi ajustado. Tal ajuste, forneceu parâmetro de dispersão estimado em 1,3649 ($EP=0,036$) e desvio residual de 6295,8 (5401 graus de liberdade), indicando um ajuste mais adequado. O envelope simulado (Figura 10), mostra que as discrepâncias entre a distribuição dos resíduos e a distribuição normal padrão são pequenas, aparecendo somente nas extremidades, evidenciando a adequação da suposta distribuição. Além disso, este modelo apresentou uma grande redução nos

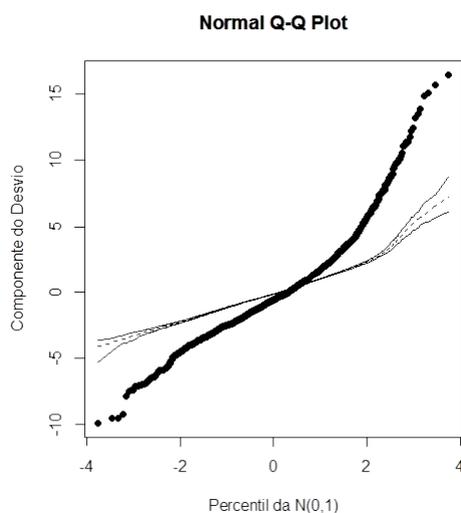


Figura 9: Envelope simulado para os Componentes do Desvio no modelo de Poisson.

valores de AIC e BIC, cujos valores são, respectivamente, 31958 e 32554. Quanto a predição de zeros, a estimativa se aproximou bastante do valor observado. Por não se

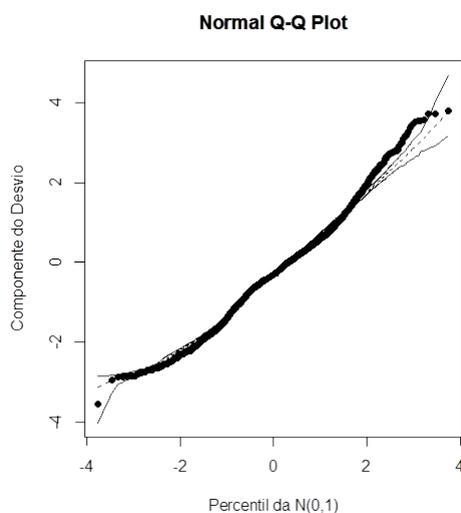


Figura 10: Envelope simulado para os Componentes do Desvio no modelo GLM binomial Negativo.

saber com precisão a quantidade de observações nulas que o modelo de Poisson e o binomial negativo suporta, ajustamos o modelo ZIP na tentativa de acomodar o excesso de zeros presente nos dados e seus efeitos. Tal ajuste, forneceu uma log-verossimilhança de -22913 com 216 graus de liberdade, AIC=46238 e BIC=47603, apresentando, portanto, um ganho pequeno em relação ao modelo de Poisson usual. A partir do envelope simulado (Figura 11) podemos observar que a maioria dos resíduos de Pearson estão fora das bandas de confiança. Quanto ao número de zeros preditos, o modelo ZIP

apresentou valor, 1118, que é praticamente idêntico ao número de zeros observados que é de 1110. Esta superioridade na qualidade do ajuste sobre o modelo de Poisson tradicional foi comprovada pelo teste de Vuong 3.11, onde $V=-12,48$ ($p<0.01$).

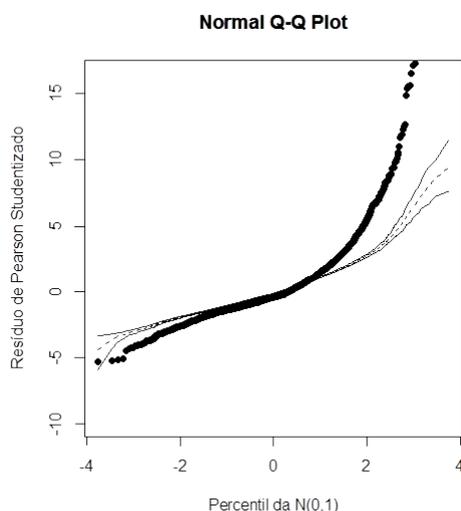


Figura 11: Envelope simulado para os resíduos de Pearson Studentizado no modelo ZIP.

Ainda na tentativa de melhorar o ajuste promovido pelos modelos anteriores, ajustamos um modelo ZINB aos dados. O modelo ZINB forneceu uma log-verossimilhança de -15986 com 103 graus de liberdade, estimativa do parâmetro de dispersão igual a 1,6348 (maior sobredispersão que o NB), $AIC=32178$, $BIC=32860$ e número de zeros preditos 1139, ligeiramente maior que o número de zeros observados. Comparativamente ao modelo de Poisson e ZIP, o modelo ZINB se mostrou bastante superior, reduzindo significativamente o AIC e BIC e assimilando melhor a sobredispersão presente, provocada pelo excesso de zeros. Mesmo assim, o ZINB continuou inferior ao Binomial Negativo, apresentando AIC, BIC superiores e número de zeros preditos sobre estimados. O envelope simulado para o modelo ZINB (Figura 12) revela que a maioria dos resíduos de Pearson estão fora das bandas de confiança, indicando inadequação da distribuição suposta. Esta superioridade na qualidade do ajuste do modelo NB sobre o ZINB foi comprovada pelo teste de Vuong, onde $V= 3.45$ ($p<0.01$).

3.5.4 Modelo utilizado na padronização da CPUE do espadarte

Dentre os quatro melhores GLM ajustados (Poisson, binomial negativo-BN, Poisson inflacionado de zeros-ZIP, e binomial negativo inflacionado de zeros-ZINB), o modelo mais parcimonioso, com menor número de parâmetros, e que melhor descreve os dados

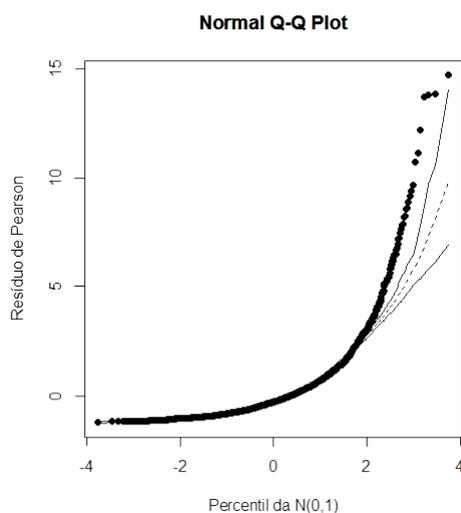


Figura 12: Envelope simulado para os resíduos de Pearson Studentizado no modelo ZINB.

de captura do espadarte no Atlântico Sul, dispostos na base selecionada, foi o binomial negativo, como mostra a Tabela 2. Tal modelo apresentou os menores valores de AIC e BIC, predizendo 82% dos zeros observados.

Tabela 2: Comparativo dos modelos ajustados quanto ao logaritmo da verossimilhança ($\log\text{Lik}$), critério de informação de Akaike (AIC), critério de informação bayesiano (BIC), e o número de zeros preditos, sendo o número de zeros observados igual a 1110.

| | Poisson | NB | ZIP | ZINB |
|------------------|---------|--------|--------|--------|
| Df | 103 | 90 | 206 | 103 |
| $\log\text{Lik}$ | -26375 | -15889 | -22913 | -15986 |
| AIC | 52956 | 31958 | 46238 | 32178 |
| BIC | 53638 | 32554 | 47603 | 32860 |
| Theta | — | 1.3 | — | 1.6 |
| Zeros | 349 | 890 | 1118 | 1139 |

3.5.5 Gráficos de diagnóstico do modelo binomial negativo

O gráfico Componente do Desvio (t_{D_i}) versus valores ajustados (Figura 13), aponta uma dispersão maior perto dos zeros, evidenciando que o modelo, apesar das qualidades apontadas anteriormente e da sua superioridade em relação aos demais, não foi capaz de incorporar toda a extra-variação provocada pelo excesso de zeros ou a falta de independência entre as observações. Quanto ao gráfico da Distância de Cook (LD_i) contra as ordens das observações (distância de Cook), revela a presença de 11 observações (dentre 5579) consideradas influentes. O ajuste realizado sem a presença destas

observações não alterou os resultados inferenciais.

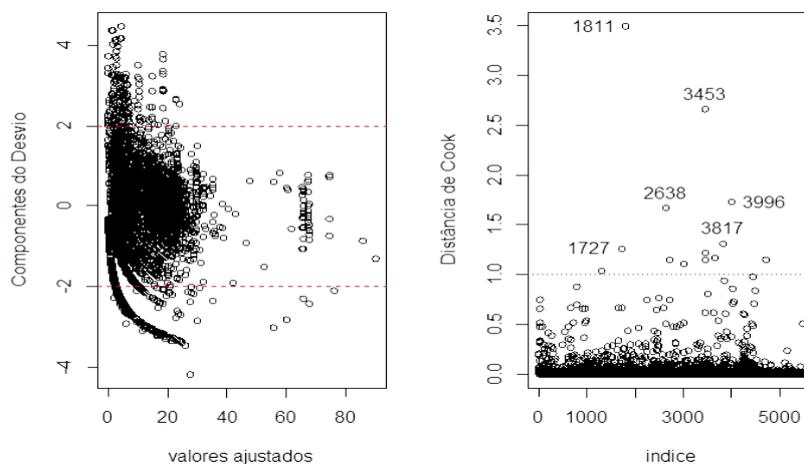


Figura 13: Gráficos de diagnósticos referentes ao modelo binomial negativo utilizado na padronização da CPUE de espadarte. (a) Componente do Desvio versus valores ajustados e (b) Distância de Cook versus a ordem das observações.

3.5.6 Análise de deviância e estimativas dos parâmetros do modelo binomial negativo

A tabela de análise de deviância, referente ao GLM binomial negativo utilizado para a padronização da CPUE de espadarte (Tabela 3), mostra a significância da inclusão sequencial dos termos do modelo ajustado, relativa à redução provocada na deviância. A proporção de deviância explicada pelo modelo foi de aproximadamente, 54,5%, sendo que o fator ano explicou 37% da deviância que foi explicada pelo modelo, seguido da variável número de anzóis entre boias, com 26% da variabilidade.

As estimativas de máxima verossimilhança dos parâmetros do modelo binomial negativo são apresentadas na Tabela 4. De maneira geral, tais estimativas foram significativamente diferentes de zero, ao nível de 5%. As conclusões a respeito do efeito de cada um dos termos do modelo sobre a CPUE foram realizadas a partir dos valores preditos com os coeficientes estimados, fixado os demais. Para dar uma ideia da relação entre CPUE e cada efeito considerado, a média da CPUE marginal foi calculada em diferentes níveis de cada efeito e plotados:

- ano - os coeficientes estimados para o fator ano são todos negativos, sendo que os coeficientes para os anos de 2000 e 2001 não são significativos;

Tabela 3: Análise de deviância para o modelo binomial negativo, utilizado na padronização da CPUE de espadarte capturado pela frota espinheira brasileira, de 1998 a 2006. Para cada termo é apresentado o decréscimo provocado na deviância residual, graus de liberdade e o p-valor do teste. Siglas utilizadas: min - distância mínima à costa (x1000km), lat - latitude, hpb - n^o de anzóis entre boias, tp.ime - tempo de imersão (horas) e significâncias: (***) - 0,001, (**) - 0,01.

| | G.L | Deviância | G.L Resid | Dev. Resid. | Pr(>Chi) |
|---------------|-----|-----------|-----------|-------------|-------------|
| ano | 8 | 2784 | 5494 | 11031 | < 2e-16 *** |
| min | 1 | 49 | 5493 | 10982 | 2.3e-12 *** |
| trimestre | 3 | 57 | 5490 | 10925 | 2.8e-12 *** |
| lat | 1 | 7 | 5489 | 10918 | 0.0089 ** |
| tp.ime | 1 | 204 | 5488 | 10714 | < 2e-16 *** |
| turno | 1 | 358 | 5487 | 10356 | < 2e-16 *** |
| hpb | 1 | 1969 | 5486 | 8387 | < 2e-16 *** |
| ano:min | 8 | 67 | 5478 | 8320 | 1.9e-11 *** |
| ano:trimestre | 23 | 527 | 5455 | 7793 | < 2e-16 *** |
| ano:lat | 8 | 151 | 5447 | 7642 | < 2e-16 *** |
| ano:tp.ime | 8 | 290 | 5439 | 7352 | < 2e-16 *** |
| ano:turno | 8 | 177 | 5431 | 7175 | < 2e-16 *** |
| ano:hpb | 8 | 573 | 5423 | 6602 | < 2e-16 *** |
| min:lat | 1 | 60 | 5422 | 6542 | 7.5e-15 *** |
| min:tp.ime | 1 | 27 | 5421 | 6515 | 2.4e-07 *** |
| min:turno | 1 | 67 | 5420 | 6448 | 2.8e-16 *** |
| min:hpb | 1 | 26 | 5419 | 6422 | 3.4e-07 *** |
| trimestre:lat | 3 | 39 | 5416 | 6384 | 2.0e-08 *** |
| lat:tp.ime | 1 | 35 | 5415 | 6348 | 2.9e-09 *** |
| lat:hpb | 1 | 46 | 5414 | 6302 | 1.3e-11 *** |

- distância mínima à costa - as CPUEs são maiores próximo à costa, decrescendo até 1000 km, voltando a crescer até 2000 km, onde atinge seu valor máximo (Figura 14A);
- trimestre - as CPUEs tem comportamento decrescente ao longo dos trimestres, apresentando valor máximo no primeiro e mínimo no quarto (Figura 14B);
- latitude - as CPUEs são maiores próximo ao Equador decrescendo exponencialmente até 40°S (Figura 14C);
- tempo de imersão do espinhel (tp.ime) - as CPUEs são praticamente nulas até aproximadamente 15 horas de imersão, ascendendo a partir daí até atingir valor máximo em 24hs (Figura 14D);
- turno - as CPUEs apresentaram maiores valores quando se lançou o espinhel entre 0:00 e 12:00 horas (turno am) (Figura 14E);

Tabela 4: Estimativa dos coeficientes dos parâmetros do modelo binomial negativo, utilizado na padronização da CPUE de espadarte capturado pela frota espinheira brasileira. Siglas utilizadas: min - distância mínima à costa (x1000km), lat - latitude, hpb - n° de anzóis entre boias, tp.ime - tempo de imersão (horas) e significâncias: (***) - 0,001, (**) - 0,01, (*) - 0,05 e (.) - 0,1.

| | Estimativa | Erro Padrão | Valor z | Pr(> z) |
|--------------------|------------|-------------|---------|-------------|
| (Intercept) | -4.600227 | 0.937730 | -4.91 | 9.3e-07 *** |
| ano1999 | -3.956174 | 0.974367 | -4.06 | 4.9e-05 *** |
| ano2000 | -1.005021 | 0.967941 | -1.04 | 0.29913 |
| ano2001 | -1.046253 | 0.962029 | -1.09 | 0.27679 |
| ano2002 | -2.964873 | 1.128953 | -2.63 | 0.00863 ** |
| ano2003 | -6.311090 | 1.111623 | -5.68 | 1.4e-08 *** |
| ano2004 | -2.113316 | 1.030432 | -2.05 | 0.04028 * |
| ano2005 | -4.559422 | 1.125408 | -4.05 | 5.1e-05 *** |
| ano2006 | -5.328908 | 1.185548 | -4.49 | 7.0e-06 *** |
| min | 1.947008 | 0.493805 | 3.94 | 8.1e-05 *** |
| trimestre2 | 0.004288 | 0.142229 | 0.03 | 0.97595 |
| trimestre3 | -0.355443 | 0.159404 | -2.23 | 0.02576 * |
| trimestre4 | -1.076842 | 0.149176 | -7.22 | 5.3e-13 *** |
| lat | -0.095225 | 0.017900 | -5.32 | 1.0e-07 *** |
| tp.ime | 0.555452 | 0.080861 | 6.87 | 6.5e-12 *** |
| turnopm | -2.378292 | 0.294687 | -8.07 | 7.0e-16 *** |
| hpb | -0.614372 | 0.047372 | -12.97 | < 2e-16 *** |
| ano1999:min | -0.243763 | 0.281289 | -0.87 | 0.38616 |
| ano2000:min | -0.211121 | 0.285747 | -0.74 | 0.46000 |
| ano2001:min | 0.445662 | 0.275655 | 1.62 | 0.10593 |
| ano2002:min | 0.010746 | 0.361941 | 0.03 | 0.97632 |
| ano2003:min | 1.255981 | 0.372404 | 3.37 | 0.00074 *** |
| ano2004:min | 0.572985 | 0.288283 | 1.99 | 0.04686 * |
| ano2005:min | 0.298777 | 0.287279 | 1.04 | 0.29833 |
| ano2006:min | 0.362092 | 0.282638 | 1.28 | 0.20015 |
| ano1999:trimestre2 | 0.393792 | 0.168359 | 2.34 | 0.01934 * |
| ano2000:trimestre2 | -0.301914 | 0.160299 | -1.88 | 0.05964 . |
| ano2001:trimestre2 | 0.274829 | 0.161424 | 1.70 | 0.08866 . |
| ano2002:trimestre2 | -0.975491 | 0.391986 | -2.49 | 0.01283 * |
| ano2003:trimestre2 | 0.162671 | 0.259359 | 0.63 | 0.53052 |
| ano2004:trimestre2 | 0.297301 | 0.247842 | 1.20 | 0.23031 |
| ano2005:trimestre2 | 0.394219 | 0.211424 | 1.86 | 0.06224 . |
| ano2006:trimestre2 | 0.389301 | 0.185835 | 2.09 | 0.03618 * |
| ano1999:trimestre3 | 0.600919 | 0.188613 | 3.19 | 0.00144 ** |
| ano2000:trimestre3 | -0.305794 | 0.176857 | -1.73 | 0.08380 . |
| ano2001:trimestre3 | 0.284775 | 0.176744 | 1.61 | 0.10713 |
| ano2002:trimestre3 | -0.743880 | 0.355624 | -2.09 | 0.03646 * |
| ano2003:trimestre3 | 1.011539 | 0.499058 | 2.03 | 0.04267 * |
| ano2004:trimestre3 | 1.301685 | 0.230143 | 5.66 | 1.5e-08 *** |
| ano2005:trimestre3 | 0.346633 | 0.234268 | 1.48 | 0.13897 |
| ano2006:trimestre3 | 0.331576 | 0.217222 | 1.53 | 0.12690 |
| ano1999:trimestre4 | 1.705152 | 0.169724 | 10.05 | < 2e-16 *** |
| ano2000:trimestre4 | 0.940072 | 0.173436 | 5.42 | 6.0e-08 *** |
| ano2001:trimestre4 | 0.789673 | 0.171084 | 4.62 | 3.9e-06 *** |

- n° de anzóis entre boias (hpb) - as CPUEs são maiores quando se utiliza quatro anzóis, decrescendo a partir daí (Figura 14F).

Tabela 5: Continuação da tabela 4. Estimativa dos coeficientes dos parâmetros do modelo binomial negativo, utilizado na padronização da CPUE de espadarte capturado pela frota espinheleira brasileira. Siglas utilizadas: min - distância mínima à costa (x1000km), lat - latitude, hpb - nº de anzóis entre boias, tp.ime - tempo de imersão (horas) e significância: 0^{***}, 0.001^{**}, 0.01^{*}, 0.05[.].

| | Estimativa | Erro Padrão | Valor z | Pr(> z) |
|--------------------|------------|-------------|---------|-------------|
| ano2002:trimestre4 | 0.764407 | 0.310436 | 2.46 | 0.01380 * |
| ano2004:trimestre4 | 2.101471 | 0.237897 | 8.83 | < 2e-16 *** |
| ano2005:trimestre4 | 1.171458 | 0.217563 | 5.38 | 7.3e-08 *** |
| ano2006:trimestre4 | 0.838769 | 0.184976 | 4.53 | 5.8e-06 *** |
| ano1999:lat | -0.046712 | 0.010682 | -4.37 | 1.2e-05 *** |
| ano2000:lat | -0.046752 | 0.010813 | -4.32 | 1.5e-05 *** |
| ano2001:lat | -0.041858 | 0.010032 | -4.17 | 3.0e-05 *** |
| ano2002:lat | -0.032545 | 0.011663 | -2.79 | 0.00527 ** |
| ano2003:lat | 0.006973 | 0.016120 | 0.43 | 0.66530 |
| ano2004:lat | -0.017106 | 0.012111 | -1.41 | 0.15783 |
| ano2005:lat | -0.052034 | 0.011231 | -4.63 | 3.6e-06 *** |
| ano2006:lat | -0.034018 | 0.010331 | -3.29 | 0.00099 *** |
| ano1999:tp.ime | -0.294608 | 0.085642 | -3.44 | 0.00058 *** |
| ano2000:tp.ime | -0.486835 | 0.084110 | -5.79 | 7.1e-09 *** |
| ano2001:tp.ime | -0.425795 | 0.085009 | -5.01 | 5.5e-07 *** |
| ano2002:tp.ime | -0.391095 | 0.088683 | -4.41 | 1.0e-05 *** |
| ano2003:tp.ime | -0.358718 | 0.107113 | -3.35 | 0.00081 *** |
| ano2004:tp.ime | -0.449140 | 0.089450 | -5.02 | 5.1e-07 *** |
| ano2005:tp.ime | -0.274707 | 0.089601 | -3.07 | 0.00217 ** |
| ano2006:tp.ime | -0.218761 | 0.088733 | -2.47 | 0.01369 * |
| ano1999:turnopm | 2.466443 | 0.334227 | 7.38 | 1.6e-13 *** |
| ano2000:turnopm | 2.102864 | 0.331145 | 6.35 | 2.1e-10 *** |
| ano2001:turnopm | 2.994916 | 0.345983 | 8.66 | < 2e-16 *** |
| ano2002:turnopm | 3.373227 | 0.433638 | 7.78 | 7.3e-15 *** |
| ano2003:turnopm | 2.415347 | 0.646036 | 3.74 | 0.00018 *** |
| ano2004:turnopm | 3.122316 | 0.355663 | 8.78 | < 2e-16 *** |
| ano2005:turnopm | 3.458456 | 0.373817 | 9.25 | < 2e-16 *** |
| ano2006:turnopm | 1.932028 | 0.554518 | 3.48 | 0.00049 *** |
| ano1999:hpb | 0.541812 | 0.050385 | 10.75 | < 2e-16 *** |
| ano2000:hpb | 0.509892 | 0.054305 | 9.39 | < 2e-16 *** |
| ano2001:hpb | 0.267557 | 0.048077 | 5.57 | 2.6e-08 *** |
| ano2002:hpb | 0.613840 | 0.069125 | 8.88 | < 2e-16 *** |
| ano2003:hpb | 1.458261 | 0.119222 | 12.23 | < 2e-16 *** |
| ano2004:hpb | 0.613787 | 0.052707 | 11.65 | < 2e-16 *** |
| ano2005:hpb | 0.641334 | 0.082855 | 7.74 | 9.9e-15 *** |
| ano2006:hpb | 0.971427 | 0.083940 | 11.57 | < 2e-16 *** |
| min:lat | -0.056195 | 0.006803 | -8.26 | < 2e-16 *** |
| min:tp.ime | -0.241524 | 0.041043 | -5.88 | 4.0e-09 *** |
| min:turnopm | 1.087345 | 0.215627 | 5.04 | 4.6e-07 *** |
| min:hpb | -0.136076 | 0.030331 | -4.49 | 7.2e-06 *** |
| trimestre2:lat | 0.025048 | 0.004915 | 5.10 | 3.5e-07 *** |
| trimestre3:lat | 0.015457 | 0.005356 | 2.89 | 0.00390 ** |
| trimestre4:lat | 0.003076 | 0.005168 | 0.60 | 0.55173 |
| lat:tp.ime | 0.008708 | 0.001194 | 7.29 | 3.1e-13 *** |
| lat:hpb | 0.006364 | 0.000926 | 6.87 | 6.4e-12 *** |

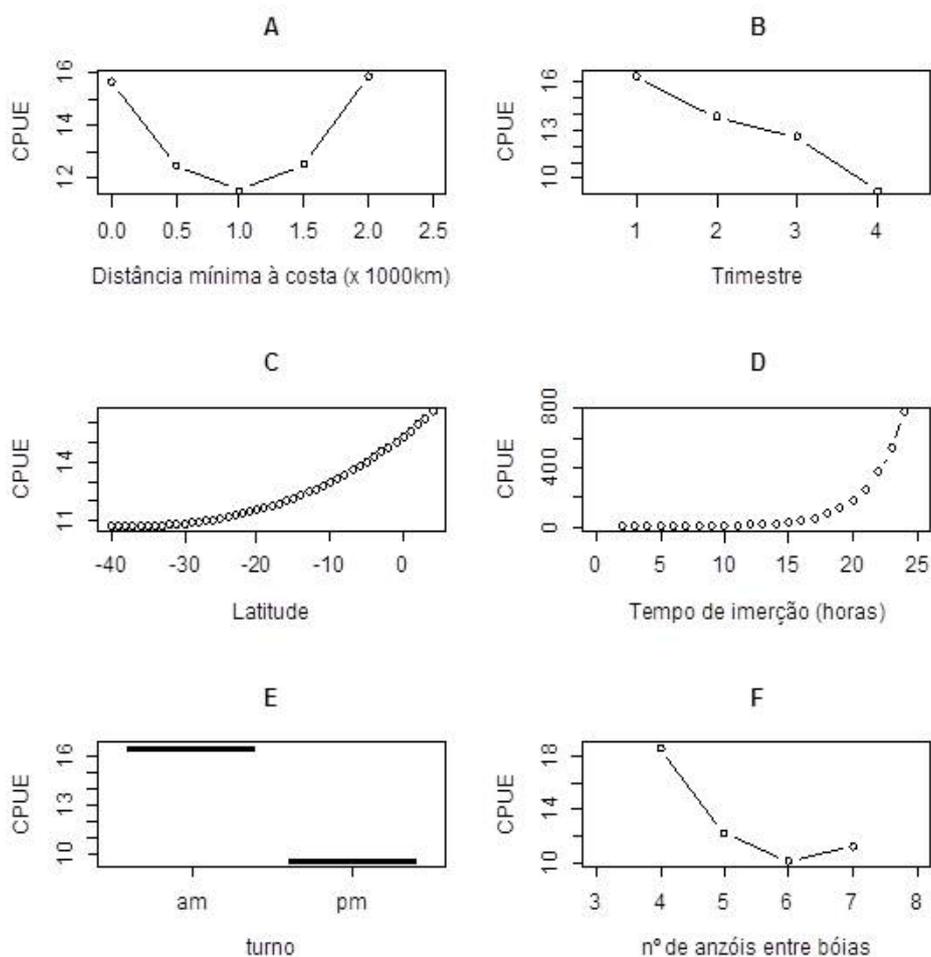


Figura 14: Efeito isolado dos termos principais do modelo binomial negativo sobre captura por unidade de esforço (CPUE) padronizada de espadarte (n° de peixes/1000 anzóis): A) distância mínima à costa (x1000 km), B) trimestre, C) latitude, D) tempo de imersão (em horas), E) turno e F) n° de anzóis entre bóias.

- ano:min - com exceção dos anos de 1998, 2003 e 2004, as CPUEs anuais não se diferenciaram significativamente quanto à distância mínima à costa (Figura 15);
- ano:trimestre - de modo geral, as CPUEs trimestrais dependem dos anos em que ocorreram, por exemplo, em 1998 a CPUE foi significativamente maior no primeiro, em 2003 foi maior no terceiro, em 2004 a CPUE foi maior no quarto e nos demais anos não se diferenciaram significativamente (Figura 16);
- ano:lat - com exceção de 1998, as CPUEs anuais não se diferenciaram significativamente com a latitude;
- ano:tp.ime - com exceção de 1998, as CPUEs anuais não se diferenciaram significativamente em se tratando do tempo de imersão do espinhel na água;

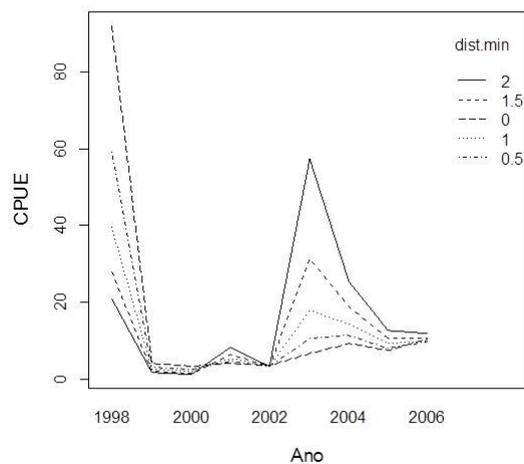


Figura 15: Efeito da interação ano:distância mínima à costa sobre captura por unidade de esforço (CPUE) padronizada de espadarte (nº de peixes/1000 anzóis), derivado de um modelo binomial negativo.

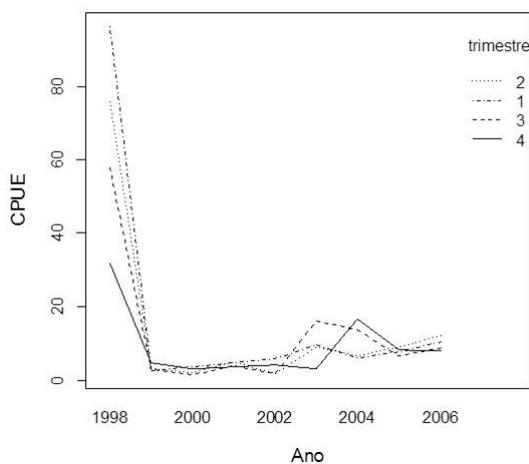


Figura 16: Efeito da interação ano:trimestre sobre captura por unidade de esforço (CPUE) padronizada de espadarte (nº de peixes/1000 anzóis), derivado de um modelo binomial negativo.

- ano:turno - com exceção de 1998, as CPUE anuais apresentaram maiores resultados quando os lançamentos ocorreram no turno após o meio dia (Figura 17);
- ano:hpb - em 1998 a CPUE foi significativamente maior quando se utilizou quatro anzóis, em 2003 quando se utilizou sete e nos demais anos não se diferenciaram quanto ao número de anzóis entre boias (Figura 18);

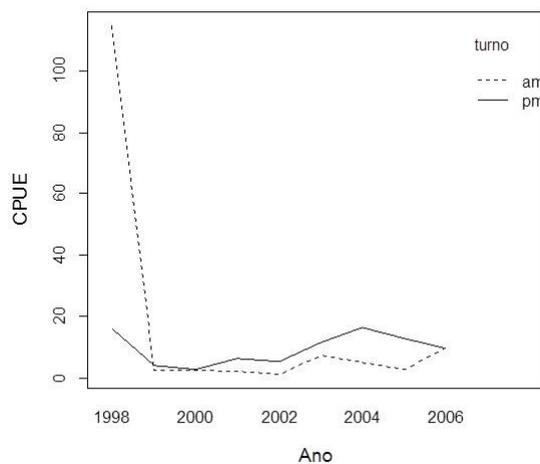


Figura 17: Efeito da interação ano:turno sobre captura por unidade de esforço (CPUE) padronizada de espadarte (n° de peixes/1000 anzóis), derivado de um modelo binomial negativo.

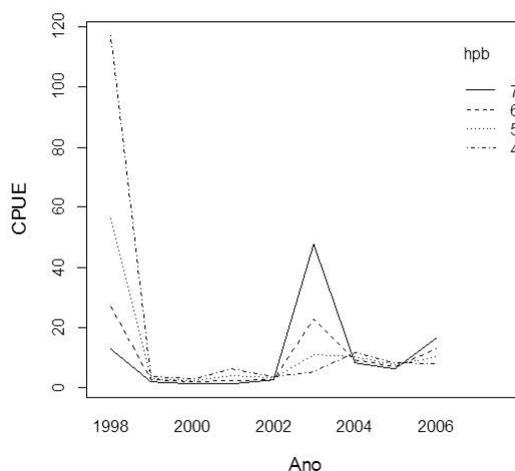


Figura 18: Efeito da interação ano: n° de anzóis entre boias sobre captura por unidade de esforço (CPUE) padronizada de espadarte (n° de peixes/1000 anzóis), derivado de um modelo binomial negativo.

- min:lat - entre as latitudes 10°S e 40°S as CPUEs foram maiores a 2000 km de distância mínima à costa, já entre 10°S a 5°N foram maiores próximo à costa (Figura 19);
- min:tp.ime - para o tempo de imersão superior a 15 horas, as CPUEs são maiores quando se aproxima da costa e para tempos inferiores foram maiores a 2000km da costa (Figura 20);

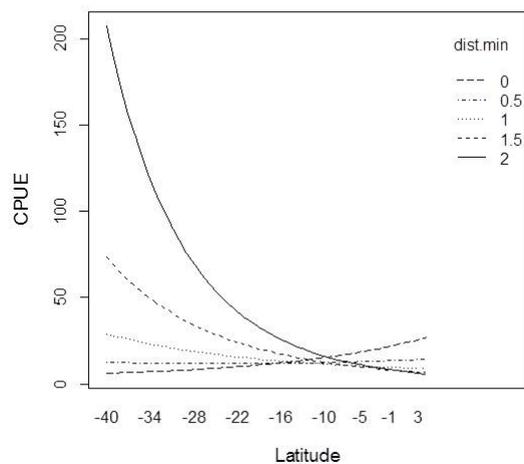


Figura 19: Efeito da interação distância mínima à costa:latitude sobre captura por unidade de esforço (CPUE) padronizada de espadarte (n° de peixes/1000 anzóis), derivado de um modelo binomial negativo.

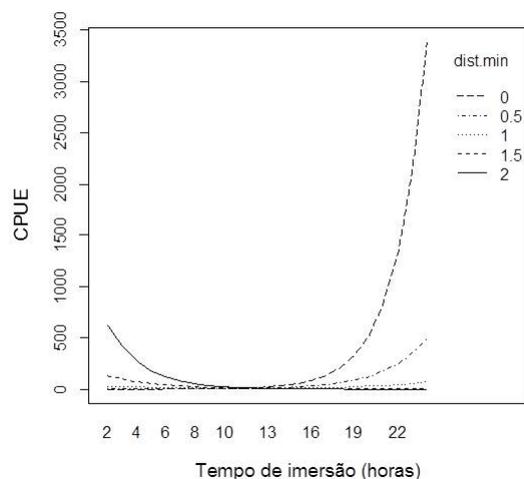


Figura 20: Efeito da interação distância mínima à costa:tempo de imersão do espinhel sobre captura por unidade de esforço (CPUE) padronizada de espadarte (n° de peixes/1000 anzóis), derivado de um modelo binomial negativo.

- min:turno - até aproximadamente 800 metros de distância mínima à costa, as CPUEs são maiores quando se lança o espinhel entre 0:00 hora e 12:00hs, após essa distância as maiores CPUEs ocorrem no turno inverso (Figura 21);
- min:hpb - até 1500 metros de distância mínima à costa as CPUEs são maiores quando se utiliza quatro anzóis entre boias, após esta distância são maiores quando se utiliza sete (Figura 22);

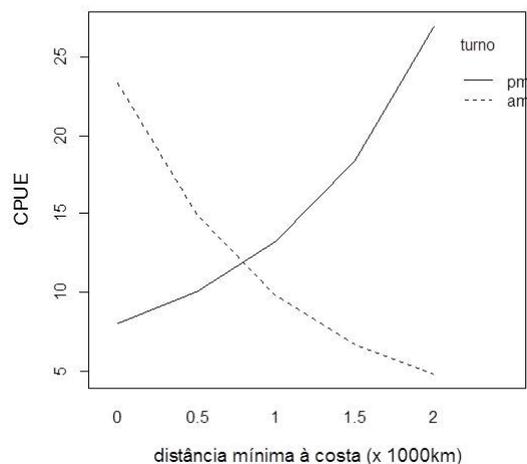


Figura 21: Efeito da interação distância mínima à costa:turno sobre captura por unidade de esforço (CPUE) padronizada de espadarte (n° de peixes/1000 anzóis), derivado de um modelo binomial negativo.

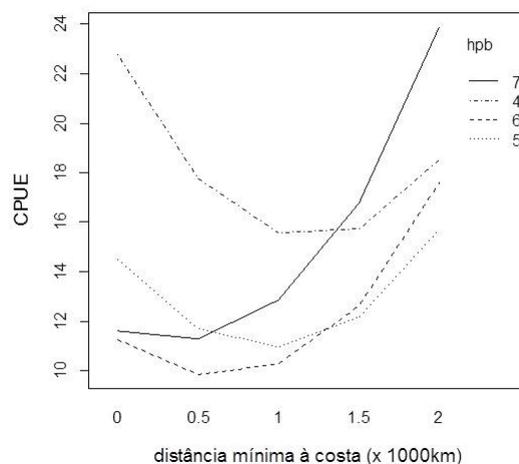


Figura 22: Efeito da interação distância mínima à costa: n° de anzóis entre boias sobre captura por unidade de esforço (CPUE) padronizada de espadarte (n° de peixes/1000 anzóis), derivado de um modelo binomial negativo.

- trimestre:lat - entre as latitudes 5°S e 40°S as CPUEs são maiores no primeiro trimestre e próximo ao Equador, são maiores no segundo (Figura 23);
- lat:tp.ime - em qualquer latitude as CPUEs são maiores para 24 hs de imersão;
- lat:hpb - em qualquer latitude as CPUEs são maiores quando se utilizou quatro anzóis entre boias(Figura 24);

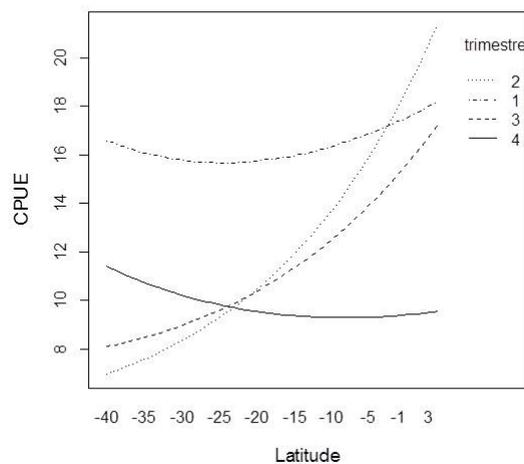


Figura 23: Efeito da interação trimestre:latitude sobre captura por unidade de esforço (CPUE) padronizada de espadarte (n° de peixes/1000 anzóis), derivado de um modelo binomial negativo.

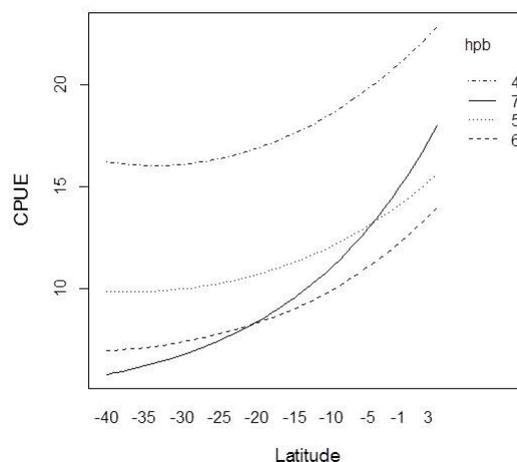


Figura 24: Efeito da interação latitude:nº de anzóis entre boias sobre captura por unidade de esforço (CPUE) padronizada de espadarte (n° de peixes/1000 anzóis), derivado de um modelo binomial negativo.

3.5.7 CPUE padronizada

Como o modelo utilizado para obtenção da CPUE padronizada contém interações envolvendo o fator ano, o processo de padronização foi executado considerando a média ponderada de tais interações ao longo dos anos, conforme descrito em Maunder e Punt (2004). A tendência da CPUE nominal difere moderadamente do índice de abundância padronizado, com as maiores discrepâncias em 1998, onde a CPUE padronizada

foi bem maior que o valor nominal, e em 2003, onde o crescimento da CPUE nominal foi mais acentuado, mantendo-se estável e superior ao índice padronizado até 2006 (Figura 25B). Ressalta-se que, a tendência da CPUE nominal não reflete verdadeiramente as mudanças ocorridas na biomassa do estoque de espadarte do Atlântico Sul, mas sim uma mistura de fatores relacionados com o coeficiente de capturabilidade e a abundância.

A série de CPUE padronizada assumiu valor extremamente alto no início da série histórica estudada, em 1998, seguindo-se de uma queda acentuada até 1999. A partir daí, observou-se pouca variação até 2002, um crescimento gradativo em 2003 e 2004, a partir daí o índice oscilou em torno de um patamar estável, apresentando decréscimo e crescimento discretos, em 2005 e 2006, respectivamente. Contudo, a estimativa do índice padronizado para 1998 é bastante imprecisa (Figura 25A).

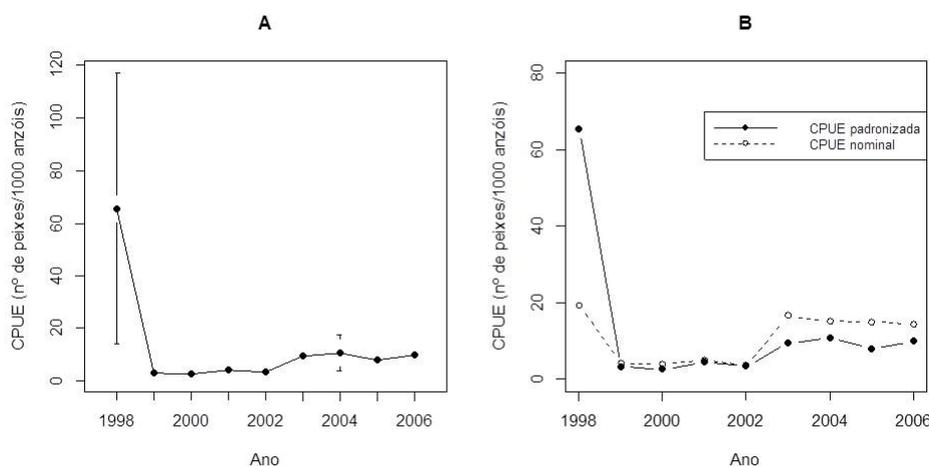


Figura 25: Captura por unidade de esforço (CPUE) nominal e padronizada obtida através do modelo binomial negativo para o espadarte capturado pela frota atuneira brasileira no Atlântico Sul, de 1998 a 2006.

3.6 Discussão

O modelo com a distribuição binomial negativa apresentou os melhores resultados, com respeito ao critério de informação AIC, sendo superior aos modelos inflacionados de zeros. Esta superioridade é devida a acomodação da elevada variabilidade presente nos dados, pelo parâmetro de dispersão existente no modelo e ao fato de que a alta incidência de capturas zero não é provavelmente devida a zeros verdadeiros, já que os pescadores só efetuam pescarias em habitats onde supostamente ocorrem espadartes.

A proporção de deviência explicada pelo modelo binomial negativo (54,5%) pode ser considerada elevada em comparação com outros trabalhos envolvendo a padronização de CPUE de espadarte, visto que os dados aqui utilizados na padronização da CPUE estão desagregados (sem perda de informação), ou seja, são referentes a cada evento de pesca. PUNT et al. (2000) e MAUNDER e PUNT (2004) comentam que o agrupamento dos dados (por ano ou trimestre, por exemplo) elimina grande parte da variabilidade, resultando em altas estimativas para o pseudo R^2 .

Para uma melhor comparação visual com outras series de CPUEs padronizadas para o Espadarte do Atlântico Sul apresentadas em trabalhos recentes, as séries de índices padronizados foram escalonadas a partir da divisão de cada valor pelas respectivas médias (Figura 26). Apesar do valor extremamente alto em 1998, a série CPUE padronizada gerada nesta análise, apresenta tendência similar com as séries obtidas em Hazin et al. (2007) e Hazin et al. (2014) entre 1998 e 2003, atingindo valor máximo em 1998, seguindo-se uma queda acentuada até 1999, a partir daí, pouca variação até 2002 e crescimento em 2003. De 2003 a 2005 as tendências são bastante contraditórias, e o crescimento no final da série em 2006, apesar de discreto, assemelha-se aos apresentados por Hazin et al. (2010) e Hazin et al. (2014)

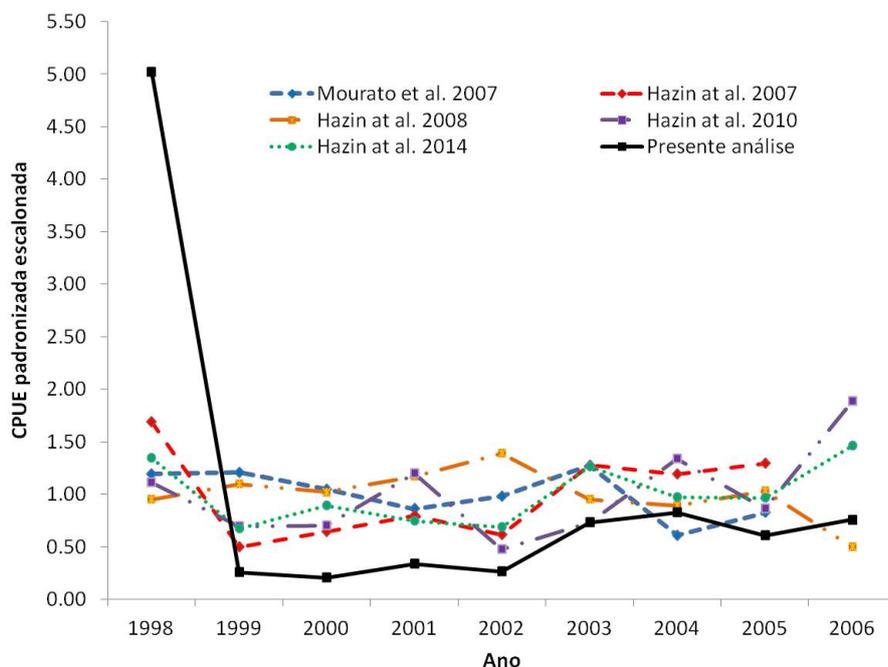


Figura 26: Comparação entre CPUEs padronizadas para o espadarte capturado pela frota atuneira brasileira no Atlântico Sul, de 1998 a 2006.

Atribuímos o elevado e impreciso valor da estimativa do índice padronizado em

1998, ao comportamento atípico de dois barcos amostrados, pertencentes à frota arrendada da Espanha e Estados Unidos no referido ano. Conclui-se que, apesar das oscilações e falta de precisão da estimativa em 1998, os resultados da presente análise confirmam o cenário otimista de uma tendência estável na abundância do estoque da espécie no Atlântico Sul, nos últimos quatro anos da série, apresentado pelo Comitê de Pesquisa e Estatística (SCRS) na reunião da ICCAT de avaliação do estoque de espadarte capturado no oceano Atlântico, realizada em 2006 (ICCAT, 2007).

Os valores preditos com os coeficientes estimados do modelo revelaram que as maiores CPUEs ocorreram durante o primeiro trimestre. Este resultado não corrobora com o apresentado em Mourato et al. (2007), que afirmam que o terceiro trimestre é mais produtivo para a captura no Atlântico Sul. Contudo, os dados analisados pelos autores são advindos da frota atuneira de Santos e Guarujá atuantes no Sudeste e Sul do Brasil. Hazin (2006) afirma que o espadarte apresenta um ciclo sazonal bem definido no oeste do oceano Atlântico, distribuindo-se em águas mais quentes no segundo trimestre, migrando para águas mais frias ao sul onde permanecem durante o terceiro trimestre. Este fato se deve a maior disponibilidade de nutrientes promovida pela convergência subtropical presente na região Sul, e o conseqüentemente aumento de presas potenciais para o espadarte, como as lulas (*llex argentinus*) que são abundantes na região e lá permanecem do final do segundo ao terceiro trimestre (Santos e Haimovici, 2002; REVIZEE, 2006).

De maneira geral, as CPUEs foram maiores quando se lançou o espinhel antes do meio dia (turno am), contudo, quando levou-se em conta as CPUEs anuais, observou-se que este fato ocorreu somente em 1998, nos demais anos houve um aumento da CPUE quando se lançou o espinhel no após o meio dia (turno pm), em particular, a distâncias superiores a 800 metros da costa, quando se utilizou quatro anzóis entre boias. Com esta configuração, os anzóis se posicionam mais superficialmente na coluna da água. Hazin (2006) comenta que o espadarte migra para a superfície durante a noite, e em vista disso, os lançamentos de pesca iniciados no final da tarde, com anzóis posicionados mais superficialmente na coluna da água, com um número de anzóis entre boias igual a quatro, resultam em elevadas capturas da espécie.

É difícil saber qual a causa exata da sobredispersão presente nos dados, entretanto, Hinde e Demetrio (1998) afirmam que a sobredispersão pode ser devido à falta de independência entre as observações. Este pode ser o caso dos dados de captura analisados, visto que as observações provenientes dos mesmos barcos em momentos diferentes, provavelmente, são correlacionadas e isso não é levado em conta quando se utiliza de

modelos GLM. Ainda segundo os autores, à violação da suposição de independência imposta no GLM pode trazer como consequência, inferências incorretas dos parâmetros da regressão devido a erros padrões subestimados e estimadores ineficientes.

Segundo os últimos relatórios sobre avaliação do estoque de espadarte no Atlântico Sul elaborado pelo Comitê Permanente de Pesquisa e Estatística da ICCAT (ICCAT, 2007, 2010 e 2014), apesar da melhoria no nível de informação disponível, as CPUEs padronizadas disponíveis para o espadarte do Atlântico Sul, obtidas por diferentes técnicas, ainda apresentam um indesejável elevado grau de incerteza e tendências contraditórias. Portanto, recomenda-se que mais pesquisas sejam realizadas para diminuir a elevada incerteza no status das ações e mensuração da abundância dos estoques.

Diante desta problemática, realizamos no próximo capítulo, uma análise do tipo longitudinal, em que supõe-se que exista uma correlação não nula entre as observações aninhadas em uma mesma embarcação e um padrão de variação individual na análise das informações, possibilitando assim, um aumento na qualidade das interpretações de casualidade envolvidas na padronização da CPUE de espadarte do Atlântico Sul.

3.7 Conclusão

Considerando a estimativa de CPUE padronizada de espadarte como um índice de abundância relativo, os resultados sugerem que a biomassa do estoque de espadarte do Atlântico Sul apresenta um padrão de estabilidade. Sendo assim, a pesca atuneira no Atlântico Sul parece não estar afetando fortemente o estoque da espécie, resultado que concorda com as conclusões e recomendações da ICCAT quanto à avaliação do referido estoque. Contudo, esta interpretação deve ser tomada com cautela, pois se analisaram dados relativos apenas ao período 1998 a 2006. Para se obter uma interpretação do estado atual do estoque, seria necessário estudos incluindo os anos mais recentes.

O modelo binomial negativo se mostrou adequado para a padronização de dados de captura e esforço de pesca com porcentagem moderada de capturas nulas e presença de sobredispersão, particularmente comum para espécies pelágicas capturadas por espinhéis de superfície, possibilitando a mensuração de efeitos de variáveis espaciais, temporais e operacionais sobre a CPUE de espadarte.

4 Modelagem da CPUE de espadarte sob o enfoque longitudinal

4.1 Introdução

Os modelos lineares generalizados (GLMs) têm sido o método mais largamente utilizado para modelagem de dados provenientes de eventos de pesca e padronização da captura por unidade de esforço (CPUE), útil na avaliação de estoques, necessários para a gestão sustentável das pescarias. No âmbito dos GLMs, dentre outras características, supõe-se que a variável resposta tenha uma distribuição pertencente à família exponencial e a amostra seja composta por observações independentes. Entretanto, esta suposição de independência pode não ser verificada em determinadas situações, como em pesquisas que envolvem observações correlacionadas, provenientes de dados longitudinais. Os dados longitudinais são caracterizados pelo fato de que observações repetidas para uma mesma unidade tendem a ser mais semelhantes do que observações realizadas em indivíduos diferentes, e, portanto, tendem a estar correlacionadas (ZEGGER *et. al.*, 1988).

Em estudos como os que envolvem dados de captura de peixes pelágicos, em particular de espadarte (*Xiphias gladius*), objeto de estudo deste trabalho, é óbvio que o conhecimento empírico dos pescadores sobre o fenômeno, influencia na procura de áreas mais produtivas para a pesca, e que a eficiência de uma determinada embarcação em um determinado evento de pesca, pode ser influenciada pelos eventos anteriores. Sendo assim, é razoável supor que, apesar das embarcações estudadas serem independentes, exista uma correlação não nula entre as observações aninhadas em uma mesma embarcação em ocasiões distintas, levando à violação da suposição de independência imposta nos GLMs. DIGGLE *et al.* (2002) e BALLINGER (2004) mostram que, ao modelar dados longitudinais com o GLM tradicional, isto é, com a suposição de independência

entre observações, tem-se como consequência inferências incorretas dos parâmetros da regressão devido a erros padrões subestimados e estimadores ineficientes.

Neste capítulo será desenvolvida uma análise dos dados de captura de espadarte, sob o enfoque longitudinal. Para tanto, iremos utilizar a teoria das Equações de Estimacão Generalizadas (GEE), proposta por LIANG e ZEGER (1986) e ZEGER e LIANG (1986), para estender os modelos lineares generalizados (NELDER E WEDDERBURN, 1972) e incorporar a correlação existente entre as observações intraunidades de investigação, realizadas ao longo do tempo, na análise. Esta extensão do GLM via GEE, é conhecida na literatura estatística como modelo marginal ou modelo GEE, esta última denominação será utilizada daqui em diante. Neste tipo de modelo a avaliação da relação entre a variável resposta, de natureza discreta ou contínua, e as variáveis explicativas, tempo-dependentes ou não, é realizada em um contexto populacional, e não individual, onde a média e a covariância são modeladas separadamente (DIGGLE *et al.*, 2002). Os GEEs permitem modelar a estrutura de dependência entre as observações intraunidades de investigação, incluindo na análise, uma matriz de correlação de "trabalho" para a obtenção de estimativas consistentes e não-viciadas.

Neste capítulo, além da compreensão das potencialidades dos modelos GEE como metodologia para modelagem dos dados de captura e esforço de pesca, referentes à pescaria brasileira do espadarte com espinhel de superfície, objetivamos caracterizar a trajetória da CPUE de espadarte no período estudado, avaliando sua tendência, assim como mudanças globais e individuais ao longo do período; e, a produção de uma série de CPUE padronizada para o espadarte, em escala mensal, a partir dos modelos GEE e o confronto dos resultados com os obtidos a partir do GLM, a partir da mesma base de dados.

4.2 Dados Longitudinais

Em estudos longitudinais, os dados associados a cada unidade de investigação podem ser expressos em termos de um vetor com os valores observados das variáveis respostas em cada instante e de uma matriz com os correspondentes valores das variáveis explicativas ou covariáveis. A estrutura básica de um estudo longitudinal é apresentada na tabela (6). Nesse caso, suponha que a unidade de investigação i de uma amostra aleatória, $i = 1, 2, \dots, K$, seja observada em t_i instantes de observação, segundo o vetor, denotado *perfil individual de resposta*, $\mathbf{Y}_i = (y_{i1}, y_{i2}, \dots, y_{it_i})'$, de dimensão $t_i \times 1$, e a matriz de covariáveis, $\mathbf{X}_i = (\mathbf{x}_{i1}, \mathbf{x}_{i2}, \dots, \mathbf{x}_{it_i})'$, de dimensão

$t_i \times p$, sendo $\mathbf{x}_{ij} = (x_{ij1}, x_{ij2}, \dots, x_{ijp})'$ o vetor que contém os valores das p covariáveis, sob estudo, para o indivíduo i , em seu j -ésimo instante de observação, t_{ij} , $j = 1, 2, \dots, t_i$. Neste caso, o número total de observações será dado por $N = \sum_{i=1}^K t_i$. Para $t_{1j} = t_{2j} = \dots = t_{nj}$, $\forall j \in \{1, 2, \dots, t_i\}$, teremos um estudo balanceado em relação ao tempo. Para facilitar a notação, vamos assumir, sem perda de generalidade, que o número de observações da medida longitudinal é constante para todos os indivíduos, ou seja, $t_i = t$, $\forall i \in \{1, 2, \dots, n\}$. Quanto às variáveis explicativas, estas podem ser constantes no tempo, apenas variando entre unidades de observação ou podem ser tempo-dependentes, variando também dentro das unidades de observação.

Finalmente, admitimos que os vetores de respostas multivariadas, \mathbf{Y}_i , sejam mutuamente independentes, de esperança marginal $E(y_{ij}) = \mu_{ij}$, com $\boldsymbol{\mu}_i = (\mu_{i1}, \mu_{i2}, \dots, \mu_{it_i})'$ o correspondente vetor de médias e a variância seja definida como uma função da média, $Var(y_{ij}) = \phi^{-1}v(\mu_{ij})$.

Tabela 6: Estrutura de dados básica para estudos com desenho longitudinal.

| Unidade Indivíduos | Instante da Avaliação | Resposta Y | Covariáveis | | | |
|-----------------------|--------------------------|---------------|-------------|-----------|---------|-----------|
| | | | X_1 | X_2 | \dots | X_p |
| 1 | t_{11} | y_{11} | x_{111} | x_{211} | \dots | x_{p11} |
| 1 | t_{12} | y_{12} | x_{112} | x_{212} | \dots | x_{p12} |
| \vdots | \vdots | \vdots | \vdots | \vdots | | \vdots |
| 1 | t_{1t} | y_{1t} | x_{11t} | x_{21t} | \dots | x_{p1t} |
| 2 | t_{21} | y_{21} | x_{121} | x_{221} | \dots | x_{p21} |
| 2 | t_{22} | y_{22} | x_{122} | x_{222} | \dots | x_{p22} |
| \vdots | \vdots | \vdots | \vdots | \vdots | | \vdots |
| 2 | t_{2t} | y_{2t} | x_{12t} | x_{22t} | \dots | x_{p2t} |
| \vdots | \vdots | \vdots | \vdots | \vdots | | \vdots |
| K | t_{K1} | y_{K1} | x_{1K1} | x_{2K1} | \dots | x_{pK1} |
| K | t_{K2} | y_{K2} | x_{1K2} | x_{2K2} | \dots | x_{pK2} |
| \vdots | \vdots | \vdots | \vdots | \vdots | | \vdots |
| K | t_{Kt} | y_{Kt} | x_{1Kt} | x_{2Kt} | \dots | x_{pKt} |

Um dos aspectos mais relevantes dos estudos longitudinais é que devido a forma sistemática como as medidas são repetidas nos mesmos indivíduos, é razoável supor que exista uma heterocedasticidade das variâncias nas diversas ocasiões e uma correlação não nula entre as observações aninhadas no indivíduo, ou seja, as observações podem refletir uma tendência sistemática dentro de cada indivíduo. Essas suposições, além de exigir a modelagem da matriz de variâncias-covariâncias dos dados, que geralmente

não segue o padrão de uniformidade, viola algumas pressuposições básicas de muitas análises estatísticas como os GLMs (VONESH e CHINCHILLI, 1997).

4.3 Equações de Estimação Generalizadas

As equações de estimação generalizadas (GEE) foram desenvolvidas por LIANG e ZEGER (1986), a partir de uma generalização do método de estimação de quasi-verossimilhança de WEDDERBURN (1974) para dados longitudinais, com o objetivo de modelar a esperança da variável resposta como uma função linear de um conjunto de covariáveis, considerando a correlação entre as observações intraunidades de investigação. Assim, assumiram a variância como uma função conhecida da média e definiram uma matriz de correlação de trabalho.

Para análise de dados longitudinais, por meio das GEE, ao invés de especificarmos a distribuição conjunta do vetor de respostas multivariado para a i -ésima unidade de investigação, $i = 1, 2, \dots, K$, $\mathbf{Y}_i = (y_{i1}, y_{i2}, \dots, y_{it_i})'$, assumiremos que apenas é conhecida a distribuição marginal de y_{ij} , pertencente à família exponencial, em que:

1. A média marginal, $E[y_{ij}] = \mu_{ij}$, são modeladas como nos GLM, ou seja, $g(\mu_{ij}) = \eta_{ij}$, sendo $\eta_{ij} = \mathbf{x}'_{ij}\boldsymbol{\beta}$ o preditor linear, $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)'$, $p < n$, o vetor de parâmetros desconhecidos a serem estimados e $g(\cdot)$ uma função monótona e diferenciável, denominada função de ligação;
2. A variância marginal depende da média marginal pela relação $Var(y_{ij}) = \phi^{-1}v(\mu_{ij})$, onde $v(\cdot)$ é a função de variância, ϕ^{-1} é o parâmetro de dispersão; e
3. A correlação entre as observações de ordem j e k do vetor de respostas \mathbf{Y}_i , é função das médias marginais e, às vezes, de parâmetros adicionais α , isto é, $Corr(y_{ij}, y_{ik}) = \rho(\mu_{ij}, \mu_{ik}; \alpha)$.

Sendo assim, para a i -ésima unidade de investigação avaliada em t_i ocasiões, a matriz de variância-covariância para \mathbf{Y}_i , por definição, é dada por:

$$Var(\mathbf{Y}_i) = \phi^{-1}\mathbf{V}_i^{1/2}\mathbf{R}_i\mathbf{V}_i^{1/2} = \boldsymbol{\Sigma}_i^{-1}, \quad (4.1)$$

em que \mathbf{R}_i é a verdadeira matriz de correlação de \mathbf{Y}_i e \mathbf{V}_i é uma matriz diagonal de dimensão $t_i \times t_i$, cujo j -ésimo elemento da diagonal é igual a $v(\mu_{ij})$, $j = 1, 2, \dots, t_i$.

Perceba que para o caso de dados não correlacionados, temos simplesmente, $\mathbf{R}_i = \mathbf{I}_{t_i}$, e conseqüentemente $Var(\mathbf{Y}_i) = \phi^{-1}\mathbf{V}_i$.

Logo, a equação de estimação para obter as estimativas $\hat{\beta}$, do vetor de parâmetros de regressão, obtida a partir da introdução da estrutura de correlação, \mathbf{R}_i , na função escore utilizada em GLM, é dada por

$$S(\beta) = \sum_{i=1}^K \frac{\partial \mu_i}{\partial \beta} \Sigma_i^{-1} (Y_i - \mu_i) = 0, \quad (4.2)$$

Como na prática a verdadeira matriz de correlação, \mathbf{R}_i , é desconhecida. Para tornar este problema, LIANG e ZEGGER (1986) definiram $\mathbf{R}_i(\boldsymbol{\alpha})$ uma matriz simétrica, denominada *matriz de correlação de trabalho*, descrita completamente por um vetor de parâmetros de perturbação $\boldsymbol{\alpha}$, de comprimento finito s e que não dependem de β . Quando a unidade de investigação é observada em todas as t ocasiões, $\mathbf{R}_i(\boldsymbol{\alpha}) = \mathbf{R}(\boldsymbol{\alpha})$. No caso da i -ésima unidade de investigação não possuir todas as t observações, utiliza-se a matriz $\mathbf{R}_i(\boldsymbol{\alpha})$ obtida da matriz de correlação de trabalho $\mathbf{R}(\boldsymbol{\alpha})$, retirando-se desta as colunas e linhas correspondentes às observações faltantes.

Diante deste contexto, usando a matriz de correlação de trabalho, LIANG e ZEGGER (1986) definiram as *equações de estimação generalizadas* para estimar β , que é dada por

$$U(\beta) = \sum_{i=1}^K \frac{\partial \mu_i'}{\partial \beta} \Omega_i^{-1} (Y_i - \mu_i) = 0, \quad (4.3)$$

onde a matriz de variância-covariância de trabalho, Ω_i , do vetor de respostas Y_i , é dada por

$$\Omega_i = \phi^{-1} \mathbf{V}_i^{1/2} \mathbf{R}_i(\boldsymbol{\alpha}) \mathbf{V}_i^{1/2}. \quad (4.4)$$

O método de estimação via GEE é uma extensão natural do método de quase-verossimilhança proposto por Wedderburn (1974) para modelar respostas multivariadas em GLMs. É importante observar que, quando $R_i(\alpha)$ é a matriz identidade, $S_n(\beta)$ da expressão (4.3) reduz-se a função escore utilizada em GLM para estimação de β , ignorando-se a estrutura de correlação intraunidades de investigação. Detalhes sobre as demonstrações, condições de existência e teoria assintótica das equações GEE podem ser encontrados em HARDIN e HILBE (2003) e ARTES e BOTTER (2005).

O método GEE para estimar o vetor de parâmetros de regressão β permite que a estrutura da matriz de correlação de trabalho $R_i(\alpha)$ entre as observações do mesmo indivíduo seja especificada de diferentes formas. DIGGLE *et al.* (2002) e BALLINGER

(2004) sugerem algumas estruturas, apresentam um estimador para $\hat{\alpha}$ para cada caso e aconselham uma avaliação da natureza dos dados para uma escolha mais adequadas. Neste trabalho foram utilizaremos as seguintes estruturas:

- **Independente (*independence*):** considera as observações não correlacionadas. Nesta situação, GEE coincidem com as equações de estimação independentes, utilizadas no GLM. Isto implica que $R_i(\alpha)$ é a matriz identidade, ou seja, assume a forma

$$R_i(\alpha) = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}_{(t_i \times t_i)} .$$

- **Auto-regressiva de ordem 1 (*ar1*):** assume uma relação entre as correlações referentes aos tempos anterior e posterior, onde as correlações diminuem com o aumento das distâncias.

$$R_i(\alpha) = \begin{bmatrix} 1 & \alpha & \alpha^2 & \cdots & \alpha^{t_i-1} \\ \alpha & 1 & \alpha & \cdots & \alpha^{t_i-2} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \alpha^{t_i-1} & \alpha^{t_i-2} & \cdots & \alpha & 1 \end{bmatrix}_{(t_i \times t_i)} .$$

- **Não estruturada (*unstructured*):** tem diferentes correlações entre todos os instantes de avaliação e $t_i(t_i - 1)/2$ parâmetros a serem estimados e define-se $R_i(\alpha)$ como

$$R_i(\alpha) = \begin{bmatrix} 1 & \alpha_{12} & \cdots & \alpha_{1t_i} \\ \alpha_{12} & 1 & \cdots & \alpha_{2t_i} \\ \vdots & \vdots & \ddots & \vdots \\ \alpha_{1t_i} & \alpha_{2t_i} & \cdots & 1 \end{bmatrix}_{(t_i \times t_i)} .$$

- **Permutável (*exchangeable*):** assume que todas as correlações duas a duas sejam iguais a α . Neste caso, define-se $R_i(\alpha)$ como

$$R_i(\alpha) = \begin{bmatrix} 1 & \alpha & \cdots & \alpha \\ \alpha & 1 & \cdots & \alpha \\ \vdots & \vdots & \ddots & \vdots \\ \alpha & \alpha & \cdots & 1 \end{bmatrix}_{(t_i \times t_i)} .$$

4.3.1 Estimação dos parâmetros

Para estimação dos parâmetros β, α e ϕ , a partir do GEE, utiliza-se o processo iterativo dos mínimos quadrados ponderados que combina uma versão modificada do método score de Fisher para estimar β , com o método dos momentos para estimar os parâmetros ϕ e α . Neste processo, alterna-se em estimar β para valores fixos de $\hat{\phi}$ e $\hat{\alpha}$ e estimar (ϕ, α) , para valores fixos de $\hat{\beta}$. Tal método se encontra descrito em VENEZUELA *et al.* (2007).

Para este processo iterativo, podemos definir o seguinte algoritmo: (1) Calcula-se a estimativa inicial de $\widehat{\beta}^{(0)}$ através do modelo GLM, assumindo a independência entre as observações; (2) define-se a estrutura de correlação de trabalho $\mathbf{R}_i(\alpha)$; (3) A partir da estimativa de β , calcula-se os resíduos de Pearson, $\hat{r}_{ij} = \frac{(y_{ij} - \hat{\mu}_{ij})}{\sqrt{v(\hat{\mu}_{ij})}}$; (4) tais resíduos são então utilizados para produzir estimativas consistentes para o parâmetro de dispersão $\hat{\phi}^{(0)} = \frac{1}{(N-p)} \sum_{i=1}^K \sum_{j=1}^{t_i} \hat{r}_{ij}^2$, em que $N = \sum_{i=1}^K t_i$, e para os parâmetros de correlação $\hat{\alpha}^{(0)}$ pelo método dos momentos, considerando a suposta estrutura de $\mathbf{R}_i(\alpha)$; (5) obtém-se uma estimativa da matriz de variância-covariância, Ω_i , através da equação (4.4); e, (6) atualiza-se $\hat{\beta}$ através da equação (4.3), repetindo os passos (2) a (6) do processo até atingir a convergência.

Uma ótima característica da abordagem GEE é que, não é necessário que a matriz de correlação de trabalho de \mathbf{Y}_i esteja corretamente especificada, ou seja, que $\Omega_i = \text{Var}(\mathbf{Y}_i)$, para que $\hat{\beta}_G$, solução da equação 4.3, seja um estimador consistente e assintoticamente normal de β com matriz de variâncias, $\text{Var}(\hat{\beta}_G)$, estimada de maneira consistente por um estimador empírico, $\hat{\mathbf{V}}_G$, denominado na literatura estatística como *estimador robusto, ou sanduíche*. Entretanto, quanto mais próxima da verdadeira matriz de correlação estiver a escolha da matriz de correlação de trabalho, maior será a eficiência estatística dos estimadores de regressão, em especial quando a correlação entre as respostas for alta (FITZMAURICE, 2009). É importante ressaltar que, pelo fato do estimador robusto ser assintótico, suas propriedades são garantidas somente quando o número de unidades investigadas é grande. Para um número pequeno ($K < 20$) este estimador apresenta viés e recomenda-se usar o estimador de variância baseado no modelo, denominado na literatura estatística como *estimador naive* (PAIK, 1988).

4.4 Materiais e Métodos

4.4.1 Descrição dos dados

As análises foram realizadas a partir do Banco Nacional de Dados de Atuns e Afins (BNDA) mantido pelo Subcomitê Científico de Atuns e Afins (SCC de Atuns e Afins). O BNDA trás informações dos mapas e de observadores de bordo sobre a produção e o esforço pesqueiro, em cada evento de pesca, das embarcações pertencentes à frota atuneira brasileira (embarcações nacionais e arrendadas) que atuam com espinhel de superfície no oceano Atlântico. Após alguns dias de viagem, realizando eventos diários (geralmente consecutivos), as embarcações retornam ao porto de origem para descarregar a produção, retomando a atividade dias depois.

No banco de dados, além do nome e origem da embarcação, há informações de caráter temporal (ano, mês, dia, hora e data), geográfico (latitude e longitude), do desempenho quantitativo (número e peso para as diferentes espécies capturadas) e outras informações referentes às características da embarcação como número de intervalos e anzóis entre boias, e número total de anzóis utilizados no espinhel.

Para a análise dos dados sob o enfoque longitudinal e estudo da correlação entre os eventos de pesca diários, foi necessário definir um intervalo de tempo com um maior número de embarcações trabalhando em dias comuns e com menor período de inatividade. Com estes critérios, selecionamos 555 eventos de pesca realizados durante os meses de maio, junho e julho de 2001 (92 dias), referentes a oito embarcações arrendadas da Espanha. A partir daí, todas as observações provenientes dos 555 eventos de pesca, foram aninhadas nas embarcações que a originaram e ordenadas de acordo com o mês e dia de ocorrência. Trata-se de uma base de dados incompleta com 181 (92 dias x 8 barcos - 555) observações faltantes (*data missing*), relativas aos dias de inatividade dos barcos analisados.

4.4.2 Variáveis consideradas

Para o ajuste dos modelos GEE propostos, o número de espadartes capturados foi utilizado como variável resposta, adicionando ao preditor linear o logaritmo do número de anzóis como *offset*. Além disso, a captura por unidade de esforço (CPUE), para cada evento de pesca, foi definida pela equação

$$CPUE = \left(\frac{\text{número de espadartes capturados}}{\text{número de anzóis}} \right) \times 1000. \quad (4.5)$$

Quanto às variáveis explicativas, foram consideradas variáveis com prováveis influências na capturabilidade do espadarte. As do tipo categórica: dia (1 a 31), mês e turno (diurno: antes das 17 horas, noturno: caso contrário) em que ocorreu o evento de pesca; e as quantitativas: o tempo de imersão do espinhel (tp.ime, em horas), o número de anzóis entre boias (hpb), a latitude (lat, em graus) e a distância mínima à costa (min, em milhares de km), segundo as oito direções radiais, onde os ângulos utilizados foram 0° , 45° , 90° , 135° , 180° , 225° , 270° e 315° graus.

4.4.3 Modelos GEE propostos

Uma vez que nossa variável resposta "nº de espadartes capturados" é fruto de uma contagem, propomos o modelo GEE com distribuição marginal de Poisson considerando a ligação log. Para tanto, foi suposto que $\mathbf{Y}_i = (Y_{i1}, \dots, Y_{it_i})'$ é o vetor resposta multivariado com o número de espadartes capturados pelo i -ésimo barco, $i = 1, \dots, 8$, nos t_i instantes de pesca ($1 \leq t_i \leq 91$), assumindo que a distribuição marginal de Y_{ij} , para cada j , seja a Poisson, com $E(Y_{ij}) = \mu_{ij} = n_{ij}e^{\mathbf{x}_{ij}^T\boldsymbol{\beta}}$, sendo n_{ij} o número total de anzóis utilizados pelo i -ésimo barco no j -ésimo instante de pesca. Logo, teremos um modelo com parte sistemática

$$\log\mu_{ij} = \log n_{ij} + \mathbf{x}_{ij}^T\boldsymbol{\beta}, \quad (4.6)$$

em que $\boldsymbol{\beta}$ é o vetor de parâmetros a serem estimados e $\mathbf{x}_{ij} = (x_{ij1}, x_{ij2}, \dots, x_{ijp})'$, o vetor que contém os valores das p variáveis explicativas, sob estudo, para o barco i , em seu j -ésimo instante de pesca, $j = 1, 2, \dots, t_i$. Já o termo $\log n_{ij}$, denominado *offset*, é uma constante compensatória utilizada nos casos onde o número de capturas é utilizada como variável resposta em vez da taxa de capturas. A partir do modelo (4.6), teremos como CPUE estimada

$$\log\left(\frac{\mu_{ij}}{n_{ij}}\right) = \mathbf{x}_{ij}^T\boldsymbol{\beta},$$

$$\log(\widehat{CPUE}) = \mathbf{x}_{ij}^T\hat{\boldsymbol{\beta}}, \quad (4.7)$$

$$\widehat{CPUE}_{ij} = e^{\mathbf{x}_{ij}^T\hat{\boldsymbol{\beta}}} \quad (4.8)$$

Para o ajuste dos modelos foram utilizadas as estruturas de correlação temporal intrabarcos independente, permutável e auto-regressiva de ordem 1. Não utilizamos a não estruturada, pois a mesma é computacionalmente intensiva e causou travamento no R.

4.4.4 Critério para seleção da matriz de correlação de trabalho

Para a escolha da melhor estrutura de correlação de trabalho, utilizamos o critério de quasi-verossimilhança sob o modelo de independência (*Quasi-likelihood under the Independence model Criterion* - QIC), proposto por PAN (2001). O QIC é uma modificação do critério de informação de AKAIKE (AIC) (AKAIKE, 1974) para ser utilizado no método GEE, e é calculado a partir da comparação de um modelo com determinada matriz de correlação de trabalho com aquele utilizando a estrutura independente. Do mesmo modo que o AIC, quanto menor o valor do QIC, melhor a escolha da matriz. Alguns momentos podem ocorrer de os valores do QIC não serem muito diferentes. Visando aperfeiçoar o desempenho do QIC na escolha da matriz de correlação de trabalho, recentemente, HIN e WANG (2009) propuseram uma modificação do QIC, denotado por *Critério de Informação de Correlação* (*Correlation Injformation Criterion* - CIC). Neste trabalho utilizaremos os dois critérios.

4.4.5 Seleção e ajuste dos modelos

Para comparação de modelos GEE não podemos utilizar a deviança e o AIC, como nos GLM, pois tais medidas são baseadas na verossimilhança do modelo, e o método GEE utiliza a quasi-verossimilhança. Para solucionar este problema, PAN (2001) propôs uma versão corrigida do QIC, o QICc, que penaliza a complexidade do modelo, recompensando a parcimônia. Assim, para se definir a ordem de entrada das variáveis nos modelos investigados, cada variável explicativa foi ajustada sozinha, com a mesma variável resposta, e, em seguida, calculou-se o QICc de cada modelo. Selecionamos inicialmente a variável explicativa do modelo com menor QICc. Em seguida, os modelos foram ajustados considerando as variáveis previamente selecionadas, uma de cada vez, levando em conta sempre o modelo com a menor QICc. O processo continuou até que todas as variáveis explicativas estivessem incluídas no modelo sob investigação. Para assegurar a significância da inclusão de cada uma das variáveis explicativas, executamos o teste de Wald, proposto por ROTNITZKY e JEWELL (1990) entre os modelos encaixados com e sem cada termo. Este teste utiliza a estatística de Wald generalizada, que tem distribuição qui-quadrado com graus de liberdade igual ao número de parâmetros que são testados (VENEZUELA *et al.*, 2007). Ao final deste processo, tínhamos três modelos GEE de Poisson, não encaixados, com cada uma das estruturas de correlação propostas.

4.4.6 Técnicas de Diagnóstico

Visando identificar casos em que as estimativas do modelo GEE se distanciem muito do valor observado e detectar observações extremas que interferem desproporcionalmente nos resultados do ajuste, será realizado uma análise de diagnóstico dos modelos, a partir de uma adaptação das técnicas utilizadas para os GLM, em que considera-se a correlação intraunidades de investigação. Tais técnicas, baseada na análise dos resíduos, podem ser encontradas, para alguns dos modelos GEE, em CHANG (2000), que recomenda o uso dos resíduos de Pearson contra a ordem dos instantes de avaliação, em VENEZUELA *et al.* (2007) que consideram que as unidades experimentais são todas avaliadas nas t ocasiões, ou seja, que não há observações faltantes, e em DAMIANI (2012), que generalizou os métodos para bancos de dados com observações faltantes. Tais técnicas são baseadas no resíduo de Pearson (r_{ij}) (seção 4.3.1), na medida de alavanca, h_{ij} , elemento da diagonal principal da matriz de projeção ortogonal, na medida de alavanca referente ao i -ésima unidade h_i , definida por $h_i = \frac{1}{(t_i)} \sum_{j=1}^{t_i} h_{ij}$, e no resíduo padronizado $(r_s)_{ij}$, associado à observação y_{ij} , definido a partir da convergência do processo iterativo para estimar β e de h_{ij} , referente a j -ésima observação da i -ésima unidade de investigação.

Neste trabalho a análise de diagnóstico e a verificação da existência de observações atípicas no contexto dos Modelos GEE, será realizada por meio dos seguintes gráficos de diagnósticos:

- gráfico de r_{ij} contra os instantes de avaliação, para verificar a existência de pontos aberrantes.
- gráfico de $(r_s)_{ij}$ contra a ordem das observações, i , para verificar a existência de pontos aberrantes.
- gráfico de h_{ii} ou h_i contra a ordem das observações ou valores ajustados, para verificar a existência de pontos de alavanca.
- gráfico de $(CD)_{ij} = \frac{\hat{h}_{ij}}{p(1-\hat{h}_{ij})} r_{P_{ij}}^2$ (distância de Cook aproximada), contra a ordem das observações, para detectarmos possíveis pontos influentes.

4.4.7 Procedimento Computacional

Todas as análises foram realizadas com o programa R 3.0.3 (R Development CORE TEAM, 2014), por meio do pacote **geepack** (*Generalized Estimating Equation Package*) que permite, através da função `geeglm`, o ajuste dos Modelos GEE para dados longitudinais. Mais detalhes sobre algumas funções deste pacote, podem ser encontradas em HOJSGAARD *et al.* (2006). Utilizamos também a função `QIC` para obtenção do QIC e CIC dos modelos e a função `drop1`, ambas do pacote **MESS**, onde esta última retorna testes de Wald sequenciais para cada um dos termos do modelo, na ordem em que os mesmos foram adicionados ao mesmo.

4.4.8 Tratamento de dados ausentes

A função `geeglm` assume por padrão que as observações são igualmente separadas no tempo. Por isso, é preciso informar a função, sobre a existência de diferentes separações se houver valores ausentes. Como este é o nosso caso, criamos uma variável com valor numérico inteiro para o argumento `waves` da referida função. Esta variável identifica a ordem e o espaçamento das observações, para que as mesmas observações em diferentes barcos se correspondam corretamente. Como a base de dados foi ordenada de acordo com as embarcações e os instantes observados, utilizamos uma variável `instante` correspondente ao mês e dia no modo numérico.

4.4.9 Padronização da CPUE

A padronização da CPUE de espadarte foi realizada a partir do modelo GEE de Poisson com melhor ajuste, assumindo que os valores preditos calculados exclusivamente com os coeficientes do fator "mês", seriam índices relativos que refletiriam as variações mensais reais da abundância do estoque. Os efeitos dos demais fatores que interferem no coeficiente de capturabilidade alterando a CPUE bruta ficariam caracterizados pelos coeficientes do modelo calculados para as demais variáveis explicativas que não mês.

4.5 Resultados

4.5.1 Análise Exploratória

Nos 555 eventos de pesca analisados, ocorreram variações entre 0 e 32 unidades capturadas de espadarte. Sendo que apenas 6 destes, tiveram capturas iguais a zeros, correspondendo a aproximadamente a 1% dos dados (Figura 27). Percebe-se uma distribuição assimétrica a direita, sendo a média de espadartes capturados igual a 6,2 unidades, com variância 25,9. Havendo, portanto, indício de presença de sobredispersão nos dados.

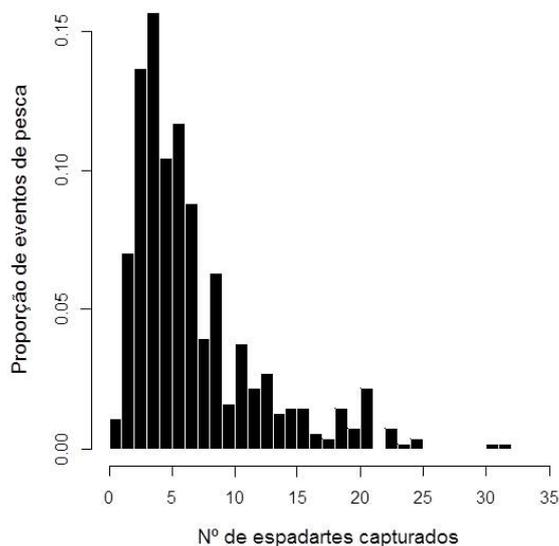


Figura 27: Histograma do número de espadartes capturados por eventos de pesca.

Os dados de captura e esforço de pesca foram agregados em blocos de $5^\circ \times 5^\circ$ de latitude e longitude e a CPUE nominal de espadarte no b -ésimo bloco foi assim calculada: $CPUE_b = (\sum C_b) / (\sum f_b) \times 1000$, em que C é o nº de espadartes capturados e f é o esforço de pesca nominal (número de anzóis). As capturas de espadarte ocorreram na porção oeste do Atlântico Sul, porém com CPUEs maiores na região localizada no bloco definido pelos limites $15^\circ - 20^\circ\text{S} / 20^\circ - 25^\circ\text{W}$ (Figura 28).

O gráfico de perfis individuais dos barcos, relativo à CPUE de espadarte (Figura 29) mostra que: para o mês de maio (instantes 1 a 31), com exceção dos barcos 1, 5 e 7, há uma alta variabilidade nas CPUEs intrabarcos, onde as mesmas são maiores nos primeiros dias, decrescendo ao longo do mês; para o mês de junho (instantes 32 a 61), com exceção dos barcos 1 e 3, onde houve decréscimo nas CPUEs ao longo do mês, as CPUEs apresentaram um padrão de estabilidade; com relação a julho (últimos 31 dias

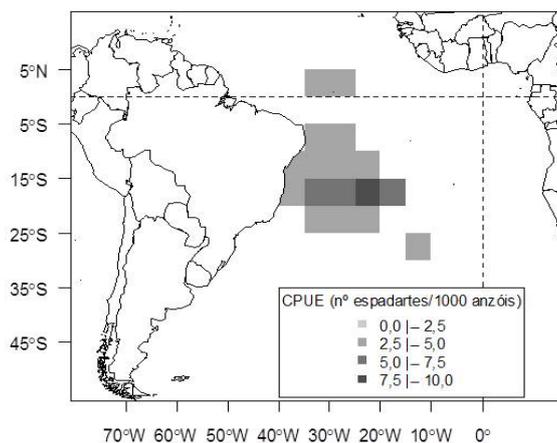


Figura 28: Distribuição geográfica da CPUE nominal de espadarte (número de peixes/1000 anzóis).

da série), de modo geral, as CPUEs se mantiveram praticamente inalteradas.

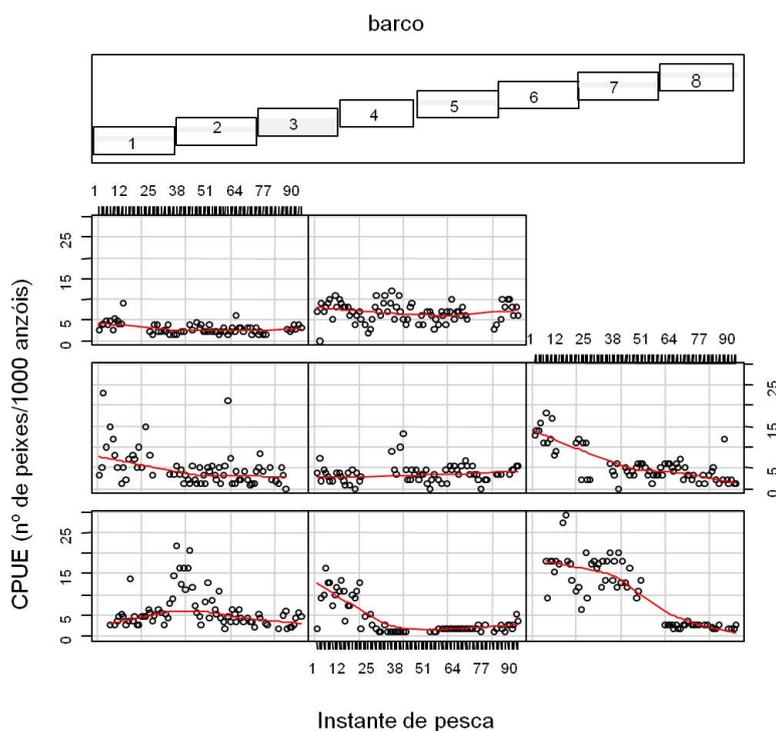


Figura 29: Perfis individuais dos barcos por instante de pesca, relativos à CPUE nominal de espadarte (nº de peixes capturados/1000 anzóis). Os instantes de 1 a 31 correspondem ao mês de maio, de 32 a 61 a junho e 61 a 92 a julho. A linha vermelha é uma linha suavizada (*Locally Weighted Scatterplot Smoother - LOWESS*), que mostra a tendência da CPUE.

Quanto ao perfil médio dos barcos, há uma tendência crescente no n^o médio de anzóis utilizados nos lances subsequentes de pesca no decorrer dos dias e meses (Figura 30A). Em paralelo, nota-se tendência geral decrescente na média das CPUEs ao longo dos instantes (Figura 30B). Há exceção são as pescarias realizadas no 3^o mês de avaliação (julho), quando houve certa estabilidade e até reversão do quadro de decréscimo nos últimos dias.

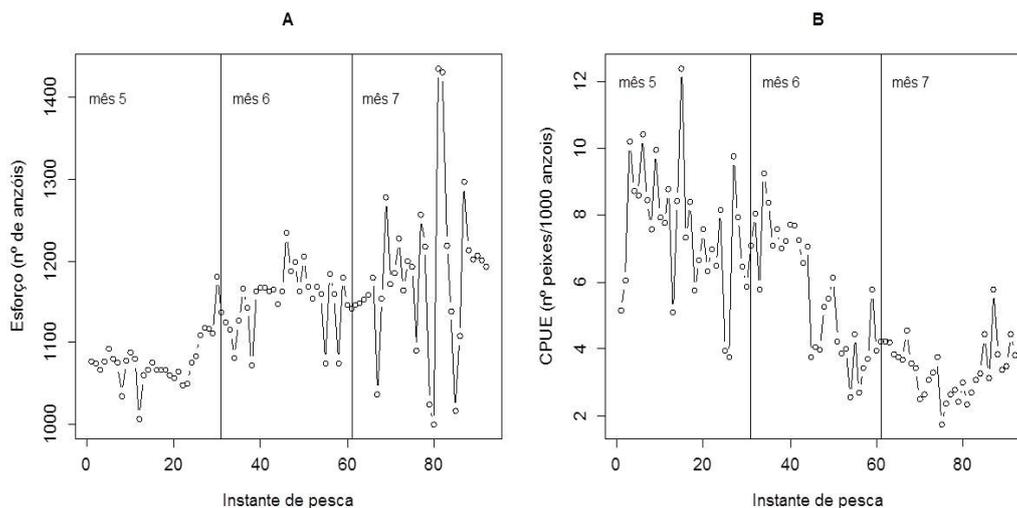


Figura 30: Perfil médio dos barcos relativo ao esforço de pesca (n^o de anzóis) (A), e relativo à CPUE (n^o de espadartes/1000 anzóis) (B).

4.5.2 Modelo GEE utilizado na padronização da CPUE do espadarte

Devido ao baixo número de barcos utilizados na análise, as estimativas dos parâmetros dos modelos de Poisson utilizados, foram calculadas a partir do estimador "naive". Além disso, devido ao forte indício de sobredispersão nos dados, o parâmetro de dispersão não foi fixado em um. Dentre os três melhores GEE ajustados com distribuição de Poisson, supondo cada uma das estruturas de correlação (independente, permutável e AR-1), o modelo com a estrutura AR-1 foi escolhido como o mais adequado para explicar a correlação entre as respostas intrabarcos, visto que apresentou o menor valor para o índice CIC. As estimativas dos parâmetros do referido modelo são apresentadas na Tabela 7.

A partir dos valores preditos com os coeficientes estimados (Tabela 7), podemos tirar as seguintes conclusões a respeito da influência de cada um dos termos, fixado os demais, sobre a CPUE de espadarte:

Tabela 7: Estimativas, erros padrão naive e p-valores relativo ao teste de Wald dos parâmetros do modelo GEE de Poisson com estrutura de correlação ar1. Siglas utilizadas: min - distância mínima à costa (x1000km), lat - latitude, hpb - n° de anzóis entre boias, tp.ime - tempo de imersão (horas) e significância: 0***, 0.001**, 0.01*, 0.05.

| Parâmetro | Estimativa | Erro padrão (naive) | Wald | Pr(> W) |
|------------------|------------|---------------------|----------|-----------|
| (Intercept) | -7.80418 | 0.90673 | 74.0797 | 0*** |
| hpb | 0.66273 | 0.18964 | 12.21227 | 0.0005 |
| min | 0.04214 | 0.09095 | 0.21466 | 0.6431 |
| turnonoturno | 0.11899 | 0.09297 | 1.63814 | 0.2006 |
| factor(dia)2 | -0.03887 | 0.0632 | 0.37832 | 0.5385 |
| factor(dia)3 | 0.19836 | 0.08148 | 5.92712 | 0.0149* |
| factor(dia)4 | 0.20204 | 0.09497 | 4.52642 | 0.0334* |
| factor(dia)5 | 0.10672 | 0.10734 | 0.98833 | 0.3201 |
| factor(dia)6 | 0.18951 | 0.11456 | 2.73637 | 0.0981 |
| factor(dia)7 | 0.00622 | 0.12497 | 0.00248 | 0.9603 |
| factor(dia)8 | 0.0501 | 0.12877 | 0.15139 | 0.6972 |
| factor(dia)9 | 0.03065 | 0.13475 | 0.05175 | 0.82 |
| factor(dia)10 | -0.08227 | 0.14338 | 0.32924 | 0.5661 |
| factor(dia)11 | -0.08348 | 0.1468 | 0.32335 | 0.5696 |
| factor(dia)12 | -0.06641 | 0.15003 | 0.19594 | 0.658 |
| factor(dia)13 | -0.05593 | 0.15393 | 0.13202 | 0.7164 |
| factor(dia)14 | -0.2388 | 0.16603 | 2.06865 | 0.1504 |
| factor(dia)15 | 0.08064 | 0.15689 | 0.26416 | 0.6073 |
| factor(dia)16 | -0.35676 | 0.1739 | 4.20871 | 0.0402* |
| factor(dia)17 | -0.17395 | 0.16616 | 1.09595 | 0.2952 |
| factor(dia)18 | -0.37586 | 0.17691 | 4.51403 | 0.0336* |
| factor(dia)19 | -0.01232 | 0.16206 | 0.00578 | 0.9394 |
| factor(dia)20 | -0.28351 | 0.16766 | 2.85934 | 0.0908 |
| factor(dia)21 | -0.30595 | 0.16646 | 3.37837 | 0.0661 |
| factor(dia)22 | -0.42582 | 0.17555 | 5.88369 | 0.0153* |
| factor(dia)23 | -0.3114 | 0.16592 | 3.52251 | 0.0605 |
| factor(dia)24 | -0.09434 | 0.15377 | 0.37641 | 0.5395 |
| factor(dia)25 | -0.70596 | 0.17967 | 15.43919 | 0.0001*** |
| factor(dia)26 | -0.31995 | 0.1562 | 4.19585 | 0.0405* |
| factor(dia)27 | -0.42534 | 0.16041 | 7.03106 | 0.008** |
| factor(dia)28 | -0.24294 | 0.14142 | 2.95083 | 0.0858 |
| factor(dia)29 | -0.47672 | 0.14495 | 10.81652 | 0.001** |
| factor(dia)30 | -0.38584 | 0.12997 | 8.8131 | 0.003** |
| factor(dia)31 | -0.47733 | 0.14148 | 11.38351 | 0.0007*** |
| factor(mes)6 | -0.19003 | 0.13064 | 2.11586 | 0.1458 |
| factor(mes)7 | -0.60324 | 0.21772 | 7.67689 | 0.0056** |
| lat | 0.00453 | 0.00905 | 0.25046 | 0.6167 |
| tp.ime | 0.04248 | 0.03324 | 1.63291 | 0.2013 |
| coef. dispersão | 3.34 | 0.453 | | |
| coef. correlação | 0.935 | 0.027 | | |

(1) a cada aumento de uma unidade no n° de anzóis entre boias (hpb), em particular, quando se aumenta de 3 para 4 anzóis, há um aumento de aproximadamente 94% ($e^{0.663}$) na CPUE; (2) a distância mínima à costa (min) não têm influencia significativa sobre a CPUE; (3) não há diferença significativa entre as CPUEs das pescarias ocorridas em ambos os turnos; (4) quanto ao fator dia, percebe-se um decréscimo

da CPUE a partir do 10º dia do mês, com poucos dias diferindo significativamente do dia de referência 1; (5) as CPUEs diminuem ao longo dos meses, sendo significativa a diferença entre o mês 7 e o mês de referência 5; (6) as CPUEs não se diferenciam significativamente com a latitude, e (7) o tempo de imersão do espinhel não tem influencia significativa sobre a CPUE;

4.5.3 Análise dos Resíduos

Na Figura (31) é apresentado o gráfico do resíduo de Pearson contra a ordem das observações e como podemos observar que praticamente todos os resíduos se distribuem simetricamente, indicando um bom ajuste do modelo com estrutura de correlação AR(1).

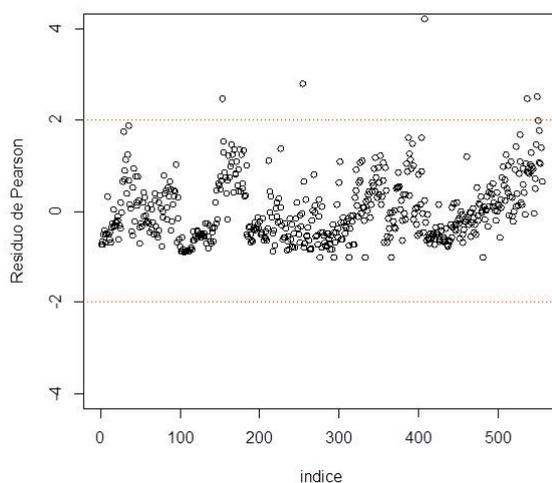


Figura 31: Gráfico do resíduo de Pearson referente ao modelo GEE de Poisson com estrutura de correlação AR(1) ajustado aos dados de capturas de espadarte.

Devido à singularidade da matriz Hessiana, alguns critérios de convergência não foram satisfeitos, o que impossibilitou a obtenção das medidas de alavanca e consequentemente do resíduo padronizado e da distância de Cook aproximada.

A partir do coeficiente de correlação estimado para a estrutura de correlação ar1, temos que $Corr(Y_{ij}, Y_{i(j+t)}) = 0.935^{t-1}$, $t = 1, \dots, 92$, evidenciando que as CPUEs observadas em dias consecutivos são altamente correlacionadas, e portanto, tendem a ser semelhantes. Contudo, esta correlação decresce com o aumento das distâncias relativas ao tempo. A matriz a seguir se refere ao cenário de um barco com as 92 observações referentes aos instantes de 1 a 92.

$$\mathbf{R}_i(\hat{\alpha}) = \begin{bmatrix} 1 & 0.935 & 0.874 & \cdots & 0.0020 \\ 0.935 & 1 & 0.935 & \cdots & 0.0024 \\ 0.874 & 0.935 & 1 & \cdots & 0.0025 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0.0020 & 0.0024 & 0.0025 & \cdots & 1 \end{bmatrix}_{(92 \times 92)} .$$

4.5.4 CPUE padronizada via GEE

A Figura 32 apresenta as séries de CPUE nominal e CPUE padronizada escalonadas (i.é. divididas pelas respectivas médias), em número de espadartes por lance com 1000 anzóis, obtida a partir do GEE de Poisson. As tendências de variação da CPUE padronizada tem padrão bastante similar à variação mostrada pela CPUE nominal, decrescendo ao longo dos meses. Contudo, a padronizada decresce de maneira menos acentuada do primeiro para o segundo mês de avaliação.

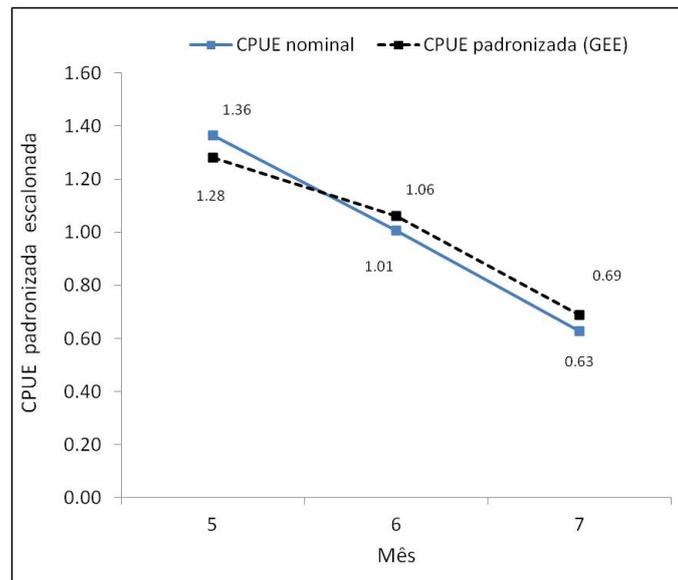


Figura 32: Captura por unidade de esforço (CPUE) nominal e padronizada escalonadas (nº de peixes/1000 anzóis) obtida através do modelo GEE de Poisson para o espadarte capturado pela frota atuneira brasileira no Atlântico Sul, de maio a julho, 2001.

4.5.5 Modelagem via GLM

A partir da mesma base de dados, mesma função de ligação e preditor linear utilizados para o ajuste dos modelos GEE, ajustamos o modelo GLM de Poisson onde é suposto a independência entre as observações intrabarcos. As estimativas dos parâmetros a partir deste modelo são apresentadas na Tabela 8.

Tabela 8: Estimativas, erros padrão e p-valores relativo aos parâmetros do modelo GLM de Poisson, ignorando as correlações entre as observações intrabarcos. Siglas utilizadas: min - distância mínima à costa (x1000km), lat - latitude, hpb - n° de anzóis entre boias, tp.ime - tempo de imersão (horas) e significância: 0^{***}, 0.001^{**}, 0.01^{*}, 0.05[.].

| Parâmetro | Estimativa | Erro padrão | z-valor | Pr(> z) |
|---------------|------------|-------------|---------|------------------------|
| (Intercept) | -6.37936 | 0.32894 | -19.39 | < 2e-16 ^{***} |
| hpb | 0.39892 | 0.04226 | 9.44 | < 2e-16 ^{***} |
| min | 0.15016 | 0.09423 | 1.59 | 0.111 |
| turnonoturno | 0.13452 | 0.04093 | 3.29 | 0.0010 ^{**} |
| factor(dia)2 | -0.04245 | 0.13377 | -0.32 | 0.751 |
| factor(dia)3 | 0.26421 | 0.12664 | 2.09 | 0.0370 [*] |
| factor(dia)4 | 0.20559 | 0.12176 | 1.69 | 0.0913 [.] |
| factor(dia)5 | 0.1177 | 0.12624 | 0.93 | 0.3512 |
| factor(dia)6 | 0.2031 | 0.12148 | 1.67 | 0.0945 [.] |
| factor(dia)7 | 0.06592 | 0.12455 | 0.53 | 0.5966 |
| factor(dia)8 | 0.09007 | 0.12085 | 0.75 | 0.4561 |
| factor(dia)9 | 0.13369 | 0.12374 | 1.08 | 0.28 |
| factor(dia)10 | 0.03845 | 0.12766 | 0.3 | 0.7633 |
| factor(dia)11 | 0.01968 | 0.12596 | 0.16 | 0.8758 |
| factor(dia)12 | 0.01611 | 0.12673 | 0.13 | 0.8989 |
| factor(dia)13 | -0.09586 | 0.1372 | -0.7 | 0.4848 |
| factor(dia)14 | -0.25021 | 0.14244 | -1.76 | 0.0790 [.] |
| factor(dia)15 | 0.12777 | 0.13842 | 0.92 | 0.3559 |
| factor(dia)16 | -0.24267 | 0.1405 | -1.73 | 0.0841 [.] |
| factor(dia)17 | -0.07537 | 0.13645 | -0.55 | 0.5807 |
| factor(dia)18 | -0.26037 | 0.14386 | -1.81 | 0.0703 [.] |
| factor(dia)19 | -0.07002 | 0.15102 | -0.46 | 0.6429 |
| factor(dia)20 | -0.26142 | 0.13754 | -1.9 | 0.0573 [.] |
| factor(dia)21 | -0.37682 | 0.13929 | -2.71 | 0.0068 ^{**} |
| factor(dia)22 | -0.27741 | 0.15227 | -1.82 | 0.0685 [.] |
| factor(dia)23 | -0.41569 | 0.15068 | -2.76 | 0.0058 ^{**} |
| factor(dia)24 | -0.12095 | 0.1394 | -0.87 | 0.3856 |
| factor(dia)25 | -0.66016 | 0.15348 | -4.3 | 1.7e-05 ^{***} |
| factor(dia)26 | -0.30112 | 0.14648 | -2.06 | 0.0398 [*] |
| factor(dia)27 | -0.14969 | 0.14666 | -1.02 | 0.3074 |
| factor(dia)28 | -0.14343 | 0.13384 | -1.07 | 0.2839 |
| factor(dia)29 | -0.36205 | 0.13951 | -2.6 | 0.0095 ^{**} |
| factor(dia)30 | -0.27244 | 0.13525 | -2.01 | 0.0440 [*] |
| factor(dia)31 | -0.18696 | 0.16019 | -1.17 | 0.2432 |
| factor(mes)6 | -0.26546 | 0.03952 | -6.72 | 1.9e-11 ^{***} |
| factor(mes)7 | -0.87791 | 0.05215 | -16.83 | < 2e-16 ^{***} |
| lat | 0.02869 | 0.00433 | 6.62 | 3.5e-11 ^{***} |
| tp.ime | 0.03508 | 0.016 | 2.19 | 0.0283 [*] |

A partir dos valores preditos com os coeficientes estimados (Tabela 7), podemos tirar as seguintes conclusões a respeito da influência de cada um dos termos, fixado os demais, sobre a CPUE de espadarte:

(1) a cada aumento de uma unidade no n^o de anzóis entre boias (hpb), em particular, quando se aumento de 3 para 4 anzóis, há um aumento de aproximadamente 50% ($e^{0.3989}$) na CPUE; (2) a distância mínima à costa (min) não têm influencia significativa sobre a CPUE; (3) a CPUE das pescarias ocorridas no turno noturno (após às 17:00hs) é 14% ($e^{0.134}$) maior do que a do diurno; (4) quanto ao fator dia, percebe-se um decrescimento das CPUEs a partir do 13^o dia do mês, com poucos dias diferindo significativamente do dia de referência 1; (5) as CPUEs ocorridas no mês 6 e 7 são significativamente menores do que a ocorrida no mês de referencia 5, decrescendo 23.3% no mês 6 e 58.5% no mês 7.

4.5.6 CPUE padronizada via GEE e GLM

A Figura 33 apresenta as séries de CPUEs padronizadas escalonadas via GLM e GEE de Poisson, em número de espadartes por lance com 1000 anzóis. Apesar de ambas apresentarem tendências decrescentes, a CPUE via GLM apresenta decrescimento mais acentuado. Enquanto a obtida via GLM cai 58% do primeiro para o terceiro mês, a obtida via GEE cai 46%.

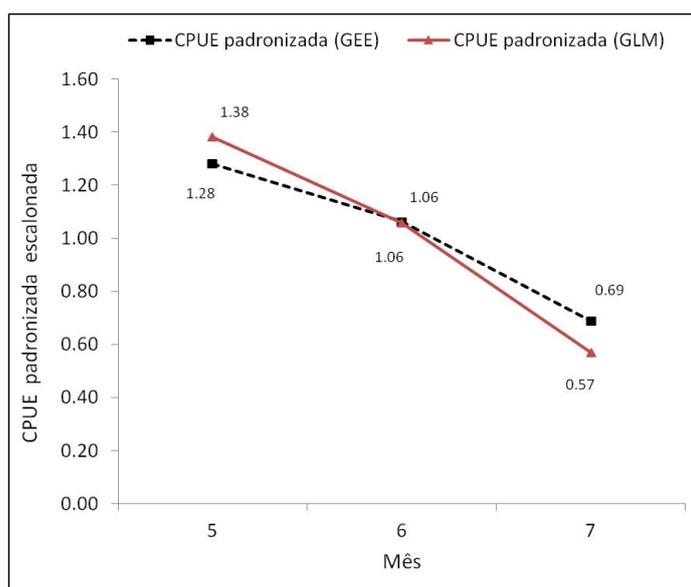


Figura 33: Séries de CPUEs padronizadas escalonadas (n^o de peixes/1000 anzóis), obtidas através dos modelos GEE e GLM de Poisson para o espadarte capturado pela frota atuneira brasileira no Atlântico Sul, de maio a julho, 2001.

4.6 Discussão

Em estudos envolvendo dados de capturas e esforços de pesca de peixes pelágicos por espinhéis de superfície, em particular de espadarte, as informações são provenientes dos mesmos barcos em pescarias realizadas ao longo do tempo. Neste sentido, supõe-se que a eficiência de um determinado barco em uma determinada ocasião, influencia os rendimentos obtidos em eventos posteriores, caracterizando um cenário longitudinal. Um das explicações para este fato, é que o conhecimento empírico dos pescadores sobre o fenômeno influencia na procura de áreas mais produtivas para a pesca (ARFELLI, 1996; HAZIN, 2006). Esta suposta correlação entre as observações intrabarcos de pesca ao longo do tempo inviabiliza a utilização dos GLM na modelagem da CPUE, visto que o mesmo é alicerçado na suposição de independência. Diante deste contexto, desenvolvemos uma análise dos dados de capturas e esforços de pesca do espadarte, baseada na teoria das equações de estimação generalizadas que incorpora o aspecto longitudinal e a correlação existente entre as observações intrabarcos.

Os resultados deste estudo, realizados a partir do modelo GEE de Poisson com estrutura de correlação AR-1, indicam que as maiores CPUEs de espadarte ocorrem quando o espinhel é acomodado mais superficialmente na coluna d'água, com um número de anzóis igual a quatro. Resultado que concorda com o apresentado em HAZIN (2006). Os efeitos da latitude e da distância mínima à costa não se mostraram significativos na abundância do espadarte. Acreditamos que este fato se deve a baixa variação na posição geográfica dos lançamentos de pesca reportados na base de dados utilizada na análise. Quanto ao fator dia, os valores negativos dos coeficientes, confirma uma redução na CPUE após nove dias consecutivos de pesca. Acreditamos que isto de deva a diminuição do poder atrativo das iscas, visto que os barcos mantiveram suas atividades de pesca, praticamente nas mesmas localidades em cada viagem de pesca.

A série de CPUE padronizada via modelo GEE mostrou uma tendência de queda na abundância relativa do espadarte ao longo dos meses, com valor significativamente inferior no mês de julho. Este fato se deve a maior disponibilidade de nutrientes promovida pela convergência subtropical presente na região Sul, e o conseqüentemente aumento de presas potenciais para o espadarte, como as lulas (*llex argentinus*) que são abundantes na região e lá permanecem do final do segundo ao terceiro trimestre, diminuindo a abundância da espécie na área sob estudo (SANTOS e HAIMOVICI, 2002; REVIZEE, 2006).

Uma das principais dificuldades do estudo foi a determinação da base de dados a

ser utilizada na análise, visto que, as informações do BNDA são advindas de eventos de pesca que ocorrem em tempos não controlados, gerando dados não balanceados e incompletos. Como a análise de dados longitudinais requer o acompanhamento de várias unidades em tempos simultâneos, foi necessário definir um intervalo de tempo com um maior número de embarcações trabalhando em dias comuns e com menor período de inatividade. Apesar do esforço empenhado, a base de dados selecionada para estudo, composta por oito embarcações acompanhadas durante o período de três meses, apresentou muitos dados faltantes (24,6%), em virtude da inatividade de alguns dos barcos ao longo dos 92 dias avaliados.

A combinação destas características da base de dados, poucos barcos em relação ao número de instantes de avaliação e praticamente 1/4 de observações faltantes, limitou bastante a utilização do modelo GEE utilizados na padronização da CPUE para o espadarte. Dentre as limitações, podemos citar: complicações computacionais, a impossibilidade da inclusão de interações no modelo, a singularidade da matriz Hessiana (necessária na verificação da qualidade do ajuste dos modelos), a não convergência do algoritmo para obtenção dos valores preditos a partir de combinações dos valores e níveis das variáveis explicativas e a diminuição da eficiência do estimador da matriz de covariância dos parâmetros do modelo. Além disso, o alto valor da estimativa do parâmetro de dispersão sugere a utilização do modelo binomial negativo, que é mais adequado para acomodar a sobredispersão (Hinde e Demétrio, 1998). Contudo, não foi possível utilizá-lo nas análises, visto que a função `geeglm` do pacote `geepack` do R, utilizada no ajuste de modelos GEE, não se encontra implementada para o uso do mesmo.

Apesar das limitações aqui relacionadas, a aplicação dos modelos GEE tornou possível a distinção entre o grau de variação na variável resposta para cada barco ao longo do tempo e a variação entre diferentes barcos, e com isso possibilitaram um aumento na qualidade das interpretações de casualidade a respeito da CPUE de espadarte. Diante deste contexto, se faz necessário que novas pesquisas sejam realizadas no âmbito da aplicação dos modelos GEE a dados de captura e esforço pesqueiro. Como proposta de trabalho futuro sugerimos o preenchimento dos dados faltantes com valores plausíveis através da técnica de imputação de dados, evitando assim, a complexidade gerada pelos dados faltantes. Entretanto, as incertezas associadas à imputação precisam ser tratadas apropriadamente porque os dados imputados não são dados reais.

4.7 Conclusão

A análise exploratória por meio do gráfico de perfis individuais dos barcos relativo à CPUE de espadarte revelou, de modo geral, a presença de uma alta variabilidade nas CPUEs intrabarcos, onde as mesmas são maiores nos primeiros dias e decresce ao longo de cada mês; Quanto ao perfil médio dos barcos analisados, embora tenha se constatado uma tendência crescente no esforço de pesca utilizado ao longo dos dias de cada mês avaliado, houve tendência decrescente na média das CPUEs de espadarte ao longo dos instantes de avaliação.

Em meio aos diversos modelos ajustados utilizando-se das equações de estimação generalizadas e das diversas estruturas de correlações, o modelo GEE com distribuição de Poisson e estrutura de correlação AR-1 se mostrou o mais adequado para modelagem dos dados de captura e esforço de pesca do espadarte, revelando alta correlação entre as CPUEs intrabarcos, entretanto, esta correlação diminui com o afastamento temporal das observações. A série de CPUE padronizada via modelo GEE mostrou uma tendência de queda na abundância relativa do espadarte ao longo dos meses de maio, junho e julho, sendo significativamente inferior no mês de julho. Entretanto, esta interpretação deve ser tomada com cautela devido ao baixo esforço de pesca empregado pelas embarcações amostradas e a sua pequena área de atuação em relação à distribuição do estoque de espadarte no Atlântico Sul.

Quanto à comparação da CPUE padronizada por ambos os métodos GLM e GEE, verificou-se que apesar de terem em comum uma tendência decrescente nos índices relativos de abundância, a série de CPUE padronizada via GLM apresenta decréscimo mais acentuado. Enquanto a obtida via GEE cai 46% do primeiro para o terceiro mês, a obtida via GLM cai 58%. Esta diferença na declividade das séries pode promover uma avaliação incorreta do estoque da espécie e medidas regulatórias equivocadas por parte da ICCAT.

5 Considerações Finais

O modelo binomial negativo se mostrou adequado para a padronização de dados de captura e esforço de pesca com porcentagem moderada de capturas nulas e presença de sobredispersão, particularmente comum para espécies pelágicas capturadas por espinhéis de superfície, possibilitando a mensuração de efeitos de variáveis espaciais, temporais e operacionais sobre a CPUE de espadarte.

Considerando a estimativa de CPUE padronizada de espadarte como um índice de abundância relativo, os resultados sugerem que a biomassa do estoque de espadarte do Atlântico Sul apresenta um padrão de estabilidade. Sendo assim, a pesca atuneira no Atlântico Sul parece não estar afetando fortemente o estoque da espécie, resultado que concorda com as conclusões e recomendações da ICCAT quanto à avaliação do referido estoque. Contudo, esta interpretação deve ser tomada com cautela, pois se analisaram dados relativos apenas ao período 1998 a 2006. Para se obter uma interpretação do estado atual do estoque, seria necessário estudos incluindo os anos mais recentes.

O GLM possibilitou a avaliação dos efeitos de vários fatores sobre a abundância dos estoques de espadarte, sendo um método eficiente para a padronização de dados de CPUE de espadarte. Entretanto, a não inclusão da correlação entre as observações intrabarcos na análise de padronização, possível causa da sobredispersão, sugere que as variações por ela produzidas sejam atribuídas a variações na abundância dos estoques e, portanto, as séries de CPUE que vierem a ser desenvolvidas podem não ser de boa qualidade.

A série de CPUE padronizada via modelo GEE mostrou uma tendência de queda na abundância relativa do espadarte ao longo dos meses de maio, junho e julho, sendo significativamente inferior no mês de julho. Entretanto, esta interpretação deve ser tomada com cautela devido ao baixo esforço de pesca empregado pelas embarcações amostradas e a sua pequena área de atuação em relação à distribuição do estoque de espadarte no Atlântico Sul.

O GLM é um processo estatisticamente mais robusto que o GEE e tem rotinas computacionais bem consolidadas. Contudo, apesar das limitações ocasionadas pela base de dados utilizada, a aplicação dos modelos GEE tornou possível a distinção entre o grau de variação na variável resposta para cada barco ao longo do tempo e a variação entre diferentes barcos, e com isso possibilitaram um aumento na qualidade das interpretações de casualidade a respeito da CPUE de espadarte. Concluí-se, portanto que, para se ter uma boa avaliação de estoque é necessário dispor de um modelo adequado para descrever a dinâmica do sistema, mas é também de fundamental importância a qualidade dos dados aos quais o modelo vai ser ajustado. Do contrário haverá a ocorrência de distorções nas estimativas dos parâmetros e erros padrão, levando a inferências incorretas sobre o índice de abundância do Espadarte.

A expectativa a partir do desenvolvimento deste trabalho, é de que as informações geradas possam subsidiar novas pesquisas no âmbito da aplicação dos modelos GEE a dados de captura e esforço pesqueiro, de modo a promover a diminuição de incertezas sobre os índices de abundância, na avaliação da biomassa e para a sustentabilidade na exploração do estoque de espadarte no Atlântico Sul. Apesar deste estudo ter como foco o espadarte, os métodos desenvolvidos podem ser utilizados para produzir índices de relativos de abundância para outras espécies capturadas por espinhéis de superfície.

Referências

- [1] ABID, N.; IDRISSE, M. **ICCAT MANUAL**, 1.ed, 2010. Capítulo 2.1.9: SWORDFISH, 2010.
- [2] AKAIKE, H. Information theory and an extension of the maximum likelihood principle. In: Petrov, B.N., Csaki, F. (Eds.), Proceedings of the 2nd International Symposium on Information Theory. **Publishing House of the Hungarian Academy of Sciences**, Budapest, p.268-281, 1973.
- [3] AKAIKE, H. A new look at the statistical model identification. **IEEE Trans. Automat. Control**, V.19, P.716-723, 1974
- [4] ALLEN, R.; PUNSLY, R. Catch rates as indices of abundance of yellow fin tuna, *Thunnus albacares*, in the eastern Pacific Ocean. **Inter-Amer. Trop. Tuna Comm. Bull.**, v.18, p.301-379, 1984.
- [5] ARFELLI, C. A. **Estudo da pesca e aspectos da dinâmica populacional de espadarte, *Xiphias gladius*, L. 1758, no Atlântico Sul**. 175 p. Tese de Doutorado Instituto de Biociências, Universidade Estadual Paulista, Rio Claro (SP), 1996.
- [6] ARTES, R; JORGENSEN, A. Longitudinal data estimating equations for dispersion models. **Scand. J. Stat. Theory Applications**, Stockholm, v.27, n.2, p. 321-334, 2000.
- [7] ARFELLI, C. A. ; AMORIN, A. F. Description of the brazilian swordfish fishery in Santos. **Collective Volume of Scientific Papers ICCAT**, v.27, p.315-317, 1988.
- [8] ARFELLI, C. A. ; AMORIN, A. F. Analysis of Santos (SP) longliners from southern Brazil. **Collective Volume of Scientific Papers ICCAT**, v.51, n.4, p.1359-1366, 2000.
- [9] ARREGUÍN-SANCHEZ, F. **Catchability: a key parameter for fish stock assessment**. Reviews in Fish Biology and Fisheries, v.6, p.221-242, 1996.
- [10] ASANO-FILHO, M. et al. A short note on the catch composition and weight-length relationship of tunas, billfishes and sharks of north Brazil. **Brazilian Journal of Oceanography**, v.52(3/4), p.249-253, 2004.
- [11] BANNEROT, S.P.; AUSTIN, C.B. Using frequency distributions of catch per unit effort to measure fish-stock abundance. **Trans. Am. Fish. Soc.** v.112, p.608-617, 1983.
- [12] ALLINGER, G. A. Using Generalized Estimating Equations for Longitudinal Data Analysis. **Organizational Research Methods**, v.7, n2, p:127-150, 2004.

- [13] BELLIDO, J.M.; PIERCE, G.J.; WANG, J. Modelling intra-annual variation in abundance of squid *Loligo forbesi* in Scottish waters using generalised additive models. **Fisheries Research**, v.52, p.22-39, 2001.
- [14] BERKELEY, S. A ; HOUDE, E.D. Age determination of broadbill swordfish, *Xiphias gladius*, from the Straits of Florida, using anal fin spine sections. **U.S. Dep. Comm., NOAA, Tech. Rep. NMFS**, v.8, p.137-143, 1983.
- [15] BEVERTON, R. J. ; HOLT, S. J. On the dynamics of exploited fish populations. **Fish. Invest. Ser. II. Mar. Fish. G.B. Minist. Agric. Fish. Food**, v.19, p.533, 1957.
- [16] BIGELOW, K.A.; BOGGS, C.H.; HE, X. Environmental effects on swordfish and blue shark catch rates in the US North Pacific longline fishery. **Fish. Oceanogr.** v.8, p.178-198, 1999.
- [17] BRANDÃO, A.; BUTTERWORTH, D.S.; JOHNSTON, S.J.; GLAZER, J.P. Using GLMM to estimate the somatic growth rate trend for male South African west coast rock lobster, *Jasus lalandii*. **Fisheries Research**, p.339-349, 2004.
- [18] BRESLOW, N. E. Extra-Poisson variation in log-linear models. **Applied Statistics**, v.33, p.38-44, 1984.
- [19] BRESLOW, N. E. ; CLAYTON, D. G. Approximate inference in generalized linear mixed models. **Journal of the American Statistical Association**, v.88, n.9-25, 1993.
- [20] BURNHAM, K.P.; ANDERSON, D.R., 2002. **Model Selection and Multi-model Inference: A Practical Information-theoretic Approach**, 2nd ed. Springer-Verlag, New York, 2002.
- [21] CAMERON, A. C.; TRIVEDI, P. K. Regression Analysis of Count Data, **Cambridge University Press**, Cambridge, 1998.
- [22] CAREY, F.G.; ROBISON, B.H. Daily patterns in the activities of swordfish, *Xiphias gladius*, observed by acoustic telemetry. **Fish. Bull.** v.79, p.277-292, 1981.
- [23] CHANG, Y. C. Residuals analysis oh the generalized estimating equations. **Stat Med.**; v.19, n.10, p.1277-93, 2000.
- [24] CHANG, S.K. Analysis of Taiwanese white marlin catch data and standardization of catch rates. **Collective Volume of Scientific Papers ICCAT** v.55, p.453-466, 2003.
- [25] COLLETTE, B.B. Xiphiidae. Peces espada. p. 1651-1652. In W. Fischer, F. Krupp, W. Schneider, C. Sommer, K.E. Carpenter and V. Niem (eds.) **Guia FAO para Identificación de Especies para lo Fines de la Pesca**. Pacifico Centro-Oriental. 3 Vols. FAO, Rome. 1995.
- [26] COOKE, J. G ; BEDDINGTON, J. R. The relationship between catch rates and abundance in fisheries. **IMA Journal of Mathematics Applied Medicine and Biology**, v.1, p.391-405, 1984.

- [27] CORDEIRO, G. M. **Modelos Lineares Generalizados**. Livro texto de minicurso, VII Simpósio Nacional de Probabilidade e Estatística, UNICAMP, Campinas, SP, 1986.
- [28] CORDEIRO, G.M. ; LIMA NETO, E.A. **Modelos Paramétricos**. Recife: Universidade Federal Rural de Pernambuco, Departamento de Estatística e Informática, 2006.
- [29] COX, D.R.; SNEL, E.J. A general definition of residuals. **Journal of the Royal Statistical Society B**, v. 30, p.248-275, 1968.
- [30] DALLAGNOLO, R. ; ANDRADE, H.A. Comprimento do espadarte (*Xiphias Gladius*) capturado pela frota de espinhel de superfície no sudoeste do Atlântico. **Braz. J. Aquat. Sci. Technol.**, v.10, n.1, p.51-55, 2006.
- [31] DAMIANI, L. P. **Diagnóstico para modelos lineares generalizados para dados com medidas repetidas e valores faltantes**. Dissertação de mestrado apresentada ao Instituto de Matemática e Estatística da Universidade de São Paulo, 2012.
- [32] DEAN, C. Testing overdispersion in poisson and binomial regression models. **Journal of the American Statistical Association**, v.87, p.451-457, 1992.
- [33] DIGGLE, P. J. et al. **Analysis of longitudinal data**. 2 ed., Oxford, Oxford University Press, 2002.
- [34] DOBSON, A. J. **An Introduction to Generalized Linear Models**. 2 ed., Chapman & Hall/CRC, London, 225p, 2002.
- [35] DOMINGO, A. et al. **ICCAT Manual**, chapter 3.1.2: Longline, 2014.
- [36] FAO. **The State of World Fisheries and Aquaculture**. Rome. 209 pp, 2012.
- [37] FARAWAY, J. J. **Extending the Linear model with R: Generalized Linear, Mixed Effects and Nonparametric Regression Models**. Chapman & Hall/CRC, London, 2006.
- [38] FITZMAURICE, G.M. et al. **Longitudinal Data Analysis**. Boca Raton: Chapman and Hall/CRC, 2009.
- [39] FRITSCHES, K. A.; BRILL, R. W. ; WARRANT, E. J. Warm eyes provide superior vision in swordfish. **Current Biology**, v.15, n.1, p.55-58, 2005.
- [40] GAVARIS, S. Use of a multiplicative model to estimate catch rate and effort from commercial data. **Can. J. Fish. Aquat. Sci.**, v.37, p.2272-2275, 1980.
- [41] GREENE, W. H. **Econometric analysis**. Prentice Hall, USA, 2002.
- [42] HALL, D. B. Zero-inflated Poisson and binomial regression with random effects: A case study. **Biometrics**, v.56, p.1030-1039, 2000.
- [43] HARDIN, J.W. ; HILBE, J.M. **Generalized Estimating Equations**, Chapman and Hall/CRC: New York, 2003.

- [44] HASTIE, T. ; TIBSHIRANI, R. **Generalized Additive Models**. Chapman and Hall, London, 1990.
- [45] HASTIE, T.; TIBSHIRANI, R.; FRIEDMAN, J. **The Elements of Statistical Learning: Data Mining, Inference, and Prediction**. Springer-Verlag, New York, 2001.
- [46] HAZIN, F. H. V. ; TRAVASSOS, P. E. A Pesca Oceânica no Século 21. **Ver. Bras. Eng. Pesca**, Pernambuco v.2, n.1, p.60-75, 2007.
- [47] HAZIN, F. H. V.; H. G. HAZIN; C. E. BOECKMANN; P. TRAVASSOS. Preliminary study on the reproductive biology of swordfish, *Xiphias gladius* (Linnaeus 1758), in the southwestern equatorial Atlantic ocean. **Collective Volume of Scientific Papers ICCAT**, v.54, n.5, p.1560-1569, 2002.
- [48] HAZIN, H. G. et al. Standardization of swordfish cpue series caught by Brazilian Longliners in the Atlantic Ocean, by GLM, using the targeting strategy inferred by cluster analysis. **Collective Volume of Scientific Papers ICCAT**, v.60, n.6, p.2039-2047. 2007.
- [49] HAZIN, H.G. et al. Standardized CPUE of swordfish (*Xiphias gladius*) caught by tuna longliners operating off southern Brazil (1998-2006). **Collective Volume of Scientific Papers ICCAT**, v.62, n.4, p.1175-1181, 2008.
- [50] HAZIN, H. G. et al. Standardized CPUE series of swordfish, *Xiphias gladius*, caught by Brazilian tuna fisheries in the Southwestern Atlantic Ocean. **Collective Volume of Scientific Papers ICCAT**, v.65, n.1, p.274-284, 2010.
- [51] HAZIN, H. G. et al. Standardized catch rates of swordfish (*Xiphias gladius*) caught by the Brazilian Fleet (1978-2012) using generalized linear mixed models (GLMM) using delta log approach. **Collective Volume of Scientific Papers ICCAT**, v.70, n.4, p.1875-1884, 2014.
- [52] HELSER, T.E.; PUNT A.E.; METHOT, R.D. A generalized linear mixed model analysis of a multi-vessel fishery survey. **Fisheries Research**. v.70, p.251-264, 2004.
- [53] HEDEKER, D.; GIBBONS, R.D. **Longitudinal Data Analysis**. New York, John Wiley & Sons, 2006.
- [54] HILBORN, R. ; WALTERS C. J. **Quantitative Fisheries Stock Assessment Choice, Dynamics and Uncertainty**, New York. Chapman & Hall, 1992.
- [55] HINDE, J.P. ; DEMÉTRIO, C.G.B. Overdispersion: models and estimation. **Computation Statistics and Data Analysis**, v.27, p.151-170. 1998a.
- [56] IN L. Y. ; WANG, Y. G. Working correlation structure identification in generalized estimating equations. **Stat Med.**, v.28, n.4, p.642-58, 2009.
- [57] HOJSGAARD, S.; HALEKOH, U.; YAN J. The R Package geepack for Generalized Estimating Equations, **Journal of Statistical Software**, v.15, n.2, p.1-11, 2006.

- [58] ICCAT. Recommendation by ICCAT regarding the implementation of an alternative option for the conservation of undersized Atlantic swordfish and the reduction of fishing mortality. **ICCAT Report for biennial period, 1996-97**. v.1, n.1, p.87-88, 1996.
- [59] ICCAT. Report of the 2006 Atlantic Swordfish Stock Assessment Session (Madrid, September 4 to 8, 2006). SCRS/2006/015 - SWO ATL Stock Assessment. **Collective Volume of Scientific Papers ICCAT**, v.60, n.6, p.1787-1896, 2007.
- [60] ICCAT. Report of the 2009 Atlantic Swordfish Stock Assessment Session (Madrid, September 7 to 11, 2009). SCRS/2009/016 - SWO ATL Stock Assessment. **Collective Volume of Scientific Papers ICCAT**, v.65, n.1, p.1-123, 2010.
- [61] ICCAT. **Annual ICCAT Meeting Press Release**, 2013.
- [62] ICCAT. **Report of the 2013 Atlantic Swordfish Stock Assessment Session**. Doc. No. SCI-036, 132p, 2014.
- [63] ICCAT. **ICCAT Statistical Bulletin (1960-2012)**, v.42, n.1, 187p, 2014a.
- [64] JACKMAN, S. **pscl: Classes and Methods for R Developed in the Political Science Computational Laboratory**, Stanford University. Department of Political Science, Stanford University, Stanford, California, 2008. R package version 0.95, URL <http://CRAN.R-project.org/package=pscl>.
- [65] KIMURA, D. K. Standardized measures of relative abundance based on modeling $\log(\text{CPUE})$ and their application to Pacific ocean perch (*Sebastes alutus*). **J. Cons. Int. Explor. Mer.**, v.39, p.211-218, 1981.
- [66] KOTAS, J. E. et al. A pesca de emalhe e de espinhel de superfície na Região Sudeste-Sul do Brasil, São Paulo: Instituto Oceanográfico USP. **Série documentos Revizee: Score Sul**, 2005.
- [67] KUHA, J. AIC; BIC: comparisons of assumptions and performance. **Sociological Methods Research**, v.33, n.2, p.188-229, 2004.
- [68] LAMBERT, D. Zero-inflated poisson regression, with an application to defects in manufacturing. **Technometrics**, v.34, p. 1-14, 1992.
- [69] LAWLESS, J. F. Negative binomial and mixed Poisson regression. **The Canadian Journal of Statistics**, v.15, p.209-225, 1987.
Leemis,
- [70] LEWSEY, J.D. ; THOMSON, W.M. The utility of the zero-inflated Poisson and zero-inflated negative binomial models: a case study of cross-sectional and longitudinal DMF data examining the effect of socio-economic status. **Community Dent Oral Epidemiol**, v.32, p.183-189, 2004.
- [71] LIANG, K. Y. ; ZEGER, S. L. Longitudinal data analysis using generalized linear models. **Biometrika**, v.73, p.13-22, 1986.

- [72] LO, N.; JACOBSON, L. ; SQUIRE, J. Indices of relative abundance from fish spotter data based on delta-lognormal models. **Can. J. Fish. Aquat. Sci.** v.49, p.2515-2526, 1992.
- [73] MAUNDER, M.N.; HARLEY, S.J. Methodological improvements to the EPO tuna stock assessments. **SCTB 16 Working Paper. Available from the Inter-American Tropical Tuna Commission**, 2003.
- [74] MAUNDER, M.N. ; PUNT, A. E. Standardizing Catch and Eeffort Data: A Review of Recent Approaches. **Fish. Res.**, v.70, p.141159, 2004.
- [75] MAUNDER, M.N.; PUNT, A.E. Standardizing catch and effort data: a review of recent approaches. **Fish. Res.**, v.70, p.141159, 2004.
- [76] MCCULLAGH, P. ; NELDER, J. A. **Generalized Linear Models**, Second Edition. Chapman and Hall, London, 1989.
- [77] MCCULLOCH, C. E. ; SEARLE, S. R. **Linear and Generalized Linear Mixed Models**. Wiley, New York, 2001.
- [78] MEJUTO, J.; HOEY, J.J. An approach to a stock hypothesis for the swordfish, *Xiphias gladius*, of the Atlantic Ocean. **Collective Volume of Scientific Papers ICCAT**. v.35, p.482501, 1991.
- [79] MELLO, R.M. **Análise de conteúdo estomacais, intensidade de alimentação, idade e crescimento do espadarte, *Xiphias gladius* (Xiphiodei: Xiphiidae), no sul do Brasil**. Rio Grande. FURG. 223p, 1992.
- [80] MOLENBERGHS, G. ; VERBEKE, G. **Models for discrete longitudinal data**, New York, Springer, 2005.
- [81] MONTOYA, A.G.M. **Inferência e diagnóstico em modelos para dados de contagem com excesso de zeros**. 107p. Dissertação (Mestrado em Estatística) - Instituto de Matemática, Estatística e Computação Científica, Universidade Estadual de Campinas, Campinas, SP, 2009.
- [82] MOUATASSIM, Y.; EZZAHID, E. H. (2012). Poisson regression and Zero-inflated Poisson regression: application to private health insurance data. **European actuarial journal**, v.2, n.2, 2012.
- [83] MOURATO, B.L. et al. Standardized catch rate of swordfish (*Xiphias gladius*) caught by Santos longliners off southern Brazil (1971-2005). **Collective Volume of Scientific Papers ICCAT**, v.60, n.6, p.1943-1952, 2007.
- [84] MPA . **Boletim Estatístico da Pesca e Aquicultura 2010** - Ministério da Pesca e Aquicultura - Brasil . Brasília, 2012.
- [85] MPA . **Boletim Estatístico da Pesca e Aquicultura 2011** - Ministério da Pesca e Aquicultura - Brasil. Brasília, 2013.
- [86] NAKAMURA, I. 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.**, v.125, p.5-65, 1985.

- [87] NELDER, J. A.; WEDDERBURN, R. W. M. Generalized linear models. **Journal of the Royal Statistical Society, Series A**, v.135, p.370-384, 1972.
- [88] ORTIZ, M.; LEGAULT, C. M. ; EHRHARDT, N. M. An alternative method for estimating by catch from the U.S. shrimp trawl fishery in the Gulf of Mexico, 1972-1995. **Fish. Bull.**, v. 98, p. 583599, 2000.
- [89] ORTIZ, M.; AROCHA, F. Alternative error distribution models for standardization of catch rates of non-target species from a pelagic longline fishery: billfish species in the Venezuelan tuna longline fishery. **Fish. Res.** v.70, p.275294, 2004.
- [90] PAIK, M. C. Repeated Measurement Analysis for Nonnormal Data in Small Samples. **Communications in Statistics - Simulation and Computation**, v.17, p.11551171, 1998.
- [91] PALKO, B.J.; BEARDSLEY, G.L.; RICHARD W.J. Synopsis of the biology of swordfish, *Xiphias gladius* Linnaeus. NOAA Tech. Rep., Seattle, NMFS/S., USA. 1981.
- [92] PAN, W. Akaike's Information criterion in generalized estimating equations. **Biometrics**, v.57, n.1, p.120-125, 2001.
- [93] PAULA, G.A. **Modelos de Regressão com Apoio Computacional**. Instituto de Matemática e Estatística, p. 252, Universidade de São Paulo, 2013.
- [94] PHANG, Y. N. ; LOH, E. F. Zero Inflated Models for Overdispersed Count Data. World Academy of Science, Engineering and Technology. **International Journal of Mathematical, Computational Science and Engineering** v.7, n.8, 2013.
- [95] POTTS, J. M.; ELITH, J. Comparing species abundance models. **Ecological Modelling**, v.199, p. 153163, 2006.
- [96] PUNSLY, R. G. ; NAKANO, H. Analysis of variance and standardization of longline hook rates of bigeye (*Thunnus obesus*) and yellowfin (*Thunnus albacares*) in the eastern Pacific Ocean during 1975 - 1987. **Inter-Amer. Trop. Tuna Comm. Bull.**, v. 20, p.165-184, 1992.
- [97] PUNT, A.E.;WALKER, T.I.;TAYLOR, B.L.; PRIBAC, F. Standardization of catch and effort data in a spatially-structured shark fishery. **Fish. Res.** v.45, p.129145, 2000.
- [98] QUINN, T.J. ; DERISO, R.B. **Quantitative fish dynamics**. Oxford University Press, Oxford, England, 1999.
- [99] R Development Core Team. 2014. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>.
- [100] REVIZEE. Avaliação do Potencial Sustentável de Recursos Vivos na Zona Econômica Exclusiva: Relatório Executivo. **Programa de Avaliação do Potencial Sustentável de Recursos Vivos na Zona Econômica Exclusiva (REVIZEE)**. Ministério do Meio Ambiente, Secretaria de Qualidade Ambiental, Brasília, DF, 279 pp, 2006.

- [101] RICHARDS, L. J. ; SCHNUTE, J. T. An experimental and statistical approach to the question: is CPUE an index of abundance? . **Can. J. Fish. Aquat. Sci.**, v.43, p.12141227, 1986.
- [102] RIDOUT, M.; DEMETRIO, C. G. B; HINDE, J. Models for Count Data with Many Zeros, **presentation a International Biometric Conference**, 1998.
- [103] ROBSON, D. S. Estimation of the relative fishing power of individual ships. **ICNAF Res. Bull.**, 3: 514, 1966.
- [104] ROTNITZKY, A.; JEWELL, N. P. Hypotesis testing of regression parameters in semi parametric generalized linear models for cluster correlated data. **Biometrika**; v.77, n.3, p.485-97, 1990.
- [105] RODRÍGUEZ-MARÍN, E. et al. Standardization of bluefin tuna, *Thunnus thynnus*, catch per unit effort in the baitboat fishery of the Bay of Biscay (Eastern Atlantic). **ICES J. Mar.Sci.**, v.60, p.12161231, 2003.
- [106] SARKAR, D. **Lattice: multivariate data visualization with R**. Springer, Nova Iorque. p.265 , 2008.
- [107] SCHWARZ, G. Estimating the dimension of a model. **Annals of Statistics**, v.6, p.461-464, 1978.
- [108] SINGER, J. M. ; ANDRADE, D. A. **Analysis of Longitudinal Data**. In *Handbook of Statistics* (Eds., P. K. Sen and C. R. Rao). v.18, p.115-160, 2000.
- [109] SMYTH, G. K. **Pearsons Goodness of Fit Statistic as a Score Test Statistic**. Monograph Series, Volume 40, Institute of Mathematical Statistics, Hayward, California, 2003.
- [110] VENABLES, W. N. ; RIPLEY, B. D. **Modern Applied Statistics with S**. 4th ed. Springer Verlag, New York, 2002.
- [111] VENABLES, W. N. ; DICHMONT, C. M. GLMs, GAMs and GLMMs: an overview of the theory for applications in fisheries research. **Fish. Res.**, v.70, p.319-337, 2004.
- [112] VENEZUELA, M.; BOTTER, D.; SANDOVAL, M. Diagnostic techniques in generalized estimating equations, **Journal of Statistical Computation and Simulation**, v.77, p.879888, 2007.
- [113] VONESH, E. F. ; CHINCHILLI, V. M. **Linear and nonlinear models for the analysis of repeated measurements**, New York: Marcel Dekke, 1997.
- [114] VUONG, Q.H. Likelihood ratio tests for model selection and non-nested hypotheses. **Econometrica**, p.307334, 1989.
- [115] WAGENMAKERS, E-J ; FARRELL, S. AIC model selection using Akaike weights (Notes and Comment). **Psychonomic Bulletin & Review**, 2004, v.11, n.1, p.192-196, 2004.

- [116] WALSH, W.A. ; KLEIBER, P. Generalized additive model and regression tree analyses of blue shark (*Prionace glauca*) catch rates by the Hawaii-based commercial longline fishery. **Fisheries Research**, v.53, p.115-131, 2001.
- [117] WALSH, W.A. et al. Analysis of logbook accuracy for blue marlin (*Makaira nigricans*) in the Hawaii-based longline fishery with a generalized additive model and commercial sales data. **Fisheries Research**, v.75, p.175-192, 2005.
- [118] WEDDERBURN, R. W. M. Quasi-likelihood functions, generalized linear models and the Gauss-Newton method. **Biometrika**, v.61, p.439-447, 1974.
- [119] WEISS, R. E. **Modeling Longitudinal Data**. New York: Springer. 2005.
- [120] WELSH, A. H. et al. Modelling the abundance of rare species: statistical models for counts with extra zeros. **Ecological Modelling**, v.88, p.297-308, 1996.
- WILLIAMS, D. A. Generalized linear model diagnostic using the deviance and single case deletion. **Applied Statistics**, v.36, p.181-191, 1987.
- [121] YAU, K.K.W.; WANG, K. & LEE, A.H. Zero-inflated negative binomial mixed regression modeling of over-dispersed count data with extra zeros. **Biometrical journal**, v.45, p. 437-452, 2003.
- [122] XIE, M., HE, B. ; GOH, T. N. Zero-inflated Poisson model in statistical process control. **Computational Statistics & Data Analysis**. Amsterdam, v.38, p.191-201, 2001.
- [123] ZEGER S.L. ; LIANG K.Y. Longitudinal Data Analysis for discrete and continuous outcomes. **Biometrics**, v.42, p.121-130, 1986.
- [124] ZEGER S.L.; LIANG K.Y. ; ALBERT, P.S. Models for Longitudinal Data: A Generalized Estimating Equation Approach. **Biometrika**, v.44, p.1049-1060, 1988.
- [125] ZEILEIS, A.; KLEIBER, C.; JACKMAN, S. Regression Models for Count Data in R. **Journal of Statistical Software**, v.27, n.8, p.1-25, 2008.